

# **Multilingual Access to the European Cultural Heritage**

**Multilingual Websites and Thesauri**

*“It is time to think multilingual”*

**Multilingual Access  
to the European Cultural Heritage  
Multilingual Websites and Thesauri**

**General co-ordination**

Rossella Caffo (MINERVA and MINERVA Plus Project Manager)  
Antonella Fresa (MINERVA and MINERVA Plus Technical Coordinator)  
Pier Giacomo Sola (MINERVA and MINERVA Plus  
Organisation Manager)

**Secretariat**

Marzia Piccininno (Ministero per i beni e le attività culturali, Italy)

**Web version**

Maria Teresa Natale and Andrea Tempera  
<http://www.minervaeurope.org/publications/multilingualismandthesaurus.htm>

**Editorial Committee**

Stephan Conrad (Germany)

Christophe Dessaux (France), Kate Fernie (United Kingdom),  
Antonella Fresa (Italy), Allison Kupietzky (Israel), Marzia Piccininno (Italy),  
Martina Rozman Salobir (Slovenia), Gabriella Szalóki (Hungary)

**Contributors** (alphabetic order by countries)

Jitka Zamrzlová (Czech Republic)  
Marju Reismaa (Estonia)  
Minna Kaukonen (Finland)  
Véronique Prouvost (France)  
Dimitrios A. Koutsomitropoulos (Greece)  
Astrid Müller (Germany)  
Giuliana De Francesco (Italy)  
Domitilla Fagan (Ireland)  
Laila Valdovska (Latvia)  
Guy Frank (Luxembourg)  
Pierre Sammut (Malta)  
Jos Taekema (The Netherlands)  
Lars Egeland (Norway)  
Piotr Ryszewski (Poland)  
Maria Sliwinska (Poland)  
Ana Alvarez Lacambra (Spain)  
Martina Roznan Salobir (Slovenia)  
Elena Kuzmina (Russian Federation)  
Martin Katuscak (Slovak Republic)



## Acknowledgements

*We dedicate this report  
to the memory of the late Stephen Conrad*

From February 2004 ten new member states (plus Russia and Israel) have been participating in the joint European initiative of MINERVA Plus working with MINERVA to coordinate digitization efforts and activities. Since then MINERVA Plus supplementary working groups (SWG) started operation and Hungary became the coordinator of SWG Multilingual thesauri. The issue of multilingualism is becoming more and more important in making the digital cultural heritage of Europe available. Language is one of the most significant barriers to access of websites and, because of this barrier, great parts of the European digital cultural heritage cannot be found on the Internet.

MINERVA Plus conducted a major survey to get an overview of the situation concerning language usage in cultural websites. The aim of the survey was to see to what extent cultural websites and portals are available for users of different language communities and also whether websites use more languages than the language they were originally created in. Furthermore the survey intended to find out if cultural websites are using retrieval tools such as controlled vocabularies or thesauri and whether multilingual tools are available for use.

The methodology used for our survey included a questionnaire completed on a voluntary basis by our target group: libraries, museums, archives and other cultural institutions operating websites. The selection of the websites was not scientifically founded and so the sampling is not statistically representative. Nevertheless, the survey yielded a general picture of multilingualism of cultural websites and the findings will be a good starting point for more systematic and statistically valid research in the future.

I would like to thank our Israeli colleagues for letting us use their questionnaire (*Registry of Controlled Vocabularies related to Jewish Cultural Heritage and Israel*) as basis for our survey.

I am also very grateful to our respondents for collecting and mailing the requested information.

Last but not least I would like to express my gratitude to the editorial board of this document.

Iván Rónai  
NRG member for Hungary



# Table of Contents

<b>Executive summary</b>	9
<b>1 Introduction</b>	11
1.1 What is multilingualism? The European context	11
1.2 Why multilingualism is important?	13
1.3 Population languages spoken in the Member States	15
<b>2 The survey of multilingual websites and thesauri</b>	25
2.1 About the working group of multilingualism	25
2.2 Overview of aims, scope and methodology of the survey	28
2.3 Findings and final results	29
2.3.1 In the first run	29
2.3.2 In the second run	31
<b>3 Thesauri and controlled vocabularies used in the different countries</b>	37
<b>4 Best practice examples</b>	51
4.1 Best practice examples for multilingual thesauri	51
4.2 Best practice examples for multilingual websites	59
4.2.1 Best practice of multilingual websites with thesaurus	60
4.2.2 Best practice of multilingual websites with free-text indexing	65
<b>5 Conclusions</b>	73
<b>6 Future perspectives</b>	77
<b>Annex 1</b> Definition of terms used in the survey	79
<b>Annex 2</b> Interesting projects	82
<b>Annex 3</b> Registered thesauri on the survey's website	83





## Executive summary

This document was created for cultural institutions to emphasize the importance of multilingualism, and to provide them information and tools for establishing multilingual access to their collections.

In the **Introduction** we highlight the meaning and importance of multilingualism. It goes back to the historical times with its well known symbol *The Tower of Babel*. But now, it is time to “*think multilingual*”. The first lesson we have learned during the joint work, that it is substantial to learn about the official and minority languages, and their legislation within the different countries. That is why we started to collect country reports, to present the complexity of the multilingualism in details. This information should be the starting point in each European Union project, because it helps to understand the differences between the countries.

In **The survey of multilingual websites and thesauri** chapter we summarize the whole survey process carried out by the WP3 working group in the scope of the MINERVA Plus Project. The aim of the survey was mapping the multilingualism of the cultural sites and collecting information on multilingual thesauri in use. The survey lasted for a year from June 2004 to June 2005 in two runs. The findings of the Survey of multilingual websites and thesauri are presented here. We have statistical information about the types of institutions, which registered their websites, how many of them monolingual, and multilingual, how many of them are available in English. And finally how many of them use controlled vocabularies for information retrieval.

One of the practical aims of the MINERVA Project is to share the **Best practice examples**. The representatives were asked to nominate the best practice examples for multilingual websites and thesauri of their countries. We have summarized the results of the nominations for **Best practice examples for multilingual thesauri** and introduced some of them in detail, which are already in use in many different countries. In the survey we collected 657 multilingual websites<sup>1</sup> from all over Europe. We present the **Best practice examples of multilingual cultural websites**, which

---

<sup>1</sup> MINERVA Institutions: <<http://www.minervaeurope.org/institutions.htm>>.

are available in two or more languages, and meet the requirements of the 7th chapter of the *Quality Principles for cultural websites: a handbook*<sup>2</sup> published by the MINERVA Plus WP5 working group. Some of them implements thesaurus for information retrieval.

From the results, and findings we derived the **Conclusions** about the importance of multilingualism, and the use of multilingual thesauri.

We also made same proposal for the future in the **Future perspectives** about supporting the translation of the well-tested thesauri, the quality test beds for thesauri, and the further collection of multilingual thesauri.

---

<sup>2</sup> <<http://www.minervaeurope.org/publications/qualitycriteria.htm>>.

# 1 Introduction

## 1.1 What is multilingualism? The European context

*Immer werden jene vonnöten sein,  
die auf das Bindende zwischen den Völkern  
jenseits des Trennenden hindeuten und im Herzen  
der Menschheit den Gedanken eines kommenden  
Zeitalters höherer Humanität gläubig erneuern*

*There will always be necessary those  
who look on the binding parts between  
peoples beyond the separating ones,  
reinvigorating, in the heart of mankind,  
the thought of a forthcoming century of  
superior humanity.*

Stefan Zweig,  
*Triumph und Tragik des Erasmus von Rotterdam*

"Multilingualism refers to both a person's ability to use several languages and the co-existence of different language communities in one geographical area."<sup>3</sup> In fact, the more languages you know, the more of a person you are, says the proverb that opens the Commission's communication on multilingualism.

The European Commission adopted in November 2005 the communication to the Council "New Framework Strategy for Multilingualism" document<sup>4</sup>, which underlines the importance of

---

<sup>3</sup> Communication from the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions - A New Framework Strategy for Multilingualism COM(2005) 596 final Brussels, 22.11.2005, <<http://europa.eu.int/languages/servlets/Doc?id=913>>.

<sup>4</sup> European Commission press release, <<http://europa.eu.int/rapid/pressReleasesAction.do?reference=IP/05/1451&format=HTML&aged=0&language=EN&guiLanguage=en#fn1>>.

multilingualism and introduces the European Commission's multilingualism policy.

"The Commission's multilingualism policy has three aims:

to encourage language learning and promoting linguistic diversity in society;  
to promote a healthy multilingual economy, and  
to give citizens access to European Union legislation, procedures and information in their own languages."<sup>5</sup>

Ever since the European Year of Languages in 2001<sup>6</sup> was organised by the European Council, the European Day of Languages has been held every September 26 to help the public appreciate the importance of language learning, to raise awareness of all the languages spoken in Europe and to encourage lifelong language learning. It is a celebration of Europe's linguistic diversity.

The European Commission has also launched recently a new portal for European languages<sup>7</sup>, which is available in all the 20 official languages of the European Union. It is a useful information source of multilingualism and can be a starting point for every project. The resource given has been prepared for the general public and covers a range from the Union's policies to encourage language learning and linguistic diversity. The main areas covered are:

- linguistic diversity
- language learning
- language teaching
- translation
- interpretation
- language technology

A wide range of information is given for each of them from EU and national rules to a round up of employment opportunities for professional linguists with the Union's institutions. In fact, the Communication also stresses the importance of language skills to worker mobility and the competitiveness of the EU economy. The Commission will publish a study next year on the impact on the European economy of shortages of languages skills.

It is worth mentioning the Eurobarometer<sup>8</sup> survey published on the web site that was carried out between May and June 2005 among European

---

<sup>5</sup> <<http://europa.eu.int/languages/servlets/Doc?id=913>>.

<sup>6</sup> <[http://europa.eu.int/comm/education/policies/lang/awareness/year2001\\_en.html](http://europa.eu.int/comm/education/policies/lang/awareness/year2001_en.html)>.

<sup>7</sup> <<http://europa.eu.int/languages/>>.

<sup>8</sup> Europeans and languages. A survey in 25 EU Member States, in the accession countries (Bulgaria and Romania), the candidate countries (Croatia

citizens including those of the accession countries (Bulgaria and Romania), of candidate countries (Croatia and Turkey) and the Turkish Cypriot Community. One of the most interesting results is that half of the people interviewed say that they can hold a conversation in a second language apart from their mother tongue.

## 1.2 Why multilingualism is important?

In Europe we want to live in a socially inclusive society in which diverse cultures live in mutual understanding, building at the same time a common European identity.

Language, together with the shared knowledge and traditions, which passed from one generation to another, is an important part of an individual's cultural identity.

We strongly believe that the diversity of languages, traditions and historical experiences enriches us all and fosters our common potential for creativity.

Let us make languages connect people and cultures not divide them. This is an important role for cultural institutions.

Take the case of museums; multilingualism is of significant importance. Museums define their sphere of tasks as collecting, making available, preserving, researching and exhibiting objects. A multilingual exchange of information on objects supports museums in their tasks on the one hand and on the other hand the users of the products of museum work (visitors).

Museums collect objects whose meaning renders them unique and one-of-a-kind. However, the physical objects can only be available in one place at one particular time, making them accessible only to a few people. In order to make information about museum pieces available to as wide a target group as possible, a special importance lies in the accessibility of the relevant information on the Internet and in overcoming language barriers. websites are extremely powerful mean to do that.

Nonetheless, multilingual exchange of information about museum pieces is also of interest for cultural tourism and therefore for economic reasons. A museum visitor wants to know how to access such objects, in other words, which museum is displaying the objects at what point in time. Museums need to be able to make this information available in different languages in order to reach visitors from neighbouring countries.

---

and Turkey) and among the Turkish Cypriot Community,  
<<http://europa.eu.int/languages/en/document/80/20>>.

Multilingualism is of special interest to smaller and local museums in Europe, to preserve local and national differences and to make available their peculiarities and unique characteristics to others.

Objects that originally belonged together have been spread around the world by means of exchange, purchase, division of goods and also by theft or violent conflict. To recreate relationships between the parts of collections that have been dispersed to multiple institutions and countries, it is essential to exchange relevant information and for this to happen multilingual accessibility is a prerequisite.

Further, it can be assumed that many objects can be qualified through a provenance reconstruction that crosses borders. The single objects mutually contextualise one another. And cross-border communication implies the use of multiple languages.

Another point of view is the quality and effectiveness of communication on the Internet. Information technologies dramatically changed users' behaviour at the end of twentieth century and a constant increase in demands and expectations from new services can be observed. Some countries report that the number of virtual visits to cultural institutions is becoming higher than real visits. Therefore each institution should take care about communication on the Internet and the best medium for this is an institutional website. Cultural institutions have become aware of the power of websites and have been creating their own websites since the 1990s. Beyond the problem of guaranteeing a regular maintenance of the information provided, multilingualism plays again a strategic role, The majority of websites are addressed to their own small communities, such as university members, public library readers or the citizens of a town in which a museum is located. However, the more useful information that can be found on a website, the more Internet users visit them regardless of borders. Language is the major barrier to foreigners in making use of these websites.

Whilst policies and initiatives aimed at preserving languages are the prime responsibility of Member States, community action can play a catalytic role at European level adding value to the Member States' efforts.

The development of multilingualism on the Internet has been stimulated in recent years by the European Commission by supporting trans-national projects, fostering partnership between digital content owners and language industries.

However, support for high quality multilingual resources still needs to be enhanced. A pan-European inventory and library of mature linguistic tools, resources and applications as well as qualified centres of competence and excellence would provide helpful support. Online access to this inventory, oriented towards problem-solving, providing cultural institutions with appropriate solutions for specific problems related to linguistic and cultural customisation would be beneficial for the improvement of

multilingualism in the web cultural applications. This Handbook is intended as a contribution to this pan-European inventory.

Europe's experiences in multiculturalism and multilingualism represent an enormous strength that European cultural institutions should be able to exploit by positioning themselves in the new digital sphere of information and knowledge society.

### **1.3 Population languages spoken in the member states**

As we have stated before, the European Union is a multicultural and multilingual community. We have gathered information on population and languages spoken in the member states to introduce this diversity in details. However, we asked for the same set of information from each country, the amount of the information differs depending on the complexity of the situation, and the person who provided the information. Although we tried to make it uniform, it was really hard. Comparing the large countries with the smaller ones there will always be a difference between the number of minorities and immigrants.

We illustrate with this set of information, that multilingualism is an issue in each member state, but it has to be handled differently.

Unfortunately we did not get any information on population and languages spoken from Austria, Belgium, Cyprus, Denmark, Finland, Lithuania, Portugal, and Sweden. This is because of the lack of tools for encouraging the participants for feedbacks. However, we have additional information about our observers: Israel, Norway, and Russian Federation.

The whole texts of the country reports are available on the survey's website: <<http://www.mek.oszk.hu/minerva/survey>>.

#### **Czech Republic**

The number of inhabitants in the Czech Republic is about 10 million. 90.4% of the population is Czech by nationality although many other nationalities are represented, 1% citizens speak Czech, which is the official language of the Czech Republic.

The 90% of the population is Czech, and the other 10% consists of Moravian, Slovakian, Polish, German, Ukrainian, Vietnamese, Hungarian, Russian, Romany/gypsy, Silesian, Bulgarian, Grecian, Serbian, Croatian, Romanian, Albanian minority.

### **Estonia**

Estonia has about 1.351 million inhabitants (as of January 2005). The largest ethnic groups are Estonians (68%), Russians (26%), Ukrainians (2%), Belorussians (1%) and Finns (1%).

Estonian is the only official language in Estonia in local government and state institutions. The Estonian language belongs to the Finno-Ugric language family and is closely related to Finnish. Finnish, English, Russian and German are also widely spoken and understood in Estonia.

### **Finland**

Finland has two official languages: Finnish and Swedish. It is the governmental policy that common public services must be provided in both languages where appropriate. This guideline is followed by most public offices and cultural institutions. The websites reflect this principle although in some cases only a fraction of the content is provided in Swedish. Another indigenous language in Finland is Sami, which is spoken within the small community of Sami people in Lapland (also known as Lapps). There are websites, which offer also material in Sami, both sites linked to their culture and administrative websites.

English is commonly used. Finnish is very different from other larger European languages. This is why English is commonly used in cases where international contacts are judged essential. Commonly only a fraction of the website content is available in English.

### **France**

On the basis of these criteria more than seventy-five languages of France can be counted in Metropolitan France and overseas areas. They are characterized by a great diversity. In Metropolitan France: Romance, Germanic, Celtic languages as well as Basque, a non-Indo-European language. Overseas: Creoles, Amerindian, Polynesian, Bantu (Mayotte) and Austronesian (New Caledonia) languages, among others. There is also a great demographic diversity between these languages. Three or four million people are speaking Arabic in France whereas Neku or Arhà are spoken only by a few dozen people. In between, the various Creoles or the Berber languages are spoken by about two million people in France.

The 1999 national census revealed that 26% of adults living in France had regularly practiced in their youth a language other than French – Alsatian (660 000 speakers), Occitan (610 000), Oïl languages (580 000), Breton (290 000). For each of these languages one can add an equal – at least – number of occasional speakers. However language transmission in France is almost not effective any more in the family circle and relies today mostly on the teaching of these languages and their creativity in the artistic domain.



## **Germany**

82 million people live in the Federal Republic of Germany, which is the most heavily populated nation in European Union. 75 million inhabitants possess German citizenship and about 8 million people hold foreign passports. Approximately 15 million people do not speak German as their native language. The largest population of foreigners are the Turks (1.87 million), followed by Italians (0.62 million), immigrants from the former Yugoslavia (0.56 million), Greeks (0.35 million), Poles (0.32 million), Croatians (0.23 million), Austrians (0.18 million), Bosnians (0.16 million), Americans (0.11 million), Macedonians (0.06 million) and Slovenians (0.02 million).

