**Automating Production of Cross Media Content
for Multi-channel Distribution**
www.AXMEDIS.org

DE8.3.1
Multilingual Guidelines and
Technical Solutions

Version: 1.2
Date: 30/09/2005
Responsible: ILABS (revised and approved by DSI)

Project Number: IST-2001-37168
Project Title: Automating Production of Cross Media Content for Multichannel Distribution
Deliverable Type: Report
Visible to User Groups: Yes
Visible to Affiliated: Yes
Visible to the Public: Yes.

Deliverable Number: DE8.3.1
Contractual Date of Delivery: 30/09/2005
Actual Date of Delivery: 10/11/2005
Title of Deliverable: Multilingual guidelines and technical solutions
Work-Package contributing to the Deliverable: WP8
Task contributing to the Deliverable: WP8.3
Nature of the Deliverable: Report
Author(s): ILABS, XIM, SEJER, AFI, ANSC, UNIVLEEDS, FUPF

Abstract:
The present document provides information on state of the art, references, techniques & solutions for multi-language support, not limited to, but certainly centered on, data & metadata management.

Keyword List:
Content, multi-language, support, technology, meta-data

Table of Content

AXMEDIS Copyright Notice

The following terms (including future possible amendments) set out the rights and obligations licensee will be requested to accept on entering into possession of any official AXMEDIS document either by downloading it from the web site or by any other means.

Any relevant AXMEDIS document includes this license. PLEASE READ THE FOLLOWING TERMS CAREFULLY AS THEY HAVE TO BE ACCEPTED PRIOR TO READING/USE OF THE DOCUMENT.

1. **DEFINITIONS**

    i. "**Acceptance Date**" is the date on which these terms and conditions for entering into possession of the document have been accepted.

    ii. "**Copyright**" stands for any content, document or portion of it that is covered by the copyright disclaimer in a Document.

    iii. "**Licensor**" is AXMEDIS Consortium as a de-facto consortium of the EC project and any of its derivations in terms of companies and/or associations, see www.AXMEDIS.org

    iv. "**Document**" means the information contained in any electronic file, which has been published by the Licensor's as AXMEDIS official document and listed in the web site mentioned above or available by any other means.

    v. "**Works**" means any works created by the licensee, which reproduce a Document or any of its part.

2. **LICENCE**

    1. The Licensor grants a non-exclusive royalty free licence to reproduce and use the Documents subject to present terms and conditions (the **Licence**) for the parts that are own and proprietary property the of AXMEDIS consortium or its members.

    2. In consideration of the Licensor granting the Licence, licensee agrees to adhere to the following terms and conditions.

3. **TERM AND TERMINATION**

    1. Granted Licence shall commence on Acceptance Date.

    2. Granted Licence will terminate automatically if licensee fails to comply with any of the terms and conditions of this Licence.

    3. Termination of this Licence does not affect either party's accrued rights and obligations as at the date of termination.

    4. Upon termination of this Licence for whatever reason, licensee shall cease to make any use of the accessed Copyright.

    5. All provisions of this Licence, which are necessary for the interpretation or enforcement of a party's rights or obligations, shall survive termination of this Licence and shall continue in full force and effect.

    6. Notwithstanding License termination, confidentiality clauses related to any content, document or part of it as stated in the document itself will remain in force for a period of 5 years after license issue date or the period stated in the document whichever is the longer.

4. **USE**

    1. Licensee shall not breach or denigrate the integrity of the Copyright Notice and in particular shall not:

        i. remove this Copyright Notice on a Document or any of its reproduction in any form in which those may be achieved;

        ii. change or remove the title of a Document;

        iii. use all or any part of a Document as part of a specification or standard not emanating from the Licensor without the prior written consent of the Licensor; or

        iv. do or permit others to do any act or omission in relation to a Document which is contrary to the rights and obligations as stated in the present license and agreed with the Licensor

5. **COPYRIGHT NOTICES**

    1. All Works shall bear a clear notice asserting the Licensor's Copyright. The notice shall use the wording employed by the Licensor in its own copyright notice unless the Licensor otherwise instructs licensees.

6. **WARRANTY**

   1. The Licensor warrants the licensee that the present licence is issued on the basis of full Copyright ownership or re-licensing agreements granting the Licensor full licensing and enforcement power.

   2. For the avoidance of doubt the licensee should be aware that although the Copyright in the documents is given under warranty this warranty does not extend to the content of any document which may contain references or specifications or technologies that are covered by patents (also of third parties) or that refer to other standards. AXMEDIS is not responsible and does not guarantee that the information contained in the document is fully proprietary of AXMEDIS consortium and/or partners.

   3. Licensee hereby undertakes to the Licensor that he will, without prejudice to any other right of action which the Licensor may have, at all times keep the Licensor fully and effectively indemnified against all and any liability (which liability shall include, without limitation, all losses, costs, claims, expenses, demands, actions, damages, legal and other professional fees and expenses on a full indemnity basis) which the Licensor may suffer or incur as a result of, or by reason of, any breach or non-fulfilment of any of his obligations in respect of this Licence.