National minorities, or in other words "groups of German citizens who have traditionally ... resided on the territory of the German Federal Republic and who live in their historic settlement areas", include the Sorbs and Wends (60,000), the Danes (50.000), the Frisians (50.000) and the German Sinti and Roma. In accordance with the European Parliament's European Charter of Regional and Minority Languages of 05 November 1992, they are protected and supported in the context of a "threatened aspect of European cultural heritage". Protection includes the right to use a regional or minority language in the private and public spheres. At the same time, the charter includes the responsibility to facilitate or maintain the use of regional or minority languages. In 1994, a further regional language, Plattdeutsch (Low German), was recognised. According to the Law on Administrative Proceedings [Verwaltungsverfahrensgesetz § 23.1 (VwVfG)], Standard German has been designated as the official written and legal language.

In reality, linguistic and cultural diversity are significantly larger: in 2004, for example, 45.4 million overnight stays of non-German tourists were registered. In December 2004 in Berlin alone, the fourth largest city in Europe, approximately 450.000 foreigners with passports from 185 countries were registered.

## **Greece**

According to the 2001 survey of the National Statistics Agency the population of Greece consists of about 10.934.087 inhabitants. 99% are Greek and the other 1% is divided between about 5 major people groups that possess other citizenships. There are no languages or national minorities currently recognized in Greece. The only officially recognized minority is the religious minority of Greek Muslims in western Thrace.

Greek is the official written/spoken language and the vast majority of the population speaks Greek. However, some very small language groups speak other languages and dialects such as Romanika, Vlachika or Turkish.

## **Hungary**

There are over 10 million inhabitants in Hungary. 97% of the population are Hungarian and the remaining 3% consist of 13 different nationalities: German, Roma, Slovakian, Croatian, Romanian, Ukrainian, Slovenian, Greek, Serbian, Polish, Ruthenian, Bulgarian and Armenian.

The official language, and the majority language is Hungarian, which is part of the Finno-Ugric language family. There are another 5 million Hungarians in the surrounding countries as minorities, and many of them live scattered all over the world. The minorities of Hungary live in small sporadic communities within majority language environments. They are free to use their mother tongue, but due to strong assimilation the usage of minority languages is decreasing in social communications. The minority languages are mainly used in self-government, TV programmes, schools and informal communication. Macedonian, Osetin and Yiddish are also spoken in Hungary, but the numbers of native speakers are very small. There is no education in these minority languages in Hungary.

## **Ireland**

The 2002 Census reported that Ireland has a population of some 3,917,203 inhabitants. There are two official languages: English and Irish. English is the most widely written and spoken language. But 42.8% of the population speaks Irish. The highest proportion of Irish speakers are amongst students in the 10-19 age group and in Galway County in the West of Ireland, which has 52.7% Irish speakers.

## **Israel**

There are over 6.3 million inhabitants in Israel, the majority are Jewish with other religions and languages present. It is a multi-cultural country with various communities living in Israel - both Jewish (stemming from North Africa, Asia, Europe and America) and non-Jewish (Arabs: Moslems, Christians, Bedouins, Druzes).

Most cultural institutions strive to be bilingual in English and Hebrew with some including Russian and Arabic. Russian is supported as there has been a high immigration of Russians to Israel. In 2004, the Israeli government supported 994 cultural institutions and projects. These included 107 museums, 220 libraries, 4 archives and 327 educational facilities.

## **Italy**

Italy has a population of 58,462,375 citizens (recorded on 31 December 2004), which includes 1,990,159 foreigners. Italian is the official language of the Republic, but there are several cultural and linguistic minorities.

Italian legislation (laws n. 482/1999 and n. 38/2001; effective decree of the President of the Republic n. 345/2001) states that the Italian Republic (according to article 6 of the Constitution) values minority languages.

According to the law, the following languages and cultures are preserved and promoted: Albanian, Catalan, Croatian, French, Franco-provençal, Friulian, German, Greek, Ladin, Occitan, Slovene, and Sardinian. (this represents a population of 2,428,770 people). Law 482/1999 decrees, among other things, that these languages and cultures can be taught in schools, that official documents and acts are bilingual, and that the local language can be used for broadcasting information. This law doesn't take into account other languages that are commonly spoken in Italy among immigrant communities, such as Arab or Chinese.

### **Latvia**

In 2004, there were 2,319,203 people in Latvia according to the 2004 Year book published by the Central Statistical Bureau of Latvia. The total number of national minorities is not particularly large in Latvia, and each minority group (except Russians) is relatively small. The biggest and most active communities are Russians, Poles, Lithuanians, Jews and Roma. The majority of people of foreign descent mainly (69.2%) live in the seven major cities of Latvia: Riga, Daugavpils, Jelgava, Jurmala, Liepaja, Ventspils and Rezekne. As in many other countries there are both types of minorities in Latvia – historical, traditional minority and immigrant minority; 16% of all minorities are historical, but 27% are immigrants. 62% of Latvia's residents recognise Latvian as their native language. According the legislation (from 1989) the official language of the Republic of Latvia is Latvian.

### **Luxembourg**

The Grand Duchy of Luxembourg is located in the centre of Europe; it's often called «the heart of Europe». The territory of the country has a surface of 2.586 spkm. The population consists of 277.400 Luxembourg nationals (61,43%) and 174.200 foreigners (38,57%)<sup>9</sup>. The indigenous language is *Lëtzebuergesch* (*Luxembourgish-Frankish-Moselle dialect from the West*) and it coexists among others with High German and French. These three languages have been recognized by law in 1984<sup>10</sup>. At that moment *Lëtzebuergesch* was elevated to the rank of the Luxembourgers' national language. The government also sanctioned its status as a language in its own right, separate from German, and its importance as a symbol of the national identity of the Grand Duchy. 33% of the inhabitants speaks *Lëtzebuergesch*, 58% speaks French, 8% speaks German, and 1% speaks English. Although the status of national language has been granted to *Lëtzebuergesch*, the law stipulated that the language of legislation in the Grand Duchy is French, based on the Code Napoleon, the French civil

---

<sup>9</sup> <STATEC, <<http://www.statistiques.public.lu/stat/TableViewer/tableView.aspx?Repor tId=128>>.

<sup>10</sup> Law of 24 February 1984 related to the organization of languages.

code of law introduced in 1804. In civil and judicial affairs, it is specified that when an application was directed to a public authority in *Lëtzebuergesch*, French or German, the Administration has to respond in the language chosen by the applicant.

### **Malta**

The total population of was 399,867 in 2003. Malta consists of three inhabited islands: Malta, Gozo and Comino and two uninhabited islands, Kemmynet and Filfla. The largest island is Malta, which had a population of just over 388,867 in 2003. Circa 99% of the population are Maltese, and the remaining 1% consists of foreigners working in Malta or a few foreign residents who have retired. Besides the main islands, there are others.

The official languages of Malta are Maltese and English, Maltese being the native language and also the majority language. Other commonly spoken languages in Malta are Italian, French and German, with Italian being by far the most popular amongst these three. In the early 1900's, Italian was the favoured language, especially by the cultured classes and the Maltese aristocracy; more than the English language or the native Maltese tongue.

### **The Netherlands**

The Netherlands has about 16,300,000 inhabitants. There are two official languages: Dutch (*Nederlands*) and Frisian (*Frysk*). Both languages belong to the West Germanic language family. Frisian is spoken by some 400,000 people, mainly in the northern province of Friesland (*Fryslân*), where official/administrative documents are published in both Frisian and Dutch. The Dutch language is also spoken by the Flemish community in Belgium and in the former Dutch colony of Surinam. The total number of people for whom Dutch is the native language is estimated at 22 million. The official organisation for the Dutch language is the *Nederlandse Taalunie* (the Dutch Language Union), in which the governments of Flanders, Surinam and The Netherlands participate.

People of many nationalities live in the Netherlands. In 2004 the city of Amsterdam counted 171 nationalities among its inhabitants. There is almost as much variety of languages spoken, especially in the major cities where most immigrants have settled. The majority of the immigrants come from the Mediterranean (Turkey (357.911) and Morocco (314.699) and from the former Dutch colony of Surinam (328.312; source: *Statistics Netherlands*, <<http://www.cbs.nl>>). In order to improve their opportunities in Dutch society, immigrants are encouraged to learn Dutch, but in spite of this official policy Turkish, Arabic and Tamazight (or Berber) have developed into *de facto* minority languages. In the major cities the municipalities publish much of their information in these languages as well.

### **Norway**

Of Norway's population of 4,606, 363 (on 1.1.2005) 95 per cent speak Norwegian as their native language. Norway has two official written languages, Norwegian and Sámi. But Norwegian is really two different languages Bokmål (Dano-Norwegian) and Nynorsk (New Norwegian). Everyone who speaks Norwegian, whether it is a local dialect or one of the two standard official languages, can be understood by other Norwegians. However, the minority Sámi language is not related to Norwegian and it is incomprehensible to Norwegian speakers who have not learned it.

The two Norwegian languages have equal status, i.e. they are both used in public administration, in schools, churches, and on radio and television. Books, magazines and newspapers are published in both languages. The inhabitants of local communities decide which language is to be used as the language of instruction in the school attended by their children. Officially, the teaching language is called the *hovedmål* (primary language) and the other language the *sidemål* (secondary language). Students read material written in the secondary language and at the upper secondary level they should demonstrate an ability to write in that language. This is a consequence of the requirement for public employees to answer letters in the language preferred by the sender.

### **Poland**

According to recent statistics, Poland is inhabited by 38,230,000 people. About 251,000 (6%) of the population are the members of national and ethnic minorities. Among these the biggest minorities are: German (147.000), Belarusian (47.000), Ukrainian (27.000), Rumanian (12.000), Lemkan (5,800), Lithuanian (5,600), Russian (3,200), Slovak (1,700) and Jewish (1.000). Other, smaller minority groups include Tatar, Czech, and Armenian. In the near future other minorities will probably be identified as a growing number of immigrants from a wide range of countries are applying for Polish citizenship.

The Polish Constitution guarantees minorities members special rights, such as protection and development of their own culture and language, the right to establish educational and cultural institutions and the right to participate in the decision making process, concerning national identity. Children from the biggest minorities may learn their mother tongue language at public schools, situated in the regions settled by those minorities. The most active minorities have established associations, publish newspapers and organize cultural and scientific events. The biggest minorities – Belarusian and German, also have representation in the Polish Parliament.

### **Russian Federation**

The number of the Russian Federation (RF) inhabitants is about 140 million. According to the Russian legislation the official language of the

Russian Federation is the Russian language on the whole territory of the country. At the same time Russia is a federative state and it unites 21 National Republics and 67 regions. All National Republics have got their own Constitutions. Some of them state a national language as an official together with the Russian language with equal rights. So Altai language, Comi language, Saha language, Tyva language, Udmurdt language, Tatar language, Chuvash language and some others are the official languages on the territories of the National Republics of the Russian Federation.

The number of official and minority languages is rather approximate: 130-200 of different language groups. There is no "official linguistic" list of official and minority languages. The official list for disappeared languages "Red book" can be consulted on a good English site <<http://www.eki.ee/books/redbook/>>, yet with the data of 1989, there are 97 languages. The late census data tables are on the site <<http://www.perepis2002.ru/index.html?id=17>>. Unfortunately there is yet no information about native peoples of the Russian Federation but language diversity is well presented - about 150 languages (including 3 foreign ones that are record-holders). and more up-to-date material.

### **Slovak Republic**

Slovakia has relatively high proportion of national minorities in its total population, as to their diversity and number of varieties. Altogether, there are 10 national minorities which constitute about 15% of all citizens. According to the 2001 Census, the largest is the Hungarian minority (9,7%), followed by the Roma minority (1,7%). But in reality, the percentage of Roma people is thought to be as high as 10% of the population. The Czech (0,8%) and other minorities have a representation below 1%: the Ruthenians (0,4%), Ukrainian (0,2%), German (0,1%), Polish, Moravian, Croatian, Russian, Bulgarian and Jewish.

The mixture of languages roughly corresponds to the ethnic composition of the country. The official language of the Slovak Republic is the Slovak, which was first officially codified in 1843.

### **Slovenia**

The official language of Slovenia is Slovene. In the territories where Italian and Hungarian minorities live the Italian and Hungarian languages also have the status of official languages.

There are a number of other minority languages spoken in Slovenia. The major linguistic groups are: Croatian, Serbian, Bosnian and Macedonian.

### **Spain**

Spain has 43.67 million inhabitants (as of 1st January 2005). It is a multilingual country as the result of its cultural diversity. Spanish or Castilian is the official language of the country as recognized in the

Spanish Constitution of 1978. There are other regional languages which are co-official in their Comunidades Autónomas or regions, such as: Galician in Galicia, Catalan in Catalonia and the Balearic Islands, Valencian in the Valencia region and Basque in Navarra and Euskadi. Foreign immigration is a recent phenomenon and, though it implies an impact in multilingualism, the figures are still not very representative. Two million foreigners are recognized by the authorities, in a high percentage coming from Latin America (from Spanish speaking countries).

### **United Kingdom**

English is the most widely spoken language in the UK and it is the *de facto* official language. It is estimated that over 95% of the population of the UK are monolingual English speakers. The UK has several indigenous minority languages, which are protected under the European Charter for Regional or Minority Languages, which entered into force on 1st July 2001. Welsh, Gaelic and Irish are given the highest level of protection under the Charter with Scots, Ulster-Scots, Cornish and British Sign Language also being recognised.

Welsh is spoken by approximately 582,500 people with the number of Welsh speakers increased by 80,000 in the period between 1991 and 2001. In Scotland, Gaelic is spoken by approximately 69,500 people with the highest concentrations of Gaelic speakers living in the Highlands and Islands. In Northern Ireland, Irish is spoken by approximately 106,844 people. Ulster-Scots is spoken by approximately 35,000 people in Northern Ireland.

There are large numbers of other languages spoken in the UK, which have been brought into the country and are sustained by immigrant communities. No single UK body collects information about the numbers of languages that are spoken but some indication is available from local authorities, who translate materials into the languages spoken by inhabitants of their areas communities in their area. The most common languages in which materials are translated include: Bengali, Chinese, Gujerati, Punjabi, Somali, Turkish and Urdu.





## 2 The survey of multilingual websites and thesauri

### 2.1 About the working group on multilingualism

After accession to the European Union the new member states became a part of a multicultural and multilingual community. At present there are 20 official, and about 150 estimated minority and immigrant languages are spoken in the enlarged European Union<sup>11</sup>. The European Cultural Heritage is a common value for the member states. Since distributed search in the different collections is technically possible, it gives also an excellent opportunity for connecting different digital collections, or library catalogues: like The European Library, or The European Digital Library. Since the information and the metadata are registered in different languages, thus information retrieval whether on the Web or in a common database can be a serious problem.

That is why, at the kick off meeting of the MINERVA Plus Project in Budapest February 2004, it was decided to establish a working group specialized on multilingual issues, especially on multilingual thesauri. The working group was a follow up of the work carried out by the working group by the MINERVA Project Work Package 3 (WP3). The new working group was lead by Hungary.

#### Goals and methods

Instead of creating a brand new multilingual thesaurus for the project's purposes, we decided to make a survey of multilingual websites and thesauri. This also gave us a good opportunity to discover the usage of multilingual thesauri all over Europe. The survey was completely voluntary, and we declare that our results cannot be considered to be statistically relevant. They can be best referred to as a random sampling.

---

<sup>11</sup> Calimera Guidelines: Cultural Applications: Local Institutions Mediating Electronic Resources, Multilingualism, 2004, <<http://www.calimera.org/Lists/Guidelines/Multilingualism.htm>>.

The reason for this is explained by the different customs of the member states, different methods of circulating and gathering information implemented by the national representatives and the different social attitudes of each country towards the issue of multilingualism and consequently the different levels of maturity of the digital products in terms of multilingual features.

The coordinators' attitudes, working fields and positions made a major impact on their countries' results. Some countries, including Israel, The Netherlands and Slovakia, had just finished a survey, and were able to contribute these results offline. Other countries, including Poland, Greece and Russia, decided to send offline results because of a shortage of time or resources; these were added to the online results in the same format.

### **The survey's website**

The aim of the survey was mapping multilingual access to the European digital cultural content. To implement the survey we compiled a website <<http://www.mek.oszk.hu/minerva/survey>>, which was used for data collection and displaying the actual results. The online questionnaire could be reached from the front page "Questionnaire". It had two major parts. The first section was for auditing the multilingualism of the cultural websites. The second part could be filled out only by institutions that declared the use of controlled vocabularies for information retrieval in their database. This part was based on an Israeli questionnaire that was developed for a different survey. The results could be continuously followed online. There were separate links from the front page to the "Survey statistics", to the registered "Multilingual websites", and to the "Multilingual thesauri" grouped by the countries.

The statistics were calculated by individual countries, and also for the whole survey. The institutions' types, the number of the languages available on the site, the site availability in English and the type of searching tools were analysed. "Multilingual websites" showed the names of the registered institutions linked to their websites, so that the site could be easily reviewed. "Multilingual thesauri" show the names of the registered thesauri and their registration form.

The country reports and presentations are also available on the website.

### **The first run of the survey**

The first run of the data collection started in June 2004 and ended in August. In the first analysis there were 236 answers from 21 states. This high score indicated also the diversity of participation. From 1 to 40 institutions answered per state and registered their websites in our database. There were 67 libraries, 63 museums, 35 archives, 21 cultural

sites, and 45 other institutions. The results of the first run demonstrated that the 30% of the websites were still monolingual, 43% were bilingual, and about 26% were multilingual. There were 31 thesauri registered: 13 from Italy, 10 from the United Kingdom, 6 from Hungary, 1 from the Netherlands, and 1 from Austria.

The working group had its first meeting on 12<sup>th</sup> of November 2004 in Budapest. The members of the working group presented a short country report. The slides are available on the official website of the survey by clicking on the "*Download the slide shows*". It was clear, that there are different legislation and customs in each member states and so we planned to collect country reports of multilingual aspects. The group agreed on new rules for the survey and restrictions for the results. We started a second run of the survey for those countries that were underrepresented in the first run. We also decided to create a mailing list (WP3 list) for circulating general information and discussion. We set up the criteria for the best practices examples and agreed on definitions.

### **The second run of the survey**

The second run of the survey started in November 2004 and lasted until the end of May 2005. The combined results of the two runs of the survey doubled those of the first. There were 676 websites registered from 24 countries. Some countries, like Germany, Italy, Greece, Israel and Malta sent additional information, but no information came from Cyprus, Latvia, Lithuania or Luxembourg. There were 265 museums, 138 libraries, 98 archives, 65 cultural sites, and 129 other websites registered. 179 of them were monolingual, the majority (310) were bilingual, 123 were available in 4 languages, 14 in 5 languages, 10 in 6 languages, 4 in 7 languages, 3 in 9 languages, and 1 in 34 languages. 491 out of the 676 websites were available in English. There were 106 registered controlled vocabularies in our database: 1 from Austria, 3 from France, 22 from Germany, 6 from Hungary, 30 from Israel, 13 from Italy, 19 from Russia, 1 from Sweden, 1 from The Netherlands and 10 from the United Kingdom.

The second meeting took place in Berlin on the 8<sup>th</sup> of April 2005, during the two day WP5 meeting on quality of the websites. We gained useful experiences. We realised that it would be useful to get to know about the multilingual issues from each country in a sophisticated way and so we decided to collect country reports. This will also help us to find the best practices examples to share. We agreed on the form of the country reports and the deadline for preparing them.

The third meeting took place in Budapest on the 8th of September 2005. The participants of the meeting established an editorial board of this document. We agreed on the timeline, set up the structure of the deliverable and shared the tasks among the group.

## 2.2 Overview of aims, scope and methodology of the survey

The aim of the survey was to get of snapshot of dealing with multilingualism in the different countries: especially on cultural websites, and in online digital collections. The survey was also a campaign for highlighting the importance of the multilingual access on the Internet from the institutional point of view. It was also a good promotion of the whole MINERVA project, and its results, because the institutions participated in the survey were more eager for the different events and documents of the project.

The main objectives were:

- Mapping the multilingual access to the cultural content
- Identifying multilingual thesauri
- Sharing the best practices

The target audience were mainly the websites of different cultural institutions:

- libraries
- archives
- museums
- and other cultural sites

The methodology for the survey was:

- Compiling a questionnaire
- Identifying contact persons from each country
- Creating a website for the online data collection, and for the results, which also serves as a common database

The questionnaire had two major parts. The first section was for auditing the multilingualism of the cultural websites. The second part could be filled out only by institutions that declared the use of controlled vocabularies for information retrieval in their database. This part was based on an Israeli questionnaire that was developed for a different survey. The results could be continuously followed online.

The survey was completely voluntary, and we declare that our results cannot be considered to be statistically relevant. They can be best referred to as a random sampling. The reason for this is explained by the different customs of the member states, different methods of circulating and gathering information implemented by the national representatives and the different attitudes of each country towards the issue of multilingualism and consequently the different levels of maturity of the digital products in terms of multilingual features.

## **2.3 Findings and final results**

The statistics presented here are based both on the answers of the online questionnaire, and the offline summaries.

First we examined the institution type, which maintains the website. Depending on the national representative in some countries - museums or libraries dominated.

Secondly we asked about the languages, which are available on the website. In most of the cases not all the information on the website is translated into the other languages. It can be ranged from 5% to 95% depending on the size and financial potential of the institutions, who maintain the site.

Although we originally examined the languages of the interface, in some cases especially by the digital libraries, they may have different language books, and they reported themselves as a multilingual website.

Then we also wanted to know, how many of the registered websites available also in English. In most of the cases English is the second language of a website.

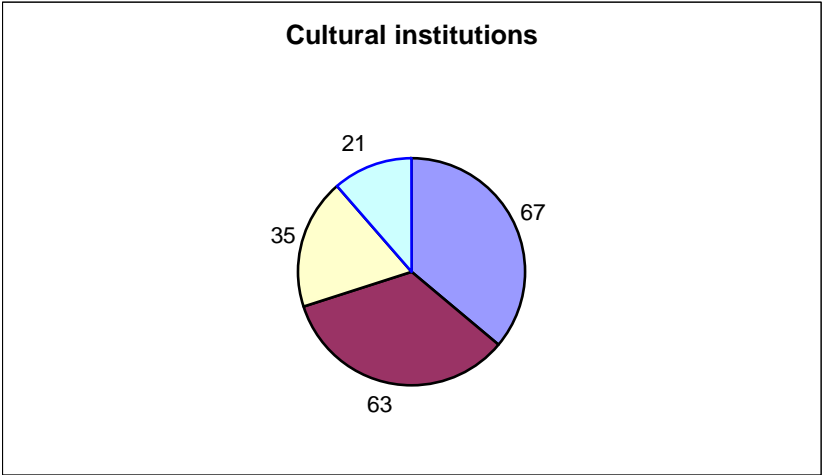
Finally we wanted to learn about the information retrieval tools on the website. In many cases it is enough to have free text indexing, but for digital collections controlled vocabularies can be very useful.

For details about the results in each country, please refer to the survey website.

### **2.3.1 In the first run**

The first run of the data collection started in June 2004 and ended in August 2004. It was a good start, there were 236 registered websites from 21 member states. This high score indicated also the diversity of participation. From 1 to 40 institutions answered per state and registered their websites into our database. Each country registered at least one website like in Ireland, Israel, Norway, but in some countries took it really serious: like from Austria 25, from Slovenia and Hungary about 40 websites were registered. The others ranged in between them. No answer came from Cyprus, Denmark, Malta, Latvia, Lithuania, Luxembourg, Russian Federation that time.

There were 67 libraries, 63 museums, 35 archives, 21 cultural sites, and 45 other institutions.



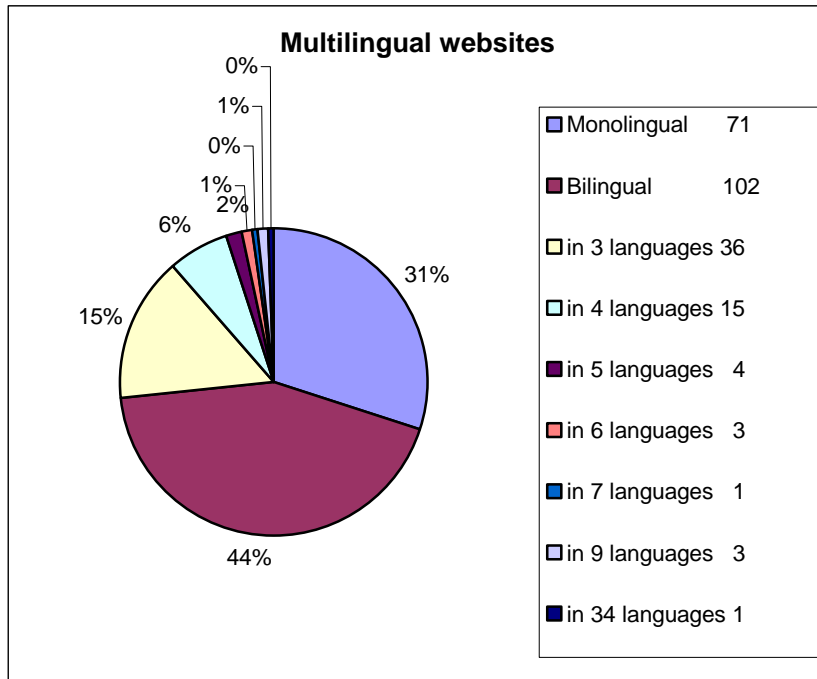
The results of the first run demonstrated that about 30% of the websites were still monolingual, 43% were bilingual, and about 26% were multilingual.

*First findings*

Monolingual	71	30.1%
Bilingual	102	43.2%

*Available*

in 3 languages	36	15.3%
in 4 languages	15	6.4%
in 5 languages	4	1.7%
in 6 languages	3	1.3%
in 7 languages	1	0.4%
in 9 languages	3	1.3%
in 34 languages	1	0.4%



There were 31 thesauri registered:

- 13 from Italy
- 10 from the United Kingdom
- 6 from Hungary
- 1 from the Netherlands
- 1 from Austria.

### 2.3.2 In the second run

The second run of the survey started in November 2004 and lasted until the end of May 2005. The combined results of the two runs of the survey doubled those of the first. There were 657 websites registered from 24 countries. Some countries, like Germany, Italy, Greece, Israel and Malta sent additional information, but no information came from Cyprus, Latvia, Lithuania or Luxembourg. After all Luxembourg sent two multilingual thesauri, and we got a country report from Lithuania.

There were 265 museums, 138 libraries, 98 archives, 65 cultural sites, and 129 other websites registered.

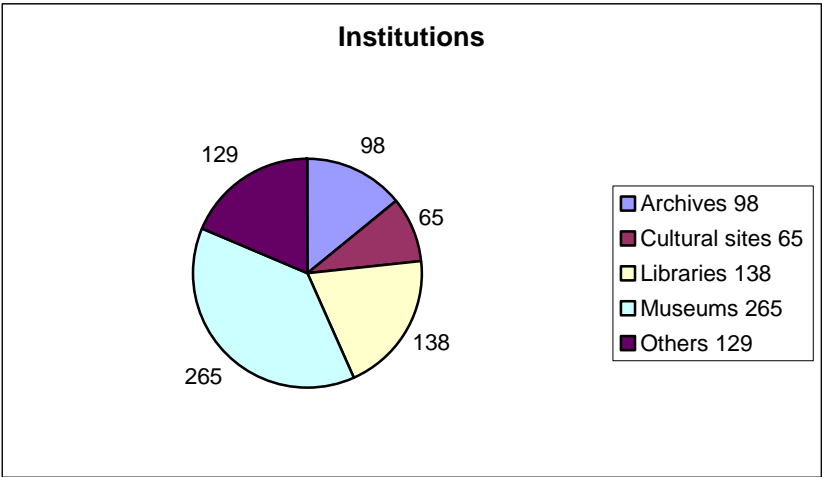
### Types of institutions

<i>Institution type</i>	<i>Number</i>
Archive	85
Archive and cultural site	4
Cultural site	47
Cultural site and other	6
Library	128
Library and archive	1
Library, archive and cultural site	2
Library and cultural site	1
Library and other	1
Museum	248
Museum and archive	4
Museum and cultural_site	3
Museum, cultural_site and other	1
Museum and library	3
Museum, library and archive	1
Museum, library, archive and cultural site	1
Museum and other	4
Other	117
<b>Total</b>	<b>657</b>

Another view of the participants to the second run:

Archive (entirely or partly)	98
Cultural site (entirely or partly)	65
Library (entirely or partly)	138
Museum (entirely or partly)	265
Other (entirely or partly)	129

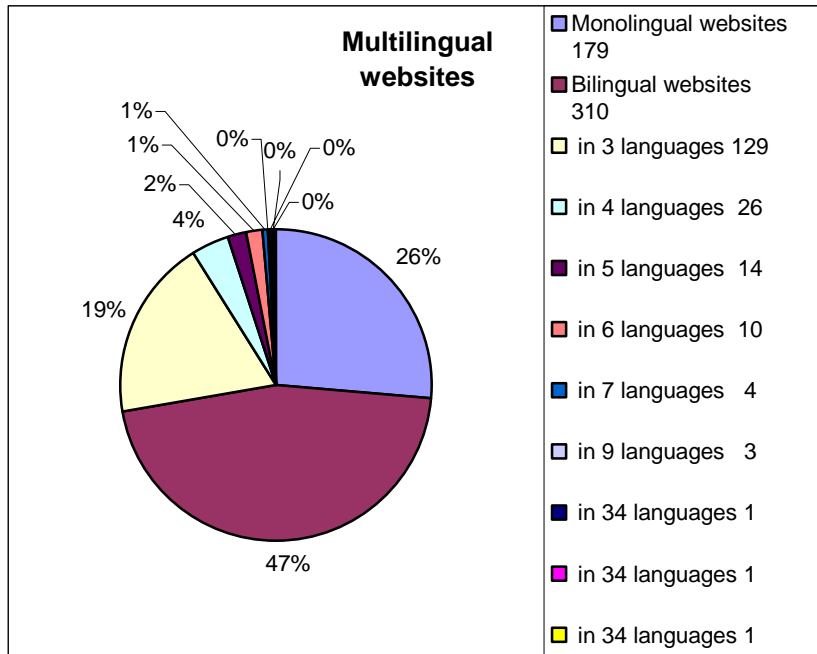




179 of them were monolingual, the majority, 310 were bilingual, 129 were available in 3 languages, 26 were available in 4 languages, 14 in 5 languages, 10 in 6 languages, 4 in 7 languages, 3 in 9 languages, and 1 in 34 languages. 491 websites were available in English.

**Languages available**

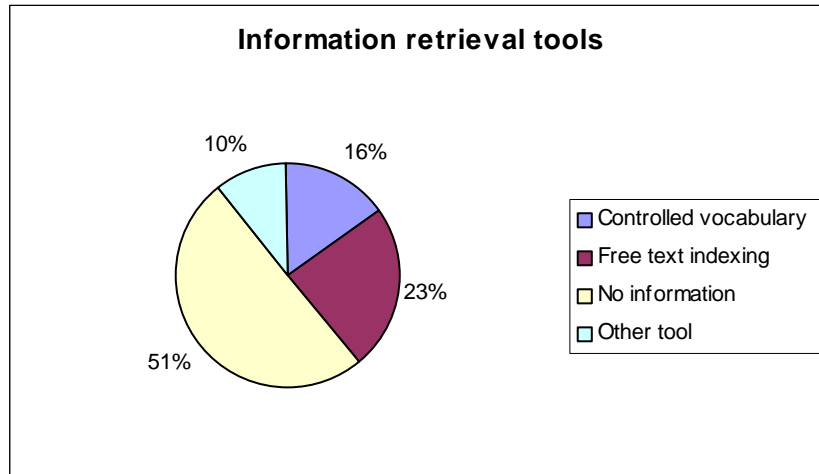
Monolingual websites	179
Bilingual websites	310
<b>Multilingual websites</b>	
- available in <b>3</b> languages	129
- available in <b>4</b> languages	26
- available in <b>5</b> languages	14
- available in <b>6</b> languages	10
- available in <b>7</b> languages	4
- available in <b>9</b> languages	3
- available in <b>34</b> languages	1



We have found, that 26% of the cultural sites are still monolingual, 47% of them bilingual, 27% are multilingual. 74% of them are available in other languages than the original one. There are 491 from 676 websites available in English, which takes 73%. Even if we do not deal with the websites registered from those countries, where English is official language like United Kingdom, Ireland, and Malta, 31 websites, it will be still 460 of them (68%), which are available in English. It means that most of the time the second language of the cultural sites is English.

#### Tools for information retrieval

	Number of institutions
Controlled vocabulary	106
Free text indexing	159
No information	345
Other tool	71



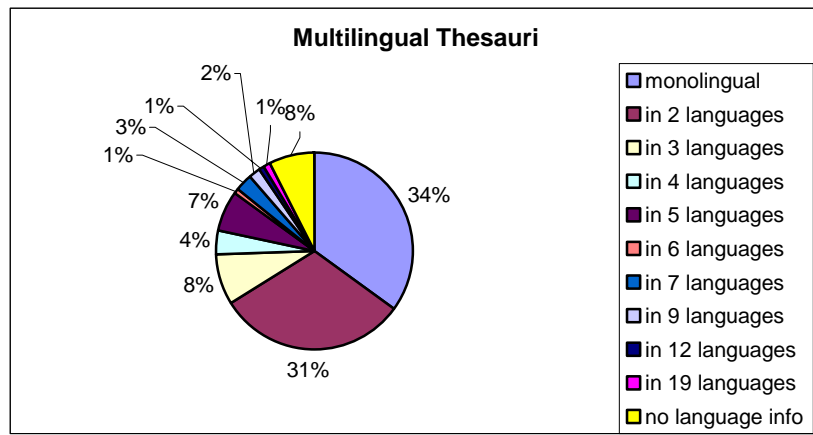
Having a lot of results coming from summaries, we only have information about the half of the websites. Only 16% percent of them use controlled vocabularies for searching their collections. Maybe there was a confusion about using information retrieval tool on the website, or in the database.

#### **Controlled vocabularies**

There are 114 registered controlled vocabularies in our database:

- 1 from Austria
- 11 from France
- 22 from Germany
- 6 from Hungary
- 30 from Israel
- 13 from Italy
- 19 from Russia
- 1 from Sweden
- 1 from The Netherlands
- 10 from the United Kingdom

monolingual	37
in 2 languages	33
in 3 languages	9
in 4 languages	4
in 5 languages	7
in 6 languages	1
in 7 languages	3
in 9 languages	2
in 12 languages	1
in 19 languages	1
no language info	8
sum	106



There were 106 controlled vocabularies registered in our database. 34% of them are monolingual, 31% of them are bilingual, and 23% of them are multilingual. About 8% of them the person, who registered them, forgot to fill out the field about the languages, or it may be the result of other technical problem.

Only 68 are bilingual or multilingual from them, which is 63% of the whole. So we can say, that multilingual thesauri are used by many institutions, and we try to encourage everyone instead of complying one thesaurus, try to find the one, which is suitable for indexing the collections.

The analysis shows, that in Israel many multilingual thesauri used with more than 5 languages. Some of them are in more than 10 languages, which proves us, they can be used quite well in international context.

### 3 Thesauri and controlled vocabularies used in the different countries

#### **Czech Republic**

No multilingual thesauri with cultural coverage were found to be available online among the institutions included in the survey. Relations between terms were mostly done using links or some other hypertext methods. Some of the institutions used free text indexing, but most did not use any sophisticated retrieval tools. The same situation is true of online controlled vocabularies or e-glossaries.

#### *Library of Congress Subject Headings (LCSH)*

Library of Congress Subject Headings (LCSH) are currently used in the Czech Republic as a source of English equivalents of subject terms, but Czech translation does not exist.

#### *UNESCO Thesaurus*

There is no Czech translation of the UNESCO thesaurus yet.

#### **Estonia**

At present there are no multilingual thesauri in use on the Web by any Estonian cultural institution. 15 sites provide free text search.