7. **INFRINGEMENT**

   1. Licensee undertakes to notify promptly the Licensor of any threatened or actual infringement of the Copyright which comes to licensee notice and shall, at the Licensor's request and expense, do all such things as are reasonably necessary to defend and enforce the Licensor's rights in the Copyright.

8. **GOVERNING LAW AND JURISDICTION**

   1. This Licence shall be subject to, and construed and interpreted in accordance with Italian law.
   2. The parties irrevocably submit to the exclusive jurisdiction of the Italian Courts.

Please note that:
- You can become affiliated with AXMEDIS. This will give you the access to a huge amount of knowledge, information and source code related to the AXMEDIS Framework. If you are interested please contact P. Nesi at nesi@dsi.unifi.it. Once affiliated with AXMEDIS you will have the possibility of using the AXMEDIS specification and technology for your business.
- You can contribute to the improvement of AXMEDIS documents and specification by sending the contribution to P. Nesi at nesi@dsi.unifi.it
- You can attend AXMEDIS meetings that are open to public, for additional information see WWW.AXMEDIS.org or contact P. Nesi at nesi@dsi.unifi.it

1 Executive Summary and Report Scope (ILABS)

To be highly attractive content should be available in multi-language format with possibility of delivery on multiple platforms. The present document focuses on the issue of multi-lingual support both at metadata and content level. Technical solutions for both issues are presented and described in detail. The performed work is related to the following Task.

WP8.3 Content Integration and multilingual support -- responsible ILABS -- For cross media content delivering the problem of integrations and multilingualism are strongly relevant. This has the goal of integrating the content identified in the WP8.2 with additional information and with multilingual aspects. The integration will have to be supported by specific tools that will be identified and developed in this WP. Thus integration involve: improving metadata, adding text in other languages, addition of other related content components, synchronization among different media (such as an audio with the text, a video with the text, sequence of image with an audio) etc. This process will have to be done in a large scale to have content for validation and to really prove the effectiveness of AXMEDIS tools for accelerating the composition and formatting and thus for providing content on demand. For the translation of text standard commercial tools will be used while for the other aspects specific activities and tools will have to be identified and developed. The development will be performed in WP5. Other details: A first version of guidelines and defined mechanisms for the production of content with multilingual support, multilingual metadata.

1.1 Responsibilities

ILABS is primary responsible for the completion of the present document, yet all content owning partners are responsible for providing the required contributions including content samples (in terms of aspect, description and metadata). Technical partners are responsible of document revision in terms of coherence with expected supported formats. Document sections responsible are reported in the section title within brackets. When a section, or subsection, has no explicit responsible is assumed that is in charge to the responsible of the previous one or is a general one to which all partners shall contribute.

2 A brief introduction to multi-language content (ILABS, XIM, ALL)

The fast globalisation of markets, the need for cooperation with publishers and distributors all over the world and the tendency to provide 7x24 service also in access to content or ODL, could already represent a major push to adopt multi-lingual solutions in terms of delivery and content, but this would be limiting and could cause serious problems every time some content has to be managed, revised or edited somewhere around the globe. It is also interesting to take into account that multi-language management is crucial to ensure proper accessibility to several kinds of contents, from movies to text. In many countries it is common to dub video content to match the taste of the local consumer. This has some clear advantages and some hidden disadvantages. For example, if we take into account a comedy, or any other audio-visual content, produced in an English speaking country relying on the implicit irony that can be achieved and exploited using English, it is apparent that it will not be possible to achieve the same result with dubbing. First of all because a literal translation of the original dialogue may present no comical / ironical aspect in the language used to dub, secondly as the achievement of a similar effect may require a totally new set of text that may be very difficult (if not impossible) to properly sync with labial movement of the characters speaking in English.

Another relevant reason that is presently pushing content owners to tackle the multi-lingual issue is the fact that, more and more often, professional practitioners, and consumers, are searching for content on the web. Search engines will retrieve lots of content and often some pieces will be in a language unknown to the user. Most efficient search engines provide a set of basic translation services, but if someone really uses them they will find that the results are still quite poor. The search and retrieval issue is not marginal as it does apply not only to content available over the Internet, but also, and more relevantly, to content located into intranets, databases, archives, CMS, LCMS of KMS. Such dispersed and varied kind of "content repositories" represent often a major asset for a company or organisation as they hold and provide info, documents and other "objects" used for work, research, etc. and therefore have to be not only accessible, but quickly accessible to

5

user. The combination of needs for a quick and reliable access to a specific relevant piece of information, document or object on the basis of a research is often (we could even say always) based on the archival procedure and the set of data available for indexing and retrieval. This latter point is not limited to a specific language, on the contrary is more and more pushing towards a multi language environment. The combination of the previously mentioned aspects is at present a huge issue and represents the object of many studies both at academic and commercial level. The primary objective of much of the abovementioned research and development is focused on the provision of a transparent way to support users in the search and retrieval process. This basically means to enable the user to place a query (even a complex one) into a selected language and retrieve content that is relevant, and available in the same language as the one used for the query, even when this implies the need to provide a translation or just a basic set of information taken form the object metadata and covering relevant aspects of the object itself.