#### **Finland**

The National Library of Finland maintains two different thesauri, which are both also available in Swedish. The Finnish General Thesaurus is called YSA and the corresponding translated one in Swedish is called Allärs. Finnish Music Thesaurus (MUSA) has also a Swedish translation (CILLA). These thesauri are available on-line and can be searched to find terms and navigate within the thesaurus structure. There are links between the terms of the Finnish and Swedish thesauri. <<http://vesa.lib.helsinki.fi/>>

#### **France**

From the overview of projects we can make out that thesauri are more and more conceived as part of complex systems in which information is

searched through a combination of methods. Sophisticated systems such as SymOntoX<sup>12</sup> allow the management of several ontologies and reduce terminological or conceptual confusion through the definition of a common structure. One of the major challenges is the use of open-source software and open source content.

While the number of multilingual cultural websites is increasing, multilingual controlled vocabularies are still scarce and the works are slow to produce quality and coherence in these vocabularies.

*In the field of architectural and archaeological policies: the HEREIN thesaurus*

The “first multilingual thesaurus in the cultural field at an international level “according to the Council of Europe is now available online<sup>13</sup>. This service is developed by the European Heritage Network (HEREIN). It aims at offering a terminological standard for national policies dealing with architectural and archaeological heritage and at helping the user of the website when surfing through the various online national reports. The Herein thesaurus is made of more than 500 terms in seven languages (English, French, German, Spanish, Bulgarian, Polish and Slovenian) but eleven other languages will soon be available.

*In the field of restoration and conservation of paintings: the NARCISSE vocabulary and the EROS project*

The Scientific Restoration Research Centre for Museums in France (C2RMF) gave the impulse to the European NARCISSE project (Network of Art Research Computer Image SystemS) in the late 1980s. This project aimed at building a multilingual database to manage museum laboratory documentation relating to painting materials.

*In the field of architecture: the Thésaurus de l'architecture<sup>14</sup>*

The Thésaurus de l'architecture is developed by the Direction de l'architecture et du patrimoine (DAPA). It groups together in a methodical way the 1 135 terms used for the denomination of architectural works.

*In the field of religious objects: the Thésaurus des objets religieux (religious objects thesaurus)*

---

<sup>12</sup> SymOntoX is a Symbolic Ontology Management System, XML based, developed at LEKS, Istituto di Analisi dei Sistemi ed Informatica – CNR. It is a prototypal software system based on the OPAL (Object, Process, and Actor Language) methodology for knowledge representation.

<sup>13</sup> <<http://www.european-heritage.net/sdx/herein/thesaurus/introduction.xsp>>.

<sup>14</sup> A description of the Thésaurus de l'architecture is available at: <<http://www.culture.gouv.fr/documentation/thesarch/Othesaurus.htm>>.

*In the field of archaeology and antiquity: the "PACTOLS"thesauri*  
PACTOLS is the acronym for "Peoples and cultures, Anthroponyms, Chronology, Toponyms, Works, Places, Subjects". These thesauri are used by the network and database FRANTIQ which is a cooperative of Research Centres (CNRS, Universities, museums of the Ministry of Culture) and a common network of databases about Sciences of Antiquity from Prehistory to Middle Ages. It is supported by the Department of Humanities and Social Science (SHS) of the National Centre for Scientific Research (CNRS).

*In the field of art works and museum objects: Museum images vocabularies*

Museum Images is a picture library dedicated to the art works and objects of the museums worldwide. Museum Images photo agency delivers to the professionals in the publishing industry, the press, and the communication and advertising industry, digital images of the collections which are part of its catalogue or of any other museum through its picture research service. The vocabulary covers art, architecture, sciences, technology, and history. It is available in five languages (English, German, Italian, French, Spanish).

*In the field of manuscripts and letters: the Malvine thesaurus*

Between 1998 and 2001 the European Malvine project (Manuscripts and Letters Via Integrated Networks in Europe) aimed at building a network of European libraries, archives, documentation centres and museums that keep and catalogue post-medieval manuscripts and letters in order to offer new and enhanced access their collections. The Malvine vocabulary allows semantic interoperability and is available in five languages (German, English, French, Spanish, Portuguese).<sup>15</sup>

*In the field of culture: the Unesco thesaurus*

The Unesco Thesaurus is a controlled and structured list of terms used in subject analysis and retrieval of documents and publications in the fields of education, culture, natural sciences, social and human sciences, communication and information. This trilingual thesaurus contains 7,000 terms in English, 8,600 terms in French and 6,800 in Spanish that are spread between seven major subject domains broken down into micro-thesauri. It is now possible search the online unesdoc / unesbib catalogue directly from the thesaurus. The thesaurus functions are Broader / Narrower Term, Used For, Related term, Scope Note, Descriptor, Non-Descriptor.

*In the field of libraries: the MACS project (Multilingual Access to Subjects)*

The MACS project aims at providing a multilingual access to subjects in the catalogues of the participants. These are Die Deutsche Bibliothek (SchagWortnormDatei), The British Library (Library of Congress Subject

---

<sup>15</sup> The Malvine thesaurus is partly available at: <<http://www.malvine.org:8100/metasearch/thesaurus.jsp?type=thesaurus&lang=fr>>.

Headings), the Bibliothèque nationale de France (Répertoire d'Autorité-Matière Encyclopédique et Alphabétique Unifié), and the Swiss National Library which was in charge of the SWD / RSWK project. 2001.

*In the field of cultural heritage and Euro-Mediterranean tourism: the STRABON thesaurus*

Strabon is a scientific and technical cooperation programme provided for three years (2002-2005) which aims at equipping the Euro-Mediterranean space with a multilingual and multimedia information system that comprises coherent units of digital resources regarding the Euro-Mediterranean cultural heritage and ethical tourism.

## **Germany**

There are three widely available and electronic Authority lists exist for cataloguing in German libraries:

the Schlagwortnormdatei SWD (German Subject Headings Authority)  
the Gemeinsame Körperschaftsdatei GKD (German Corporate Headings Authority)  
the Personennamendatei PND (German Name Authority).

These are designed and maintained by the German Library DDB in cooperation with the different library networks, and are made available online within the framework of the Integriertes Literatur-, Tonträger- und Musikalien-Informations-System (Integrated Literature, Sound Carrier and Music Information System) ILTIS via the Z39.50-Gateway.

<[http://z3950gw.dbf.ddb.de/z3950/zfo\\_get\\_file.cgi?file Name=DDB/searchForm.html](http://z3950gw.dbf.ddb.de/z3950/zfo_get_file.cgi?file Name=DDB/searchForm.html)>.

A few German museums use the German Subject Headings Authority SWD or establish links to it. The German Name Authority PND is being linked to other national Authorities via the Virtual International Authority File (VIAF) <<http://www.oclc.org/research/projects/viaf/default.htm>> to create one international Authority.

The MACS project <<http://laborix.uvt.nl/prj/macs>> has established links between three indexing languages used in national library services: the German Subject Headings Authority (SWD), the Library of Congress Subject Headings (LCSH) and the Répertoire d'autorité-matière encyclopédique et alphabétique unifié (RAMEAU) to facilitate multilingual access to library catalogues. A prototype developed by Index Data and the Tilburg University Library can be viewed at <<http://laborix.uvt.nl/prj/macs/prototyped.html>>.

The DDC-Deutsch project <<http://www.ddc-deutsch.de>> is translating the Dewey Decimal Classification system (DDC 22) into German to develop a tool for online catalogues that enables all titles



classified with DDC to be accessible, particularly Anglo-American data. A series of projects or institutions are in progress. For example, the GVK – Gemeinsamer Verbundkatalog (Common Union Catalogue <<http://gso.gbv.de>>) uses the “Dewey Decimal Classification” link in the “Titeldatenanzeige” to conduct a systematic search via DDC. DDC notations do not exist for all titles, mainly for English language works. But since 2004, the German Library has classified all titles for the National Bibliography DNB according to DDC, it makes these available to the regional libraries for their own use.

A few libraries offer classifications that are partially translated on their websites, such as the Göttingen Online Classification GOK <<http://db1-www.sub.uni-goettingen.de/gok/index-e.html>> or the originally Dutch “Basis-Klassifikation” <<http://sbbweb1.sbb.spk-berlin.de:8080/DB=1/LNG=EN/BCL>> along with PICA, which is made available by the State Library of Berlin SBB and mainly used in Lower Saxony and Saxony Anhalt.

In the past, many different and individual solutions were created for researching single projects, especially in the museum and archive areas. These are mainly monolingual, sometimes only available offline and are often not visible to the ordinary web user.

Some museums use and maintain common Authorities by sharing data, with a rather large number of descriptors. In this context, for example, the mainly German language-based “Seitendateien”(Helpfiles) of Foto Marburg <<http://www.fotomarburg.de>> are implemented in cooperation with the MIDAS-Rules. The “Geo-Seitendatei”(Geo- side files) administers Polish and German geographic terms.

A larger number of art museums use the ICONCLASS notations for iconographic description, enabling multilingual access via the Internet if the correct technical, financial and legal prerequisites are in place. (Iconclass in German <<http://194.171.152.226/libertas/ic?style=index.xsl&taal=de>>

To date, only a few German museums use the Getty vocabularies: Thesaurus of Geographic Names TGN, the Union List of Artist Names ULAN or the Art & Architecture Thesaurus AAT.

## **Greece**

Only a small percentage of the websites evaluated in the survey were found to use a thesaurus or taxonomies for thematic indexing. In 2001, Tsafou and Hatzimari reported that libraries in Greece made very little use of thesauri for the following reasons:

- limited strength of their collections

- use of software developed outside Greece that does not support non-Latin alphabets
- the absence of a national coordinating institution to undertake the development of suitable information processing tools.

Further reasons are the lack of development of appropriate thesauri for the cultural domain or of standardized translations of such resources.

Many websites (43.5%) provide a controlled vocabulary, either mono- or multilingual, as a means of describing and searching the available resources. Most of these vocabularies are proprietary, i.e. they are suitable for the documentation needs of each particular site and are not standardized. Translation into Greek of the International Standard for Archival Description (ISAD) by the Society of Greek Archivists <<http://www.eae.org.gr>> is an example of an effort towards standardization. It has been used for the on-line collection of the Hellenic Literary and Historical Archive <<http://www.elia.org.gr>>.

Out of the sites that provide some means of structuring information (ranging from vocabulary to thesaurus) 63.3% maintain bilingual versions that become available when the interface language is selected by the user.

Only 3% of the websites included in the survey support and maintain multilingual thesauri. These are based almost solely on translations of well-known international standards and classification systems. The most prominent ones in use in Greece seem to be:

LCSH. Translated versions of LCSH are used by the majority of Greek libraries. It is not always the case that there is concurrent multilingual use of LCSH, but bilingual examples include the Library of the Technological Educational Institute of Thessaloniki <<http://www.lib.teithe.gr>> (LCSH version 27) and the on-line catalogue of the Eugenides Foundation <<http://www.eugenfound.edu.gr>>.

SEARS: Translated versions exist but multilingual use of SEARS in Greece is rare. The Library of the Technological Educational Institute of Lamia <<http://www.lib.teilam.gr>> employs a bilingual version of SEARS for thematic indexing.

NLG-LCSH: The National library of Greece (NLG, <<http://www.nlg.gr>> used LCSH as the basis for developing a customized translation in Greek. NLG maintains this and makes it available to other libraries and institutions which are then able to adjust it according to their needs. The Public Central Library of Serres <<http://www.serrelib.gr/>> uses a monolingual version of NLG-LCSH blended with SEARS headings.

Although support for multilingual thematic indexing was found to be limited a twofold momentum towards overcoming this can be recognized: Cultural institutions and organizations show both awareness and willingness to make their collections accessible to non-native speakers.

There are the on-going efforts to offer some choice and guidance in multilingual description of digital cultural resources. Institutions seem steadily adapting to the multilingual challenge as a growing number enable multilingual access to their collections. This situation can only benefit from a tighter coordination at national and international level.

## **Hungary**

Information retrieval tools were reported on 21 websites. Controlled vocabularies are being used for searching databases via 6 websites. Two of these are monolingual (OSZK Thesaurus, WebKat Thesaurus), another two are bilingual (Library of Congress Subject Headings List, Thesaurus of Library Information Science), they are available both in English and in Hungarian. The Hungarian Ecoinfo Thesaurus has also English and German versions. The Hungarian Educational Thesaurus is available in French, in English, and in German.

There are 59 thesauri available in Hungarian, but only about 35 have ever been used.<sup>16</sup>

Multilingual thesauri are the following: Thesaurus of Energetics, Ecoinfo - Economical Thesaurus, Educational Thesaurus. The UNESCO International Thesaurus of Cultural Development is available in Hungarian, but it has never been used<sup>17</sup>.

Bilingual thesauri are the following: Geological Thesaurus, Thesaurus of Library Information Science, and the Library of Congress Subject Headings List. There is only one thesaurus for museums, but it has never been used.

## **Ireland**

Although there is variety of controlled vocabularies and thesauri available to the English speaking community, the survey carried out for Ireland could not list any specific document/programme for the purpose of multilingual web sites.

---

<sup>16</sup> Ungváry Rudolf: A tezauszokról: <[http://www.oszk.hu/hun/szakmai/tezausz/tezausz\\_oszk\\_hu.htm](http://www.oszk.hu/hun/szakmai/tezausz/tezausz_oszk_hu.htm)>.

<sup>17</sup> A kulturális fejlődés nemzetközi tezauszja: információkereső tezausz / [összeáll. Jean Viet; ford. és bev. Dienes Gedeon] Budapest: Művelődéskutató Intézet, 1980.

## **Israel**

As a result of the survey, 30 institutional lexicons were identified and reviewed.

These included:

- 9 archives (6 bi/multilingual)
- 8 libraries (7 bi/multilingual)
- 5 museums (4 bi/multilingual)
- 8 educational facilities (5 bi/multilingual).

17 of the 30 lexicons reviewed in the survey were available on line.

Another recent survey among the Israeli heritage community showed that institutions are using a wide variety of vocabularies while indexing and documenting. But these are internal tools and are not directly visible to the end user. These lists are shared by more than one institution:

BARCAT - Bar-Ilan Library Catalog of Bar Ilan University digital subject listing in Hebrew and English. This work is based on a translation and adaptation of Library of Congress Subject Headings

Israel Antiquities Authority Lexicon an archaeological classification system for research and the documentation of findings. "Truly Bilingual". Hebrew and English.

IMAGINE Thesaurus developed and used by the Israel Museum, Jerusalem, an encyclopaedic museum, with standards garnered from the VRA and the AAT, focused mainly on Jewish material culture. It is constructed of "legacy terms" and is multidisciplinary in its nature. The Israel Museum has benefited from the Israel Antiquities Authority lexicon, and has continued to work on the basis of their lists for certain archaeological tables. The Israel Museum inaugurated the first multilingual bi-directional museum collections database; supporting fully both Hebrew and English. The Image Search Engine of the Israel Museum, Jerusalem (IMAGINE) was installed in June of 2004 and is used by curators, restorers, and the registrar's office. A nationwide project is in the works to share the IMAGINE thesaurus with the 54 museums of Israel supported by the Department of Museums of the Ministry of Education.

## **Italy**

The Central Institute for Catalogue and Documentation (ICCD) of the Italian Ministry for Cultural Heritage and Activities produces several mono- or multilingual controlled vocabularies for cataloguing purposes.

They represent national standards for all cultural institutions (national, local or private) involved into the cataloguing of the cultural heritage. The domains covered are: architecture, art-history, archaeological objects and sites, artistic objects, architectural areas.

The ICCD presents 8 controlled vocabularies related to description of cultural areas, authors, artistic technique and artistic objects. Artistic objects (one of the most used) is available in Italian, English, German, French and Portuguese (with specific sections in other languages). The architectural areas vocabulary is available in Italian, English and French. All these vocabularies are available upon request.

Another important tool for multilingual classification for the iconography of western art, ICONCLASS, is available in Italian, English, German, French, and Finnish <<http://www.iconclass.nl>>. The ICONCLASS vocabulary is free to use; the complete software is commercially priced.

In cooperation with the Canadian Heritage Information Network (CHIN), the Getty Information Institute, and the French Ministry of Culture, ICCD has also produced the Multilingual Thesaurus of Religious Objects, which is available in English, French, and Italian. It is available on CD-ROM <[http://www.iccd.beniculturali.it/servizi/testo\\_cd5.html](http://www.iccd.beniculturali.it/servizi/testo_cd5.html)>.

ThIST (Italian Thesaurus of Earth Sciences), available in Italian and English, is maintained by the library of the national Agency for Environmental Protection and Technical Services (APAT); it covers the earth science domain and can be browsed on-line <[opac.apat.it](http://opac.apat.it)>. This thesaurus complies with ISO 2788/1986 and is developed in cooperation with an international experts working group.

An Italian to English iconographic thesaurus, is maintained by Alinari in cooperation with the University of Florence. It contains about 8,000 entries organised in 61 classes alphabetically ordered (from Agriculture to Zoology).

The system includes a geographic thesaurus, thesauri for Periods and Styles, controlled lists for Events, People, Authors (artists) and Photographers. The Alinari thesaurus is a work in progress. It has been translated into Spanish, German, and French for the European project Orpheus. The thesaurus can be purchased for use.

A working group on the semantic Web, made up of experts of various fields (universities, W3C consortium, libraries, private companies), has developed an Italian to English glossary about e-learning, available on line at the URL <<http://www.bdp.it/websemantico/>>.

The Multilingual Thesaurus of Religious Objects, the controlled vocabulary for artistic objects, and the translation into Italian of the ICONCLASS classification, all produced by the ICCD.

## **Latvia**

The survey found that:

Museums in Latvia use local developed classification schemes in Latvian and the Art & Architecture Thesaurus (AAT) in English.

Archives use the UKCAT thesaurus in English.

Libraries use four principal vocabulary tools:

UDC classification in English (this is being translated into Latvian)

MeSH in English and Latvian (part translation)

LCSH is used as the basis for developing a partly adapted translation in Latvian

AGROVOC in English.

## **Luxembourg**

Actually, no multilingual thesauri dealing with cultural domains are available online.

The initiative of National Museum of Natural History is developing interrelationships between terms from different languages (Latin, English, French, German and Luxembourgish) in collaboration with the linguistic section of the Grand-ducal Institute. This is an ongoing project, which uses the thesaurus module in the biological recording application called *Recorder*.

<<http://www.mnhn.lu/recherche/db/recorder/collmod.asp>>.

In 2003, the Public Research Institute "Gabriel Lippmann" (CRPGL) created the first computer-based spellchecker for the Luxembourgish language. In collaboration with the Permanent Council for the Luxembourgish Language (CPLL), which was responsible for the linguistic part of the project, the computer engineers at the CRPGL developed multi-platform software. This software, a plug-in that can be used in several office-suites, includes a sequential list of 130.000 Luxembourgish words and their grammatical forms. They have been extracted from the large electronic text corpus "LuxTexte".

This corpus covers a large variety of text types: contemporary literature, written and oral media, political speeches... Over 80% of the linguistic items are lemmatised and grammatically categorised. "LuxTexte" is constantly extended towards becoming a reliable representative update of the current use of the Luxembourgish language.

<<http://www.lippmann.lu/cortina>>

## **The Netherlands**

A recent study among the Dutch heritage community showed that institutions use a wide variety of controlled vocabularies while indexing

and documenting internally, but these tools are not visible to the end user of the websites.

Most search tools for the public are either based on full text searches or on query by form. Vocabulary aids are limited and mainly offer support in the form of a list of available indexing terms. Fourteen sites in the survey group (some 27%) offer controlled vocabulary/thesaurus support to the end user.

The most important vocabulary tools accessible on line are:

- AAT-NL: a translation in Dutch of the Art & Architecture Thesaurus of the Getty Institute, maintained by the Rijksbureau Kunsthistorische Documentatie/ Netherlands Institute for Art History, which is becoming a standard vocabulary in Dutch (and Flemish) museums. When the technical development is ready, a bilingual thesaurus will be available as an indexing and search aid (cf. <http://www.aat-ned.nl/index.html>).
- Ethnographical thesaurus: developed and used by the Dutch ethnological museums as an extension of the AAT, which is focused mainly on Western material culture (cf. <http://www.svcn.nl>).
- RKDartists: a standardised list of about 200.000 names and details of artists, maintained by the Rijksbureau Kunsthistorische Documentatie/ Netherlands Institute for Art History, which will also become a standard vocabulary for the Dutch museum community (cf. <http://www.rkd.nl/rkddb/default.asp?database=ChoiceArtists&action=form>).
- Iconclass: an international classification system for iconographic research and the documentation of images (cf. <http://www.iconclass.nl/>)

A more comprehensive list of the available tools is under construction (cf. <http://www.den.nl/Leidraad/AccessDatabs/Terminologiebrn.pdf>).

Vocabulary support for the non-Dutch speaking end user is very rare. Sites of many institutions offer search pages and some support in English, but except for the major and internationally renowned institutions (like the Royal Library, the International Institute of Social History, the Rijksmuseum) in most cases the end user will have to enter search terms in Dutch. Truly multilingual functionality is not yet offered by the first three tools mentioned above. Only Iconclass has a proven track record of multilingual access.

## **Poland**

A majority of cultural institutions websites in Poland do not offer any search mechanism. Information can be selected from the menu. Just nine institutions were found to offer an advanced information retrieval mechanism. Among them there were 6 libraries and 3 museums. They

offer free text search (5), Google browser search (3) and controlled vocabulary (1).

The 6 Research Libraries were:

- Wrocław University Library  
<<http://www.bu.uni.wroc.pl>>  
searching in English – Google browser
- The Ossoliński National Institute  
<<http://www.oss.wroc.pl>>  
searching in English – Google browser
- Poznań University of Technology – Main Library  
<<http://www.ml.put.poznan.pl>>  
searching in English – Google browser
- The Central Library of the University of Gdańsk  
<<http://www.bg.univ.gda.pl>>  
searching in English – free text
- University Library in Toruń  
<<http://www.bu.uni.torun.pl>>  
searching in English – free text
- Technical University of Lodz – Main Library  
<<http://www.bg.p.lodz.pl>>  
searching in English – free text

The 3 Museums were:

- Memorial and Museum Auschwitz – Birkenau in Oświęcim  
<<http://www.auschwitz.org.pl>>  
searching of the Death Books in English and German – controlled vocabulary
- The Museum of Kurpiowska Culture  
<<http://www.muzeum-ostroleka.art.pl>>  
searching in English – free text
- Wawel Royal Castle  
<<http://www.wawel.krakow.pl>>  
searching in English – free text

## **Russian Federation**

The survey of Russian Federation websites found that most search tools are links, query by form or full text searches. Vocabulary support is rare and mostly in the form of indexing terms (3 museums – over 2%).

As to the problem of controlled vocabulary, there is no Russian standard museum thesaurus or ontology that has been officially adopted or agreed by the Russian museum community. Museum terminology is concentrated in the most popular museum information systems and adjusted in the process of adaptation of the system for individual museum needs. In Russia there are two museum information systems installed in more than 100 museums, these are CAMIS (developed by AltSoft, Saint-Petersburg,



<<http://www.altsoft.spb.ru>> and "AIS Museum"(developed by the Main Computing Centre, the Ministry of Culture and Mass Communications). Each system has a set of controlled vocabularies, but these are only available in Russian. The Ministry of Culture and Mass Communications project "United Museum Catalogue"has declared that it will develop a standard museum thesaurus but this activity has not started yet.

Some Russian museums use vocabularies for indexing and documenting internally:

Classifications on materials, technique, ethnicity and topical belonging (in Russian) have been developed by the Russian State Museum of Ethnography, Saint-Petersburg; these vocabularies are also presented as an independent resource on the website <<http://www.ethnomuseum.ru>>; the same Russian classifications on materials and technique are also used in the State Historical Museum, Moscow Polytechnic vocabularies (in Russian) developing by the State Polytechnic Museum <<http://www.polymus.ru>>, these are not directly visible for the end user

The iconography thesaurus by F. Garnier (in Russian, French, English) – a Russian version of the descriptive standard vocabulary (controlled by the Ministry for Culture of France) has been developed in the State Historical Museum, Moscow.

AAT (in Russian, English): a Russian translation of part of the Art & Architecture Thesaurus of the Getty Institute (materials, technique, periods) is being developed in the State Historical Museum, Moscow.

The State Historical Museum, Moscow is working on relating terms on materials and technique in two vocabularies (the classifications of the Russian State Museum of Ethnography and AAT) in their original languages.

No multilingual thesauri with cultural coverage are published online with the relations between the terms clearly visible. The iconography thesaurus by F. Garnier (in Russian, French and English) is a multilingual controlled vocabulary available via the museum local network in the State Historical Museum.

### **Slovak Republic**

At present there are no multilingual thesauri in use on the Web by any Slovak cultural institutions. It is worth noting that the library sector uses the Universal Decimal Classification and monolingual subject headings extensively. Support for MARC 21 enables use of controlled vocabulary or thesauri in the future. Museums and galleries use their own monolingual lists of descriptors.

## **Slovenia**

All of the bilingual and multilingual websites of the cultural institutions that took part in the survey were reviewed in order to identify bilingual or multilingual lexicons and thesauri.

No bilingual or multilingual lexicon or thesaurus was found in the desktop research. In most cases the information retrieval is supported by free text indexing. Bigger databases are normally searchable only in the Slovene language although all other information on the website is bilingual or multilingual.

## **United Kingdom**

The cultural institutions that took part in the MINERVA survey also reported on the use of controlled vocabularies and information retrieval tools in their websites. These were as follows: five websites used controlled vocabularies, six used free-text indexing, seven used no vocabulary tool while one site was reported to use another tool (neither a controlled vocabulary nor free text indexing).

The vocabulary tools that were registered include:

ARENA periods - a simple vocabulary list in English, Danish, Norwegian, Icelandic, Polish and Romanian. This list is unpublished but is made available on request free of charge by the Archaeology Data Service.

ARENA top level themes – a simple vocabulary list covering the cultural heritage and sites and monuments and available in English, Danish, Norwegian, Icelandic, Polish and Romanian. This thesaurus is unpublished but is made available on request free of charge by the Archaeology Data Service.

Culturenet Cymru bilingual Welsh-English subject index – a glossary or terminology list of 1000–5000 terms relating to the cultural heritage in Wales. This list is unpublished but is made available on request free of charge by Culturenet Cymru.

Monolingual thesauri and terminology lists were registered by English Heritage, the Tate and by the Scottish Library and Information Council.

Other terminology resources exist in the UK but were not registered in the UK survey. For example, the Tate has developed glossary definitions in British Sign Language <<http://www.tate.org.uk/collections/glossary/bsl-list.jsp>> and it also offers PDA-based gallery tours in BSL.

## 4 Best practice examples

### 4.1 Best practice examples for multilingual thesauri

Creating a multilingual thesaurus can be really expensive, and highly complicated due to the semantic problems between different languages, and also it takes a long time. That is why we have decided to collect information on thesauri used by different cultural institutions all over Europe.

During the survey there were more than 100 thesauri registered by the participating countries of the MINERVA Plus project. The registration was voluntary, so of course not all the controlled vocabularies are registered in our database, which are available. We were looking for thesauri, which are currently used by cultural institutions, and may be convenient for online implementation: information retrieval in digital collections.

We present you some of them in details, which are available in more than two languages, and have already been used in many European countries. With this collection of thesauri we would like to encourage the European cultural institutions after they decided to use a thesaurus for subject indexing, consider of choosing a well-tried multilingual one. It can be very useful for example by combining different collections, which is an emerging trend in all over the world. Time to time more international joint catalogues, and digital collections are being created with multilingual interfaces, and cross-language search facilities, for example The European Library, and The European Digital Library.

#### **The UNESCO thesaurus** <<http://www.ulcc.ac.uk/unesco/>>

The UNESCO Thesaurus was created in 1977 by the United Nations Educational, Scientific and Cultural Organization (UNESCO). Its purpose was to act as the main working tool of the UNESCO Computerized Documentation System (CDS) and allow indexing and information retrieval in the UNESCO Bibliographic Database (UNESBIB) and other sub-databases that are part of the UNESCO Integrated Documentation Network.

The UNESCO Thesaurus is a controlled and structured list of terms used in subject analysis and retrieval of documents and publications in the fields of education, culture, natural sciences, social and human sciences, communication and information, politics, law and economics and countries and country groupings. This trilingual thesaurus contains 7,000 terms in English, 8,600 terms in French and 6,800 in Spanish that are spread between seven major subject domains broken down into microthesauri. There is a yearly increase of about 20 terms.

The first 1977 edition was in English only. French and Spanish translations became available in 1983 and 1984. The version now in use is the second printed edition published in 1995 – with some amendments. The thesaurus is enriched and updated regularly. For the second printed edition the frequency of occurrence of each descriptor in document indexing in the UNESBIB database was measured, in order to choose the descriptors. In case of doubt the last version of the OECD (Organisation for Economic Co-operation and Development) multilingual Macrothesaurus (html version: <http://info.uibk.ac.at/info/oecd-macroth/>) was systematically referred to. More specialized thesauri were also consulted in order to ensure better terminological compatibility with the international controlled vocabularies. The current CD-Rom version (UNESBIB Bibliographic database – UNESCO Thesaurus CD-Rom, 2004) is the 12<sup>th</sup> edition.

The structure of the Thesaurus follows the ISO 2788 and ISO 5964 standards. The thesaurus functions supported are: Broader / Narrower Term, Use / Used For, Related term, Scope Note. The thesaurus is available on the UNESCO Databases CD-Rom and through Internet <http://databases.unesco.org/thesaurus/>. A paper version is available as well. It is made of four parts: alphabetical structured and permuted list sorted by English terms, with their French and Spanish equivalents; hierarchical list by microthesaurus; French/English/Spanish index of descriptors; Spanish/English/French index of descriptors.

Users of the Thesaurus are the institutions in Member States, United Nations System and other intergovernmental organizations, international non-governmental organizations, experts and consultants, UNESCO staff and visitors to the Organization. The Thesaurus can also be used for subject indexing by libraries, archives, documentation centres. For instance, the monolingual UK Archival Thesaurus (UKAT) and UK National Digital Archive of Datasets (NDAD) have taken the UNESCO thesaurus as their starting point.

A part of the UNESCO thesaurus may be used in the future for the French catalogue of cultural digital collections within the framework of the Michael project.

The website and web interface for the UNESCO Thesaurus are maintained by the University of London Computer Centre (ULCC). It is now possible to search the online unesdoc / unesbib catalogue directly from the Thesaurus. Requests for permission to use Thesaurus data have to be directed to the UNESCO library ([library@unesco.org](mailto:library@unesco.org)). A copy of the thesaurus can be obtained for a small fee: 23 € for the CD-Rom in 2005. The softwares used are Winisis, BASIS, and wwwisis (web version). More

information is available at <<http://www.ulcc.ac.uk/unesco/>>. The contact person for the UNESCO Thesaurus is Meron Ewketu at the UNESCO Library (email: [m.ewketu@unesco.org](mailto:m.ewketu@unesco.org); phone: + 33 1 45 68 19 34/35; fax: + 33 1 45 68 56 17/98).

The UNESCO Thesaurus is a controlled and structured list of terms used in subject analysis and retrieval of documents and publications in the fields of education, culture, natural sciences, social and human sciences, communication and information. This trilingual thesaurus contains 7,000 terms in English, 8,600 terms in French and 6,800 in Spanish that are spread between seven major subject domains broken down into micro-thesauri. It is now possible search the online unesdoc / unesbib catalogue directly from the thesaurus. The thesaurus functions are Broader / Narrower Term, Used For, Related term, Scope Note, Descriptor, Non-Descriptor.

A part of the UNESCO thesaurus may be used in the future for the French catalogue of cultural digital collections within the framework of the MICHAEL Project.

In Russian, UNESCO thesaurus is used. The multilingual thesaurus attached to the HEREIN project intends to offer a terminological standard for national policies dealing with architectural and archaeological heritage, as defined in the Convention of Granada (October 1985) and Valletta (January 1992). At first, it will be conceived in English, Spanish and French; it will subsequently be possible to extend the thesaurus to other languages. This tool is intended to help the user of the website when surfing through the various on-line national reports. Thanks to its standardized vocabulary (ISO 5964 standard: *Guidelines for establishment and development of multilingual thesauri*) and to the scope notes appended to each term - which form source material - the multilingual thesaurus gives access, with one concept, to different national experiences or policies whose specific designation, administrative structure, and development provide a view over the wide-ranging extent of European cultural diversity. Besides which, the thesaurus offers the user a terminological tool which allows them to have a better understanding of all the concepts they come across when reading the reports; thanks to the hierarchical and associative interplay of terms, the users can complete or extend their knowledge of the subject. Partners: Cyprus, France, Hungary, Lithuania, Poland, Romania, Slovenia, Spain, Switzerland, United-Kingdom.

#### **Library of Congress Subject Headings (LCSH)**

The Library of Congress Subject Headings (LCSH) is a thesaurus from which the subject indices of documents (books, articles etc) are selected. It is an accumulation of the headings established at the US Library of Congress since 1898. It currently contains over 220.000 terms and its organization is based on the ISO-2788 standard.

### **The MACS project (Multilingual Access to Subjects)**

The MACS project aims at providing a multilingual access to subjects in the catalogues of the participants. These are Die Deutsche Bibliothek (Schlagwortnormdatei), The British Library (Library of Congress Subject Headings), the Bibliothèque Nationale de France (Répertoire d'Autorité-Matière Encyclopédique et Alphabétique Unifié), and the Swiss National Library which was in charge of the SWD / RSWK project. No language is used as a source language in the MACS project. Each indexing language is autonomous but linked to the others by concept clusters. The RAMEAU language has been developed since 1980 in an autonomous way from the Quebec Laval university "Répertoire de vedettes-matières"(Laval RVM) that is itself a translation of the Library of Congress Subject Headings. Some English and French equivalents therefore already exist and this allows a search of some French library catalogues with the LCSH (Service Universitaire de Documentation, Lyons Local Library, ...) but this is not the case with the German language. In the MACS project the terms of the three lists (LCSH, Rameau, SWD) are analysed in order to determine whether they are exact or inexact linguistic equivalents. A MACS prototype is being developed by Index Data (Denemark) and Tilburg University Library (Netherlands) that uses the Link management Interface (LMI). This project is likely to be used in the TEL project (The European Library), which started in 2001.

### **"Library of Congress Subject Headings"**

In France, on the model of the LCSH, the Rameau language has a structure in three levels which makes its richness but also its complexity:

- at the terminological level (= terms selected, called headings + excluded or rejected terms), Rameau is a controlled language, in particular as for the form of the vocabulary, with synonymy and the homonymy: the objective is to arrive to a homogeneous and univocal language (where 1 heading = 1 concept and 1 concept = 1 heading ), while multiplying the access points under the terms retained starting from excluded terms;
- at the semantic level (= relations between generic terms, specific and associated), Rameau is a language arranged hierarchically with the manner of a thesaurus: the objective is to allow a navigation between the terms selected in order to widen (generic terms), to refine (narrower terms) or to reorientate (associated terms) research;
- at the syntactic level (=headings+ subdivisions), Rameau is a precoordinated language obeying precise rules of construction: the objective is to allow, beside research by words, one Library of Congress Subject Headings",

In Germany, through the Multilingual access to subjects MACS project <<http://laborix.uvt.nl/prj/macs>>, links have been established between three indexing languages used in three different national library services (the German Subject Headings Authority SWD, the Library of Congress Subject Headings LCSH and the Répertoire d'autorité-matière

encyclopédique et alphabétique unifié RAMEAU) in order to facilitate multilingual access to library catalogues. A prototype developed by Index Data and the Tilburg University Library can be viewed under <http://laborix.uvt.nl/prj/macs/prototyped.html>

In Greece, LCSH <http://www.loc.gov/cds/lcsh.html>: There exist custom translated versions of LCSH which are used by the majority of Greek libraries that provide access to their items information on-line. Concurrent multilingual use of LCSH is not always the case; however some bilingual examples include the Library of the Technological Educational Institute of Thessaloniki.

In Hungary, LCSH The Library of Congress Subject Headings [http://webpac.lib.unideb.hu/corvina/nagy/term\\_search](http://webpac.lib.unideb.hu/corvina/nagy/term_search) are used by the University and National Library, University of Debrecen <http://www.lib.unideb.hu/>. It is permanently developed. There are more than 10001 terms have been translated yet.

In Israel, The BARCAT - Bar-Ilan Library Catalog [http:// library.os.biu.ac.il](http://library.os.biu.ac.il). Bar Ilan University digital subject listing in Hebrew and English. This work is based on a translation and adaptation of Library of Congress Subject Headings (LCSH).

In Poland, the information retrieval at the majority of Polish on-line catalogues and the two central catalogues includes the Library of Congress Subject Headings (LCSH) system. LCSH has been translated from English through French RAMEAU so in theory it should be possible to search those catalogues in three languages. Since nineties we can observe a growing number of on-line catalogues available. These can be found on library websites. Among them the most important two central catalogues are available: NUKAT <http://www.nukat.edu.pl/> and KARO Distributed Catalogue of Polish Libraries <http://www.nukat.edu.pl/>. In addition 10 library on-line catalogues with interface in English are accessible.

In Latvia , the Library of Congress Subject Headings (LCSH) is used also.

#### **The HEREIN thesaurus**

<http://www.europeanheritage.net/sdx/herein/thesaurus/introduction.xsp>

The "first multilingual thesaurus in the cultural field at an international level "according to the Council of Europe is now available online<sup>18</sup>. This service is developed by the European Heritage Network (HEREIN). It aims at offering a terminological standard for national policies dealing with

---

<sup>18</sup> The HEREIN thesaurus is available at: <http://www.european-heritage.net/sdx/herein/thesaurus/introduction.xsp>

architectural and archaeological heritage and at helping the user of the website when surfing through the various online national reports. Users of the Thesaurus include authorities, professionals, researchers, training specialists. A French scientific committee was set up in October 2005 in order to further define how to make French heritage policies available on the HEREIN database.

The Herein thesaurus is made of more than 500 terms in seven languages (English, French, German, Spanish, Bulgarian, Polish and Slovenian) but eleven other languages will soon be available. It was constructed from scratch and based on the use of the equivalence, hierarchical and associative relationships. The ISO 2788 thesaurus standard was followed as well as ISO 5964 except that no source language was chosen.

The three teams (from Spain, France and the UK) which constructed the thesaurus first created each a separate list of terms and then compared them. They first brought out the different classes representing the broadest level and sorted the terms into the classes. Then within each class the terms were ordered following the same hierarchical relationship for all linguistic versions of the thesaurus. Poly-hierarchy was avoided as much as possible.

When entering a query with the help of the thesaurus one can choose to specify what kind of relationships one wants to include: broader / narrower terms, related terms, preferred / non preferred terms, linguistic equivalents (exact / inexact). The thesaurus is downloadable on Internet.

The contact persons for the HEREIN thesaurus at the Cultural Heritage Division of the Council of Europe are Christian Meyer ([christian.meyer@coe.int](mailto:christian.meyer@coe.int)) and Laetitia Hamm ([laetitia.hamm@coe.int](mailto:laetitia.hamm@coe.int)).

The contributors are: in Bulgaria, the Ministry of Culture, the National Institute for Monuments of Culture, the Bulgarian National Committee of ICOMOS; in Cyprus, the Ministry of Interior, the Department of Town Planning and Housing; in France, the Direction de l'Architecture et du Patrimoine (Department of Heritage and Architecture) of the French Ministry of Culture and Communication (contact person: Orane Proisy; email: [orane.proisy@culture.gouv.fr](mailto:orane.proisy@culture.gouv.fr)); in Hungary, the Kulturális Örökségvédelmi Hivatal (National Office of Cultural Heritage); in Lithuania, the Academy of Cultural Heritage; in Poland, the Ministerstwo Kultury, Department for the Protection of Historical Monuments; in Romania, the CIMEC - Institutul de Memorie Culturala; in Slovenia, the Ministry of Culture, the National Institute for the protection of Cultural Heritage; in Spain, the Ministerio de Educación Cultura y Deporte, Subdirección General de Protección del Patrimonio Histórico, the Consejo Superior de Investigaciones Científicas - Centre for Scientific Information and Documentation; in Switzerland, the Federal Office of Culture; in the United Kingdom, the English Heritage



The first multilingual thesaurus in the cultural field at an international level "according to the Council of Europe is now available online"<sup>19</sup>. This service is developed by the European Heritage Network (HEREIN). It aims at offering a terminological standard for national policies dealing with architectural and archaeological heritage and at helping the user of the website when surfing through the various online national reports. A French scientific committee was put in place in October 2005 in order to further define how to make French heritage policies available on the HEREIN database. The Herein thesaurus is made of more than 500 terms in seven languages (English, French, German, Spanish, Bulgarian, Polish and Slovenian) but eleven other languages will soon be available. It was constructed from scratch and based on the use of the equivalence, hierarchical and associative relationships. No source language was chosen. The three teams (from Spain, France and the UK), which constructed the thesaurus first created each a separate list of terms and then compared them. They first brought out the different classes representing the broadest level and sorted the terms into the classes. Then within each class the terms were ordered following the same hierarchical relationship for all linguistic versions of the thesaurus. Poly-hierarchy was avoided as much as possible. When entering a query with the help of the thesaurus we can choose to specify what kind of relationships one wants to include: broader / narrower terms, related terms, preferred / non preferred terms, linguistic equivalents (exact / inexact).

#### **The NARCISSE vocabulary and the EROS project**

The Scientific Restoration Research Centre for Museums in France (C2RMF) gave the impulse to the European NARCISSE project (Network of Art Research Computer Image SystemS) in the late 1980s. This project aimed at building a multilingual database to manage museum laboratory documentation relating to painting materials. A multilingual controlled vocabulary proved necessary to describe the works of art, the technical data relating to the photographic archives, the restoration and study reports. It was elaborated in German, Italian, Portuguese and French from the beginning and voluntarily restricted to 300 words. From 2001 onwards the NARCISSE vocabulary was used and updated within the framework of the EROS (European Research System) project, which was launched in collaboration with the Mission for Research and Technology of the French Ministry of Culture. Currently, over 300,000 photographic and radiographic images, 10,000 technical reports, 500 3D objects, 200,000 quantitative analyses related to 56,000 works of art are accessible online in digital form on the EROS database.

The database allows research about the works depending on their fabrication technique, the materials used, their ageing process?. The

---

<sup>19</sup> The HEREIN thesaurus is available at: <<http://www.european-heritage.net/dx/herein/thesaurus/introduction.xsp>>.

EROS system uses open source softwares that use the web technologies and respect the new interoperability and content management standards. It relies on advanced content recognition techniques. It uses at the same time multilingual value lists (NARCISSE vocabulary), an SQL search engine operating on metadata tables and free text, a search engine operating on multilingual indexes extracted from full text with an English-French interface (Pertimm), a graphic 3D interface with query according to an ontological model (Sculpteur software<sup>20</sup>), a semiautomatic clustering classification system (RETIN) and image similarity research based on a vectorial tool. The NARCISSE vocabulary is now translated in German, English, Catalan, Chinese, Danish, Spanish, French, Italian, Japanese, Portuguese and Russian. It is organized as a set of dictionaries for each translatable field and for each available language. Some of them are hierarchical. In order to get a quicker answer when searching the data within the main database is stored in a compact language independent format as short codes. The system is able to handle multiple entries within a single field. The thesaurus cannot only handle a full lexical hierarchy but also synonyms and complex character sets such as Japanese and Chinese (via Unicode encoding). The EROS database has been entirely translated from the French language into English, Japanese, Chinese and partially in Portuguese. In due course the system will be set up in the French network of restoration workshops. Some controlled vocabularies used in France and available on line were developed as European projects. This is the case of the HEREIN thesaurus in the area of architectural and archaeological heritage policy and of the Malvine thesaurus which is used for searching the IMEC (Institut Mémoire de l'Édition Contemporaine) database in France in the field of manuscripts and letters. The EROS and NARCISSE databases about restoration and conservation are based on a multilingual controlled vocabulary but the EROS database is not yet available online and only a part of the NARCISSE database is available online – in French only.

#### **In the field of iconographic description: ICONCLASS**

[<http://www.iconclass.nl/>](http://www.iconclass.nl/)

A larger number of art museums uses the ICONCLASS notations for iconographic description, enabling multilingual access via the Internet if the correct technical, financial and legal prerequisites are in place

[<http://www.iconclass.nl/ICONCLASS>](http://www.iconclass.nl/ICONCLASS)

is a specific international classification that the museums can employ for iconographic research and the documentation of images contains definitions of objects, people, events, situations and ideas abstract which can be the subject of an image.

Comprise a system of classification (approximately 28 000 definitions), an alphabetical index, as well as a bibliography of 40 000 references to

---

<sup>20</sup> A description of the Sculpteur project is available at: [<http://www.sculpteurweb.org/>](http://www.sculpteurweb.org/).

books and articles in the fields of the iconography and the cultural history. ICONCLASS is available at the present time only in English but is in the course of translation in French and other languages.

A larger number of German art museums uses the ICONCLASS notations for iconographic description, enabling multilingual access via the Internet if the correct technical, financial and legal prerequisites are in place.

Iconclass in German <<http://194.171.152.226/libertas/ic?style=index.xsl&taal=de>>.

A good example of an Iconclass implementation is the site on medieval illuminated manuscripts of Museum Meermanno and the Royal Library.

<<http://www.kb.nl/kb/manuscripts/browser/index.html>>.

### **Garnier's Thesaurus Iconographique**

It is basically a development of the ICONCLASS system where the notation has been simplified. Only broad classes have notations, so such notation is limited to four or five digits. A practical approach is taken, not to enumerate every sort of variation within a scene, but to provide a string of keywords, which will facilitate retrieval of documents or images. The iconographical analysis is not as deep as that of ICONCLASS, but this is probably an advantage in a retrieval tool not intended as a document surrogate." Steven Blake Shubert: *Classification in the CHIN Humanities Databases*, 1995. *Thesaurus iconographique: système descriptif des représentations / François Garnier*. - Paris: Léopard d'or, c1984. - 239 s.: ill. ; 30 cm. ISBN: 2-86377-032-2

## **4.2 Best practice examples for multilingual websites**

Internet users form a huge multilingual community, and they can visit as many places virtually as they want to. The only problem could be, when they find a website, which is referred relevant to their search, but they don't speak the language of the site. This is a good reason for institutions to provide information in different languages on their websites, to gain more virtual visitors.

During the survey 657 multilingual websites were registered from 24 countries. We asked the national representatives to nominate some of them as a best practice example, to encourage cultural institutions to translate their websites to different languages.

For information retrieval on most of the websites free text indexing is used, but there are sites, which provide thesaurus for searching the content. There are advantages and disadvantages of both tools - as we presented in the chapter 1.4, so we introduce them in two separate sections.

#### **4.2.1 Best practice examples of multilingual websites with thesaurus**

##### **Czech Republic**

*The Museum of Decorative Arts in Prague*

<<http://www.upm.cz/>>

Description: This website is available in 2 languages, it provides a search tree as a search facility (Czech, English)

##### **France**

*Val-de Loire – patrimoine mondial*

<<http://www.valdeloire.org/>>

Description: Ease of switching between languages

*The Grandidier collection of Chinese ceramics*

(catalogue de la collection Grandidier de céramiques chinoises)

on the website of the Museum of Asian Arts

<<http://www.museeguimet.fr/>>

Description: The bilingual treatment of a controlled vocabulary: The Museum of Asian Arts – Guimet uses a French-Chinese controlled vocabulary in the areas of humanities and art history, more specifically about Asian art and fire arts. This vocabulary comprises a value list, a classification, an index and a glossary and is made of 1,000 to 5,000 terms.

*Unifrance*

<<http://www.unifrance.org/>>

Description: The volume and the level of the vocabulary processed: the Unifrance website allows a search in its database about cinema through a number of lists of terms which add up to more than 10 000 terms in four languages while the website of the City of Carcassonne offers a terminological analysis in three languages of the technical terms that are used.

##### **Germany**

*Virtual Library for Anthropology EVIFA*

<<http://www.evifa.de/>>

Description: Online resources are accessed via a search mask and browsing structure for topics (using a thesaurus provided by the International Bibliography of Anthropology IBA) and sources in English and German

*International Architecture Database - archINFORM*

<<http://www.archinform.net>>

Description: In addition to the static context of the website and the navigation, dual language tools in English and German are made available for retrieval purposes. Personal names can be located alphabetically, and subject headings and geographic terminology can also be searched in a hierarchic order. Recording of further foreign language terminology – including languages from outside the European Union – is already partially realized. A few terms can also be acoustically selected in German, English, French and Italian.

## **Greece**

*Myriobiblos, the Digital Library of the Church of Greece*

<<http://www.myriobiblos.gr>>

Description: Makes its content available using a bilingual Greek-English vocabulary.

## **Hungary**

*The Fine arts in Hungary*

<<http://www.hung-art.hu/index-e.html>>

Description: Its cultural content is professional, and it can be searched by different aspects in both languages -English and Hungarian.

## **Israel**

*The Central Database of Shoah Victims' Names of the Yad Vashem Archives*

<[http://www.yadvashem.org/wps/portal/IY\\_HON\\_Welcome](http://www.yadvashem.org/wps/portal/IY_HON_Welcome)>

Description: The Page of Testimony registry uses a thesaurus, is bi-directional and truly multilingual. The advanced searches query the following fields: Names, Places, Date, Submitter, and Family Members. With over 10 languages equated to two main searchable languages: Hebrew and English.

*IMAGINE The Image Search Engine of the Israel Museum, Jerusalem*

<<http://www.imj.org.il/Imagine/Search.asp>>

Description: The thesaurus contains over 50,000 edited bilingual terms. At present the lexicon is available in Hebrew and English. A trilingual (Hebrew, Arabic, English) searchable hierarchal database exists online in the image filled "Living Together Project" <<http://www.imj.org.il/youthwing/livingtogether/searchEng.asp>>.

*Hadashot Arkheologiyot – Excavations and Surveys in Israel online publication by Israel Antiquities Authority*

<[http://www.hadashot-esi.org.il/search\\_eng.asp](http://www.hadashot-esi.org.il/search_eng.asp)>

Description: On line journal – Excavations and Surveys in Israel (HA-ESI). The journal contains preliminary reports of excavations and surveys in Israel, as well as final reports of small-scale excavations and surveys; it also publishes archaeological finds recorded during inspection activities. The journal is bilingual, Hebrew and English; reports submitted in English are translated into Hebrew and vice versa.

*The Ketubbot Collection of the Jewish National & University Library (JNUL)*

<<http://jnul.huji.ac.il/dl/ketubbot/>>

Description: Many projects fall under the auspices of the Jewish National and University Library (JNUL) but only The Ketubbot collection uses a lexicon within its database. The lexicon interfacing is only in English but Hebrew terms can be searched as well. The collection can be accessed by a country list using, a “Graphic List” or a “Textual List”. In addition an Aleph search engine can be used to query various parameters.

## **Italy**

*Library Claudia Augusta of the provincial administration of Bolzano, Trentino-Alto Adige region*

<<http://www.bpi.claudiaugusta.it/>>

Description: The Bolzano province is bilingual Italian-German and this website is organized in 3 sections, Italian, English, and German; the catalogue is only in Italian and German).

*On-line Sardinian dictionary*

<<http://www.ditzionariu.org/home.asp?lang=sar>>

Description: Translation from Sardinian to Italian, French, English, German, and Spanish.

## **Luxembourg**

The initiative of National Museum of Natural History is developing interrelationships between terms from different languages (Latin, English, French, German and Luxemburgish) in collaboration with the linguistic section of the Grand-ducal Institute. This is an ongoing project, which uses the thesaurus module in the biological recording application called *Recorder*.

<<http://www.mmhn.lu/recherche/db/recorder/collmod.asp>>

## Malta

### *Malta Tourism Authority's Website*

<<http://www.visitmalta.com>>

Description: available in 9 language interfaces: English, French, Italian, German, Spanish, Russian, Dutch, Chinese and Japanese. Search results in English.

## Netherlands

### *Medieval Illuminated Manuscripts of Museum Meermanno and the Royal Library*

<<http://www.kb.nl/kb/manuscripts/browser/index.html>>

Description: This website is a good example of an Iconclass implementation (French, German, English).

### *The Anne Frank Museum (or Achterhuis)*

<<http://www.annefrank.org/>>

Description: has a site with complete language versions in Dutch, English, German, French, Spanish and Italian. Searches can be performed using Google, an a-z list of topics and a list of categories in all languages.

### *The Archive of the Province of Fryslân*

<<http://www.tresoar.nl/>>

Description: offers a full version in frysk, the regional language

## Poland

### *University Library in Wrocław*

<<http://www.bu.uni.wroc.pl>>

Description: 90% available in English and German with both searching and the on-line catalogue in the two languages; like the Manuscriptorium, yet the data is only in Polish? German?

### *Technical University of Lodz – Main Library*

<<http://www.bg.p.lodz.pl>>

Description: 70% available in English with both searching and the on-line catalogue available in English.

### *Auschwitz-Birkenau Museum in Oświęcim*

<<http://www.auschwitz-birkenau.oswiecim.pl>>

Description: 100% available in English and German with searching in both languages.

## Russian Federation

### *State Hermitage website*

<<http://www.hermitage.ru>>

Description: presents cultural content including a digital collection and provides excellent search facilities for the content including QBIC search - an image content search that lets to find works of art by their visual details. The site content is available in more than one language etc. The State Hermitage website was designed and developed with the help of IBM.

### *The portal "Museums in Tatarstan"*

<<http://www.tatar.museum.ru>>

Description: has Tatar, Russian and English versions and is oriented to various user communities, including the Tatar Diaspora abroad. The portal has its singularity: audio fragments (texts, Tatar poetry and music) in the Tatar, Russian and English versions of the portal.

## Slovenia

### *COBISS.SI (Co-operative Online Bibliographic System & Services)*

<<http://www.cobiss.si/>>

Description: it is a shared bibliographic database (union catalogue) created by 280 participating libraries and is developed and maintained by the Institute for Information Science Maribor. It is a network application that allows libraries and end users online access to the bibliographic databases in the COBISS system as well as to various specialised databases (of local and foreign database providers) on local servers or remote Z39.50 servers. Of the three user interfaces (Telnet, Windows and Web), the most popular is the web interface. It is fully bilingual in Slovene and English.

### *The Moderna Galerija (Gallery of the Contemporary Art)*

<<http://www.mg-lj.si/>>

Description: houses the national collection of 20th century Slovene art (paintings, sculptures, prints and drawings as well as photography, video and electronic media collections), a collection of works from the former Yugoslavia, and the international collection Arteast 2000+. The national collection presents the basic stages in the development of the Slovene tradition of modern and contemporary art from the beginning of the 20th century onwards. The web presentation of the Gallery is attractive and well organized. It is fully bilingual including the virtual collection and the database on artists, their education, bibliography, awards and exhibitions.



## **United Kingdom**

### *Gathering the Jewels*

<<http://www.gtj.org.uk/>>

Description: The full contents of this website and the underlying database are bilingual in Welsh and English

### *Multikulti*

<<http://www.multikulti.org.uk/>>

Description: This is an online information service that provides advice, guidance and learning materials in 13 community languages. The full contents of the site are available in each language. The website itself has been developed using Unicode to support non-Latin scripts but advises users that there may be some difficulty in viewing certain language texts, particularly Bengali, Farsi and Gujarati and, for these languages PDFs are delivered as well as Unicode text.

## **4.2.2 Best practices of multilingual websites with free text indexing**

## **Czech Republic**

### *Museum of Puppets in Chrudim*

<<http://www.puppets.cz/>>

Description: This website is available in 6 languages, although it does not provide sophisticated search facilities, (Czech, English, German, French, Dutch, Italian).

## **Estonia**

### *Estonian National Museum*

<<http://www.erm.ee>>

Description: The contents of the site are available in Estonian, English, Finnish and Russian (SSEARCH ONLY IN ESTONIAN, web the same for Estonian and English, less content for other languages).

## **France**

### *Musée des Augustins (Toulouse)*

<<http://www.augustins.org>>

Description: Quality and depth found of multilingual treatment (French, English, Spanish)

*The Collection of Great Archaeological Sites (Collection des Grands Sites Archéologiques)*

Published by the Mission for Research and Technology of the French Ministry of Culture

<<http://www.culture.gouv.fr/culture/arcnat/fr/>>

Description: The availability in at least three languages: the websites from the collection of great archaeological sites (Collection des Grands Sites Archéologiques) published by the Mission for Research and Technology of the French Ministry of Culture, about the Chauvet cave

<<http://www.culture.gouv.fr/culture/arcnat/chauvet/fr/>>

Spanish, English, French, the Man of Tautavel <<http://www.tautavel.culture.gouv.fr>> Spanish, English, French) and Life along the Danube <<http://www.culture.gouv.fr/culture/arcnat/harsova/fr>> English French and Rumanian.

*The City of Carcassonne*

<<http://www.carcassonne.culture.fr/>>

Description: The volume and the level of the vocabulary processed: the Unifrance website allows a search in its database about cinema through lists of terms which add up to more than 10 000 terms in four languages while the website of the City of Carcassonne offers a terminological analysis in three languages of the technical terms that are used.

*Underwater Archaeology (from the collection of great archaeological sites published by the Mission for Research and Technology of the French Ministry of Culture)*

<<http://www.archeologie-sous-marine.culture.fr/>>

Description: The processing of non-European languages: the website devoted to Underwater Archaeology (from the collection of great archaeological sites published by the Mission for Research and Technology of the French Ministry of Culture) is available in Arabic

## **Germany**

*Virtual Library of Contemporary Art ViFaArt - makes available ArtGuide, a catalogue of annotated Internet sites.*

<<http://vifaart.slub-dresden.de>>

Description: The site offers German and English language systematic for geographic regions, time and "source" types, as well as alphabetical subject headings for content documentation and linguistic labeling in English and German

## **Greece**

*Benaki Museum*

<<http://www.benaki.gr>>

Description: Makes its collections available using a bilingual Greek-English vocabulary.

*Museum of Cycladic Art*

<<http://www.cycladic-m.gr>>

Description: Makes its collections available using a bilingual Greek-English vocabulary.

## **Hungary**

*The Hungarian Museum of Ethnography*

<<http://www.neprajz.hu/english/index2.html>>

Description: A spectacular cultural site, which provides information in 3 languages, and offers virtual exhibitions with high-resolution pictures. (English, Hungarian, German)

*Embroidered Egg collection*

<<http://datan-datenanalyse.de/Tojas/index.html>>

Description: The information provided is quite limited, because of the size of the museum. Information is available in 8 languages.

## **Israel**

*The Knesset*

<<http://www.knesset.gov.il>>

Description: The Archives of the Parliament of Israel can be searched in Arabic, Hebrew, English. Although completely trilingual, the website allows different search capabilities for each language

*Ghetto Fighters' House: Holocaust and Jewish Resistance Heritage Museum*

<<http://gfh.org.il/eng/>>

Description: The museum has a multilingual website Hebrew, English, French, Russian, Arabic, searching of the archives in Hebrew and English.

## **Italy**

*Superintendence of Venice*

<<http://www.soprintendenzave.beniculturali.it>>

Description: the website is available in 8 European languages; a searchable database is available only in Italian for the photo archives.

*Ladin Cultural Institute*

<<http://www.istladin.net/web/default.asp>>

Description: the web site is available also in Italian, German, English.

*Civic network of South Tirol*

<[http://www.provinz.bz.it/index\\_i.asp](http://www.provinz.bz.it/index_i.asp)>

Description: the website is available in Ladin, Italian, German, French.

*Slovene research Institute of Trieste*

<<http://www.lscmt.univ.trieste.it/slorig/Homepage.htm>>

Description: the website is available in Slovene, Italian, and English.

*Region Valle d'Aosta*

<<http://www.regione.vda.it/default/i.asp>>

Description: the official website is in Italian and partially in French.

## **Luxembourg**

The National Museum of Natural History (MNHN, <<http://www.mnhn.lu>>) is the national portal for our natural heritage (Luxnat). The Museum currently links names of plants and animals of species living in Luxembourg to their biological record collection in 5 languages with "RECORDER - software". The work is still in progress and the aim is to put it on-line. The National Museum of Natural History participates in the European project related to a Biodiversity Collection Access Service for Europe Biocase <<http://www.biocase.org>>. The aim of this European research project is to make biological collections accessible to research and MNHN is considered to be the national hub <<http://www.mnhn.lu/recherche/db/recorder/collmod.asp>>.

The National Museum of History and Art (MNHA, <<http://www.mnha.lu>>) works closely together with National Museum of Natural History. The aim is to use "RECORDER" for the purposes of the different archaeological sections. Also both museums participate in several national programmes; a most interesting one is called "Luxembourg tomorrow" and among the strands is multilingualism. International collaboration is a very important issue for MNHA as well as for other cultural bodies.

## **Netherlands**

*The Royal Library*

<<http://www.kb.nl/index-en.htm>>

Description: English and Dutch website. Site offers search pages and some support in English.

*The International Institute of Social History*

<<http://www.iisg.nl>>

Description: Site of institution offers search pages and some support also in English

*The Rijksmuseum*

<<http://www.rijksmuseum.nl>>

Description: Site in Dutch and English with visitor information additionally in German, French and Spanish.

**Norway**

*Bazar*

<<http://www.bazar.deichman.no/>>

Description: it is a website for language minorities in Norway and is available in 14 languages and is a unique possibility to reach language minorities in their own language on their own premises. Bazar is developed and run by the Multilingual Library with funding from ABM-utvikling.

*Vadsø museum has a multilingual website*

<<http://museumsnett.no/alias/HJEMMESIDE/vadsomuseet/>>

Description: about the museum, local history and the Kvens. The text is in Norwegian, English and Finnish/Kven.

*Kulturnett Troms*

<<http://troms.kulturnett.no/samegillii/>>

Description: is part of Kulturnet.no (the "website for culture in Norway"), run by ABM-utvikling on behalf of the Ministry for Culture and Church Affairs. Kulturnett Troms – is multilingual sami and Norwegian.

*Sami radio, run by the Norwegian non-commercial broadcasting company, has a multilingual website*

<<http://www.samiradio.org/>>

Description: in North-sami, Lule-sami, South-sami and Norwegian.

*The Sami parliament runs a website*

<<http://www.samediggi.no>>

Description: with information about sami politics and government, but also information to the citizens from health to culture. It is in Norwegian and sami.

**Poland**

*The Malbork Castle Museum*

<<http://www.zamek-malbork.pl>>

Description: 100% available in English and German with searching in both languages.

*The State Archive in Siedlce*

<<http://www.archiwumpanstwowe.siedlce.com/index.html>>

Description: 80% available in English and French.

*The State Archive in Plock*

<<http://www.archiwum.plock.com>>

Description: 80% available in English and Russian.

*BWA Gallery in Bydgoszcz*

<<http://www.bwa.bydgoszcz.com>>

Description: 100% available in English and German.

*Katarzyna Napiórkowska Art Gallery*

<<http://www.galeriakn.home.pl>>

Description: 90% available in English and 70% available in German and French.

## **Slovenia**

*Narodna galerija (National Gallery)*

<<http://www.ng-slo.si/>>

Description: is the main art museum in Slovenia containing the largest visual arts collection from the late medieval period to the early twentieth century. The information on collections, exhibitions and events is bilingual in Slovene and English and in some cases also German. There are two databases (Art in Slovenia, European Paintings) containing digital images of paintings and sculptures as well as the description of artefacts available on the National Gallery web pages. The search interface and the descriptions are available in Slovene language only.

*City museum of Ljubljana*

<<http://www.mm-lj.si/>>

Description: is a comprehensive museum storing the material evidence of human existence in the area of the Ljubljana (Slovene capital) of the last five millennia. The museum keeps several hundred thousand artefacts which testify to the history of the city and the people who lived and worked there. The web presentation of the museum matches almost all quality principles criteria. It is fully bilingual in Slovene and English including the small database of the museums digital collection, called virtual room.

*The Architecture Museum of Ljubljana*

<<http://www.arhmuz.com/>>

Description: is the central Slovenian museum for architecture, physical planning, industrial and graphic design and photography. The museum collects, stores, studies and presents material from these areas of creativity at temporary and permanent exhibitions. The museum covers the entire history of these activities from the first human presence in the area of present-day Slovenia. The museums web presentation is attractive and fully bilingual. No databases of digitised content are available.

## **United Kingdom**

### *Milestones Museum*

<<http://www.milestones-museum.com>>

Description: This website is fully accessible to BSL (British Sign Language) users. BSL versions of the text are made available using video clips with captions to allow BSL users to absorb the information about the museum's collections on the website.





## 5 Conclusions

After the recent enlargement of the European Union in 2004 we became a part of a huge multicultural community of 25 countries. To take an advantage of the union of Europe, joint work between member states is the most important. The number of European projects is increasing and more and more cooperation should be attempted. To achieve an efficient collaboration, we should learn more about each other's culture, tradition, and regulations. This will take time, but it is useful to understand the different customs, otherwise we will fail in reaching common results.

In the scope of the MINERVA project, our common goal was to preserve the European cultural heritage and make it available through the Internet to the public. Although multilingualism is only one aspect of this, it is essential to the cultural institutions to reach a wider audience. Even though English is the "lingua franca" in the European Union, individuals have the right to use their mother tongue. So it is of great importance to provide information on institutional websites in different languages. Internet users can easily cross official borders and visit as many places virtually as they want. There is a telling reason for institutions to deal with different virtual visitors, because they can become actual visitors in the future.

In the 25 countries that make up the European Union currently there are 20 official languages and many other languages are spoken. But only 45% of European citizens are capable of taking part in a conversation in a language other than their mother tongue.

European citizens want to live in a socially inclusive society in which diverse cultures live in mutual understanding, building at the same time a common European identity. Language, together with shared knowledge and traditions is an important part of an individual's cultural identity. The diversity of languages, traditions and historical experiences enriches us all and fosters our common potential for creativity. Respect for linguistic diversity constitutes one of the democratic and cultural foundations of the EU, recognised by the « European Charter of fundamental rights » in article 22. The « Council resolution on linguistic diversity » of 14 February

2002 recognised the role of language in social, political and economic integration

In the field of heritage, there is a significant importance of making information available to as wide an audience as possible and to overcome language barriers. Multilingualism plays a strategic role in the quality and effectiveness of communication on the Internet. Multilingual exchange of information is of interest for cultural tourism to reach visitors from neighbouring countries and therefore for the attractiveness of different territories and their economic development.

Whilst policies and initiatives aimed at preserving languages are the prime responsibility of the Member States, European action can play a catalytic role at the European level adding value to the Member States' efforts. The development of multilingualism on the Internet has been stimulated in the last years by the European Commission, by supporting the trans-national projects, fostering partnership between digital content owners and language industries. The New Framework Strategy for Multilingualism adopted in November 2005 by the European Commission underlines the importance of the multilingualism and introduces the European Commission's multilingualism policy: three aims are pointed out:

- to encourage language learning and promoting linguistic diversity in society;
- to promote a healthy multilingual economy, and
- to give citizens access to European Union legislation, procedures and information in their own languages."

Supporting high quality multilingual resources still needs to be enhanced. The MINERVA Plus pan-European survey will be of great interest and has already allowed us to point out best practices that will help to provide standardised solutions and shared knowledge in future.

The MINERVA Plus results also highlight reasons for multilingualism in the different countries including: self-presentation, protection of minorities, cultural heritage, support for regional development and tourism, scientific and cultural exchanges.

A continuation of this work would be helpful in working towards an inventory of existing mature linguistic tools, resources and applications as well as qualified centres of competence and excellence. Language technologies are both an essential tool for safeguarding Europe's rich cultural heritage and a source of future economic growth. As new language technologies develop they will make Europe's cultural heritage available to all, irrespective of language or location. This will be a boon to Europe's cultural industries, helping to unlock the vast resource that is European culture, art and history. Language technologies in short are essential to ensuring that all European languages – and the culture, art and history with which they are inextricably entwined - maintain their place in tomorrow's globalized, interconnected world.

Europe's experiences in multiculturalism and multilingualism represent an enormous strength. European cultural institutions should be able to exploit to position themselves in the new digital sphere of the information and knowledge society.

Institutions seem steadily adapting to the multilingual challenge as a growing number enable multilingual access to their collections. They use a wide variety of controlled vocabularies while indexing and documenting internally, but these tools are not visible to the end user of the websites. But the number of languages seem not to reflect the number of languages spoken Europe wide. English is the lingua franca of the cultural sites.

From the overview of projects we can make out that thesauri are more and more conceived as part of complex systems in which information is searched through a combination of methods.

While the number of multilingual cultural websites is increasing, multilingual controlled vocabularies are still scarce and the works are slow to produce quality and coherence in these vocabularies.

Reasons for the low number of the websites evaluated in the survey were found to use a thesaurus or taxonomies for thematic indexing:

- limited use of their collections
- lack of knowledge about multilingual thesauri available
- lack of knowledge about the implementation of multilingual thesauri on websites
- lack of development of appropriate thesauri for the cultural domain or of standardized translations of such resources.

Most search tools for the public are either based on full text searches or on query by form. Vocabulary aids are limited and mainly offer support in the form of a list of available indexing terms.



## 6 Future perspectives

As we have already introduced from different approaches, it is getting more important to *think multilingual*. Due to the quick development of the Information Communication Techniques, there are more and more tools, and facilities provided to support activities in the multilingual environment - especially on the Internet. Besides the new inventions, even traditional tools, like thesauri, can be implemented within electronic environment.

The promotion of the multilingualism should be continued at institutional level. The campaign could use the symbol of the Tower of Babel. Financial support for creating multilingual websites would be very appropriate with granting. A special icon could be given for the best practice example websites.

The number of thesauri all over the world can hardly be estimated, but we are quite sure, that almost every subject area has already been covered with one - in different languages. The best approach is to identify those thesauri, which are currently in use. The results of our survey, and the testimony of the country reports suggests, that several countries have very positive attitudes towards multilingualism, but limited uptake of controlled vocabularies. This reflects the lack of availability of multilingual thesauri for many EU languages and the scale of the work that's needed to offer this level of support.

So our suggestion within European context would be, instead of supporting the creation of brand new thesauri, it would be more useful supporting the translations of the well-tried, European wide used thesauri: like UNESCO, HEREIN, ICONCLASS, Library of Congress Subject Heading List on the European Commission level.

It would be useful to create a website for European multilingual thesauri, with the assistance of international standardization bodies, which would be a good information base for cultural institutions. The best practice examples, and the freely available thesauri could be highlighted there. It would be challenging to discover the black hole of those countries, from where we haven't got enough information on controlled vocabularies. More emphasis should be placed on developments of cross-language search facilities based on multilingual thesauri.

The thesauri developed internally by cultural institutions are a valuable asset both on a national level and an international level. By identifying the currently available thesauri and standardizing their multilingual qualities, these thesauri can serve many other institutions in the future. It would be also important, to prepare quality testbeds for existing thesauri, and discovering more evaluation methods, which could help the institutions to decide which one is convenient for their purposes. During the joint work only the Israeli working group used an evaluation method for their thesauri, the GLYPH criteria<sup>21</sup>. It would be our second recommendation for the future for international experts to test and evaluate the GLYPH criteria and other quality check techniques, and then to publish as an international working methodology - for testing websites that implement thesauri (rather than websites that host standalone thesauri).

The international survey results of the working group on multilingualism of MINERVA and MINERVA Plus and its resulting knowledge of available thesauri would be best harnessed to serve the cataloguing needs of other national and international cultural institutions with the hope of allowing freely accessible content in the languages of all European Union constituents.

---

<sup>21</sup> See Annex 1: Definitions.

## Annex 1

### Definition of terms used in the survey

**Bi/ Multilingual GUI** refers to the Graphic User Interface (GUI) on the front end. The user interface may be multi or bilingual while the controlled vocabulary may not exist or be monolingual and so this fact is noted.

**Bi-directional.** This is a specific issue pertaining to Semitic languages that differs from other languages in being read right to left. In most cases lexicons that are bi-directional can be opened in mirror image, an example of this could be reflected in the “search” button on the screen. The buttons that appear on the right for searching an English term would appear on the left to search for a, Arabic or Hebrew term.

**Truly Bi / multilingual - Bi /multilingual parallel cells.** If the lexicon is truly bi/ multilingual, the same number of results would be found if the term is searched in either language. The lexicon would also be able to act a translation tool. If the data were input in English, for example, the Arabic or Hebrew equivalent would fill in the parallel cell.

**Controlled vocabulary** is a lexicon built in a linear format. This list is similar to subject headings and includes pre-coordinated terms. Searches are performed by choosing from a list (facets). Example, Library Congress Subject Headings.

Another defition: Controlled vocabulary<sup>22</sup> is a list of terms that have been explicitly enumerated. This list is controlled by and is available from a controlled vocabulary registration authority. All terms in a controlled vocabulary should have an unambiguous, non-redundant definition. This is a design goal that may not be true in practice. It depends on how strict the controlled vocabulary registration authority is regarding registration of terms into a controlled vocabulary. As a minimum the following two rules should be enforced:

---

<sup>22</sup> What are the differences between a vocabulary, a taxonomy, a thesaurus, an ontology, and a meta-model?, <<http://www.metamodel.com/article.php?story=20030115211223271>>.

If the same term is commonly used to mean different concepts in different contexts, then its name is explicitly qualified to resolve this ambiguity.

If multiple terms are used to mean the same thing, one of the terms is identified as the preferred term in the controlled vocabulary and the other terms are listed as synonyms, aliases or non-preferred.

**Cultural Site:** is a website of a cultural institution (libraries, museums, archives) or a website providing cultural information having a digital collection (virtual galleries, cultural databases, historical sites).

**Free Text**

Almost all electronic databases allow free-text or keyword searching. In this type of search, the system usually looks for your search terms in every field of the record (not just in the subject heading or descriptor fields) and it looks for those terms to occur exactly as you type them, without mapping or translating them to controlled vocabulary terms.

**Graphic Lexicon Yielding Published Hyperlink (GLYPH)** – A set of criteria (defined below) established to evaluate multilingual controlled vocabularies, the format for cataloguing terms, the accessibility of the term lists, the additional of visual or multimedia aids - independent of language – that help define the terms and the vocabularies level of translations.

**The GLYPH System**

**A Grading System for Multilingual Lexicons (one point for each criteria)**

<b>GLYPH SYSTEM</b>	<b>GLYPH defined</b>
Online	URL
Bilingual / Multilingual Lexicon	defines lexicon as bi or multi
Bi-directional	right to left and vise versa
Lexicon / Thesaurus / Classification	linear / one to many /hierarchical
Browseable lexicon access / Tree	terms accessible via browse
Bi / Multi Languages	The lexicon interfacing
Image/ multimedia	a visual aid
Bilingual parallel cells	same result in either language

**Integrated images** – an image is provided to help express the meaning of a term.

**Multilingual website:** is a website providing information in two or more languages.

**Thesaurus** is a networked collection of controlled vocabulary terms. This means that a thesaurus uses associative relationships in addition to



parent-child relationships. The expressiveness of the associative relationships in a thesaurus vary and can be as simple as “related to term” as in term A is related to term B.

A thesaurus has two kinds of links: broader/narrower term, which is much like the generalization/specialization link, but may include a variety of others (just like a taxonomy). In fact, the broader/narrower links of a thesaurus is not really different from a taxonomy, as described above. A thesaurus has another kind of link, which typically will not be a hierarchical relation, although it could be. This link may not have any explicit meaning at all, other than that there is some relationship between the two terms.

Additional information about thesauri:

What controlled vocabularies, taxonomies, thesauri, ontologies, and meta-models all have in common are:

They are approaches to help structure, classify, model, and or represent the concepts and relationships pertaining to some subject matter of interest to some community.

They are intended to enable a community to come to agreement and to commit to use the same terms in the same way.

There is a set of terms that some community agrees to use to refer to these concepts and relationships.

The meaning of the terms is specified in some way and to some degree.

They are fuzzy, ill-defined notions used in many different ways by different individuals and communities.

Thesaurus (1) Can be reflected as one word to many or (2) Can be more expanded and have classified terms set in a hierarchical manner.

Searches may be performed by choosing from a list or by typing a free text (Boolean). This list includes post-coordinated terms. An example of a classified thesaurus – Getty’s Art and Architecture Thesaurus.

## Annex 2

### Interesting projects

The EACHMED project, has developed a portal published by CNR (Italian National Centre of Research): <<http://www.eachmed.com>>. The project aims to make this site available in 32 languages, including Latin. It will implement a multilingual thesaurus about cultural heritage in the 32 languages produced by another CNR project, Progetto Finalizzato Beni Culturali <<http://www.pfbeniculturali.it>>.

The Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche, Pisa coordinates the Cross Language Evaluation Forum (CLEF) <<http://clef.iei.pi.cnr.it/>>. CLEF develops the infrastructure for the testing, tuning and evaluation of information retrieval systems operating on European languages in both monolingual and cross-language contexts.

The Accademia Europea di Bolzano per la ricerca applicata e la formazione post-universitaria ([www.eurac.edu](http://www.eurac.edu)) is member of the IST project SALT, Standards-based Access to multilingual Lexicons and Terminologies <<http://www.loria.fr/projets/SALT/>>, an open-source project that is producing ISO standards or contributing to revised ISO standards.

ITC-IRST Trento, with its The Cognitive and Communication Technologies (TCC) division, takes part to the MEANING project (Developing multilingual web-scale language technologies), <<http://www.lsi.upc.es/~rigau/meaning/meaning.html>> concerned with automatically collecting and analysing language data from the WWW on a large scale, and building more comprehensive multilingual lexical knowledge bases to support improved word sense disambiguation.

Italian private companies are partners in the IST funded project MIETTA II, A Multilingual Information Environment for Travel and Tourism Applications <<http://www.mietta.info/>>.

## Annex 3

### Registered thesauri on the survey's website

[http://www.mek.oszk.hu/minerva/survey/contr\\_vocs2.htm](http://www.mek.oszk.hu/minerva/survey/contr_vocs2.htm)

<b>Name of the thesaurus</b>	<b>Coverage</b>	<b>Languages</b>
Biologic Taxonomy	Names of species (Animals and Plants)	Latin
THESAURUS (of architecture)	Edifices and Furniture	French, English, American, Italian
THESAURUS (of religious objects)	Religious furniture and clothes	Italian, French, English
HEREIN (European Heritage Network) thesaurus	Architectural and archaeological heritage policies	English, French, German, Spanish, Bulgarian, Polish, Slovenian
MALVINE (Manuscripts and Letters via integrated Networks in Europe)	Manuscripts and moderns letters	German, English, French, Spanish, Portuguese
NARCISSE (Network Art Research Computer Image SystemS in Europe)	Preservation and restoration of paintings	German, Italian, Portuguese, French, English, Spanish, Catalan, Danish, Russian, Chinese, Japanese
UNESCO Thesaurus	Education; culture; natural sciences; social and human sciences; communication and information; politics, law and economics; countries and country groupings	English, French, Spanish
RAMEAU (Répertoire d'autorité-matière encyclopédique et alphabétique unifié)	Catalogues of libraries	French
PACTOLS ( Peuples et Cultures, Anthroponymes, Chronologie relative, Toponymes, Oeuvres, Lieux, Sujets)	Sciences of Antiquity	French; Italian and English; German and Spanish
MACS (Multilingual Access to Subject)	Catalogues of libraries	German, French, English
Museum Images themes	art, architecture, sciences, technology, history...	English, German, Italian, French, Spanish

Museum Images Artist Names	Artist names	French, English
Museum Images Periods		French, English
Objektdatenbank, OPAC Bibliothek		German
Hessische Systematik		German
Allgemeines Künstlerlexikon		German
Thesaurus of Geografic Names		English, German, French
United List of Artist Names		English, German
Iconclass-Deutsch		English, German, French
Schlagwortnormdatei		German
PKNAD (prometheus KünstlerNamensAnsetzungsDatei)	Names of Artists	German
Seitendateien		German
Basisklassifikatoin		
Personennamendatei		
Gemeinsame Körperschaftsdatei		
Dewey Dezimal Klassifikation		
Universale Dezimalklassifikation		
Ethno-Guide:Type of Sources	Sourcetypes	English, German
Thematic Index		English, German
Regensburger Verbundklassifikation		English, German
Zeitraum	time period	
geografische Region	geographic subject	English, German
Quellentyp	Sourcetype	English, German,
Schlagwörter	subject heading	English, German
Econinfo	Area: social science ; sub-areas: economics, business and management, sociology, political science, public administration, international relations, environmental	Hungarian, English, German
Hungarian Educational Thesaurus	area: social science sub-area: education science,	Hungarian, English, German, French,

	psychology	
Library of Congress Subject Headings in Hungarian	all subject areas	Hungarian, English
OSZK Thesaurus	social science, natural science, geographical names	Hungarian
Thesaurus of library and information science	library and information science and some related fields, e.g. bookselling and publishing, computerization, history of books, printing and press etc.	Hungarian, English
WebKat.hu tárgyszórendszere	Every discipline	Hungarian
Alinari		Italian, English
ambito culturale ATBD	architecture, art-history, archaeological objects and sites	Italian
autore - qualifica AUTQ	architecture, art-history, archaeological objects and sites	Lithuanian
autore - scuola d'appartenenza	architecture, art-history, archaeological objects and sites	Russian
Descrizione Iconografica DESS	architecture, art-history, archeological objects	Italian
e-learning glossary	e-learning	
ICONCLASS IN ITALIAN	The iconography of the west art from the medieval period to the contemporary art	italian, english, German, French, other: Finish
Materia e tecnica - oggetti d'arte - MTC	artistic objects	Italian
Materia e tecnica - archeological objects - MTC	archaeological field	Polish
Oggetto definizione - artistic objects - OA	artistic objects	English, Italian, French, Portuguese, other: language only some sections of these
Oggetto Tipologia - Artistic Objects - Oa	artistic objects	English, Italian, French, Portuguese, other: language only some sections of these
ThIST (Italian Thesaurus	Earth Sciences	Italian, English

of Earth Sciences)		
Tipologia dell'oggetto - Architectonical objects	architectonical area	Italian, English, French
ARTIST	ARTIST'S NAMES	Russian, English
TITLE	TITLE OR NAME	Russian, English
SCHOOL	ARTIST SCHOOL	Russian, English
STYLE	STYLE OF ARTWORK	Russian, English
TYPE OF ARTWORK	STYLE OF ARTWORK	Russian, English
COUNTRY / ORIGINAL	country where the artwork was created	Russian, English
THEME	the domain in which a searcher is interested	Russian, English
GENRE	ICONOGRAPHIC GENRES	Russian, English
PERSONAGE	ICONOGRAPHY: PERSONAGE REPRESENTED BY THE ARTWORK	Russian, English
PAINTERS	names of painters connected with the creation of works of art	Russian
Special terms	TECHNIC OF CREATION AND RELATED NOTIONS	Russian
Vocabulary of fine arts terms	painting technique and appellations	Russian
PLACE OF CREATION	country, town etc. where the artwork was created	Russian
MANUFACTURE	factory, plan, lithography, art, work association etc. that took part in the creation of the artwork	Russian
PERSONAGES	represented people, area of iconography	Russian
MATERIALS AND TECHNIQUES	materials from which the object is done and techniques that was used for its creation	Russian
THEMES	theme subdivisions of the museum	Russian
FUNDS	museum reserves	Russian
data element catalogue	The data element catalogue is supposed to cover objects from cultural history, photos, literature, archaeology, theatre, industrial history, art	Swedish

	history, technical history, buildings and environmental values	
Art & Architecture Thesaurus - Nederlandstalig	material culture in general (with a focus on art history and archaeology)	Dutch, English
ARENA Periods	Cultural Heritage	English, Danish, Norwegian, Icelandic, Polish, Romanian
ARENA Top Level Themes	Cultural Heritage Sites and Monuments	English, Danish, Norwegian, Icelandic, Polish, Romanian
AV/Webcasting search pilot tool		
Bilingual Welsh/English subject index	cultural heritage within Wales	English, Welsh
Collection Subject Search		
Glossary	arts	
Scotland's Culture Thesaurus	All aspects of Scottish Culture	English
Subject search (indexed text search)	arts etc	
Term lists from TMS (e.g. object type)	arts	English
Thesaurus of Monument Types	Archaeology - specifically archaeological monuments in England	English
The Bar-Ilan University Controlled Vocabulary		English, Hebrew, Arabic, Russian, French, Italian, German
The Beth Hatefutsoth (Museum of Jewish Diaspora) Controlled Vocabulary	history, art, folklore, ceremonial art, architecture, Jewish life, Jewish music (liturgical, para-liturgical, traditional)	Hebrew, English
The Bibliography of the Hebrew Book, 1473-1960 Controlled Vocabulary		Hebrew, English, Ladino, Judeo-Arabic
The Center for Computerized Research Services in Contemporary Jewry Controlled Vocabulary		English
The Central Zionist Archives Controlled Vocabulary		Hebrew

The eJewish Controlled Thesaurus	Jewish studies, Israel	Hebrew, English, French, Russian, Spanish
The Hadashot Arkheologiyot – Excavations and Surveys in Israel online publication by Israel Antiquities Authority Controlled Vocabulary		Hebrew, English
The Haifa University Thesaurus	all subject areas	English
The Index to Hebrew Periodicals (Haifa Univ.) Thesaurus		English
The Israel Antiquities Authority List		Hebrew, English
The Israel Antiquities Authority Controlled Vocabulary	archaeology, architecture, finds, periods of ancient Israel, periods of ancient Near East, etc, architectural elements of archaeological sites in Israel, archaeological periods of ancient Israel	Hebrew, English
The Israel Folktale Archive Thesaurus	folktales, folklore, folk-literature, literature, Jewish studies	Polish, Moroccan, Hebrew, Yemenite, Iraqi Arabic, Yiddish, Ladino, Tunisian Arabic, Kurdish, Russian, Farsi, Rumanian, Arabic - English planned
The IMAGINE Thesaurus	Artists, Materials, Object name, Keywords, Periods, Place and Technique. A special sub-table in the keywords table is the "Judaica and Ethnography categories"	Hebrew, English
The Jerusalem Virtual Library – The Academic Database On Historic Jerusalem Thesaurus		English
The Jewish National & Univ. Library, RAMBI, Index of articles in Jewish Studies Controlled Vocabulary	Jewish studies, Israel	English, Hebrew, Arabic, Russian, French, Italian, German, Ladino, Yiddish
The Knesset Controlled Vocabulary		English, Arabic, Hebrew



The MALMAD - Israel Center for Digital Information Services Controlled Vocabulary		Hebrew, Arabic and English
The MOFET Institute Thesaurus		Hebrew, English
The Musical Library, Levinsky College Controlled Vocabulary	music	Hebrew, English
The Pro Jerusalem Society Controlled Vocabulary		English and Hebrew
The Steven Spielberg Jewish Film Archive Controlled Vocabulary		English
The The Aviezer Yelin archives of Jewish education in Israel and the Diaspora Controlled Vocabulary	history of Jewish education, Jewish schools, educators	Hebrew, English
The The Ben-Gurion Research Institute Controlled Vocabulary	David Ben-Gurion, State of Israel, Diaspora, Holocaust, Israeli wars, Israeli society, Zionism	Hebrew, English, French, Arabic, Spanish, Italian, German, Yiddish, Dutch, Swedish, Russian, Polish, Danish, Greek, Romanian, Turkish, Portuguese, Bulgarian, Hungarian
The The Henrietta Szold Institute Thesaurus	social sciences, education	English, Hebrew, Arabic, Russian, French, Italian, German
The The Moshe Dayan Center Bibliographical Database Controlled Vocabulary		French Arabic and English
The The Tel-Aviv Museum of Art Controlled Vocabulary	Visual arts	Hebrew, English
The The Vidal Sassoon International Center for the Study of Antisemitism Controlled Vocabulary		English
The The Yad Ben Zvi Controlled Vocabulary		English, Hebrew, Arabic, Russian, French, Italian, German
The U. Nahon Museum of	history and art of Italian	Hebrew, English

Italian Jewish Art Thesaurus	Jews	
The Wingate Institute for PE & Sport Thesaurus	natural sciences, social sciences, humanities, sport, physical activity, physical education	Hebrew, English
The Yad Vashem Archive Thesaurus	geography, names	Hebrew, English, French, Spanish, Italian, German, Yiddish, Dutch, Portuguese
Archaeological Thesaurus compiled as part of the Luxembourg National Research Fund (FNR) 'Environment and Cultural Heritage' Project	Description of archaeological discoveries and results (various time periods, various categories of archaeological material), geological and geographical terms	Luxembourgish, French, English, German, Latin
Musée National d'histoire Naturelle – Service d'Information sur le Patrimoine Naturel / Institut Grand-Ducal section de linguistique	Names of Plant and animal species of Luxembourg	Luxembourgish, French, English, Latin
Archaeological Thesaurus compiled as part of the Luxembourg National Research Fund (FNR)	time periods, various categories of archaeological material, geological and geographical terms	Luxembourgish, French, English, German, Latin
Common Names of plants and animals	Names of plant and animal species of Luxembourg	Luxembourgish, French, German, English, Latin

Other collection of thesauri and tools  
by A.J.Miles.(a.j.miles@rl.ac.uk <a.j.miles@rl.ac.uk>  
<[http://www.w3c.rl.ac.uk/SWAD/thes\\_links.htm](http://www.w3c.rl.ac.uk/SWAD/thes_links.htm)>