Another general problem for multilingual content is the need to understand the context of the text that is being translated. For example, subtitling needs to be consistently translated throughout the video so that places, objects and people are not given different translations in different subtitles due to lack of context. This is also true for text labels used within multimedia for labelling illustrations and animations. An automated translation process might not associate the labels as being related and therefore translate the same word or concept differently in separate labels. One may think that this kind of concern is purely academic and that a good translator could always work it out, yet this is not always true on a side because context is highly relevant for translation (it is the primary link factor to the culture embedded in the source language) and therefore special care has to be placed in ensuring proper contextualisation also in the target language. This brings along a major problem related to cultural aspect of languages. It is known that populations like the Inuit have over 40 terms to describe "snow" while people coming from other part of the world may have much less or even none. In other words quoting Adeline Yen Mah[1]: "Concepts are expressed by words. If certain English words are missing in Chinese, it follows that the concept expressed by those words will be absent in China, and vice versa". This nay force translators to adopt some strategies or work around to address the problem and convey the meaning, yet if this is basically feasible for humans is still far to come for automatic translators. The problem is even more evident when the addressed content to be tackled has got some artistic value as it is almost impossible to convey in a different language some subtleness or specificities that make that specific content a masterpiece.

3 Language related issue (ILABS, XIM, ALL)

The Internet has changed considerably from the early days. Previously the bulk of online content was presented in English, however, the last 5 years have seen a burst in the volume of content in languages other than English, as well as in the number of non-English speaking users. Right now less than 35% of Internet users have English as their first language (compared to 50% in 2000). This trend is presently confirmed (mainly due to the development rates in Internet connections of countries like China and India) and it is predicted that such figures will drop to 25% by 2008. Indeed, some of the busiest sites in the world do not even have an English version like: *Daum.net*, a large Korean portal available in Korean only and that is frequently reported in the top 5 visited sites in the world. Besides absolute volume of potential site visitors as measured by language, there are several other conditions to take into account when planning a web based multilingual content production, sales and distribution.

For example if the provided Internet-based access to content is just for information provision service or branding, rather than for representing a point of sale, there are other areas to pay attention to like how do different cultures search for information. Each culture, and language, has its own underlying logic. Site navigation, keyword targeting and intuitive use will all vary between languages. Even within languages, there are strong differences. As a simple example, UK English speakers are more likely to search for the term "*vacation*" whereas USA English speakers will use "*holiday*" more frequently.

On the other hand if the site is providing a point of sale, it will be crucial to take into account under what framework will users operate, for example if they will purchase online, or if their cultural constraints will imply that they would rather communicate directly with a local distributor. Furthermore it will be crucial to figure out how will users purchase, as it may be the case that online credit card purchasing is an accepted method within this language group/culture, while for other target group it will be needed to provide other purchasing options (telegraphic transfer etc.); each of this aspect will be reflected in the localized version.

Last but not least it will be necessary to take into account the enquiry processing, as if the translated site is likely to generate an enquiry, this will need handle, yet for some sites it could easy to develop tightly structured generic enquiry forms, while for others email processing may be required.

In the present section of the document are reported the most relevant issues related to language that affect content. We have already mentioned that audio-visual content is often dubbed, but this is just an example of how to deal with the issue of having content in a different language from the one of the user. Still in the case of audio-visual content is often possible to have subtitling. This in turn presents other kinds of issues like the synchronisation one. There is then another full set of cases to be taken into account and among them the most relevant is the one related to metadata management. It will be clear how relevant it is to have not only the content accessible in several languages, but also metadata. It may seem a paradox, but in some cases if metadata are not supporting multi-lingual management, even a fully multi-lingual content may fail to be used or found by a user having a specific language issue. What follows gives a quick summary of the issues of managing multilingual content/websites. It is generally accepted that there are seven key issues to be considered in respect of multilingualism, nominally: *Translation, Localization, Culture, Feedback, Design, Workflow and Non-Latin character sets.*

Translation - is essential for multilingual website and will require qualified personnel or the use of an external translation service; in any case proof reading of translated work is required. Machine translation should be avoided and, if considered, only with extreme caution as it may be a plausible alternative for infrequently accessed pages containing non-essential content. In this case, the use of short, unambiguously structured sentences and the avoidance of idiomatic phrases are essential, and sub-editing is still a necessity as, depending on content domain errors may be introduced despite the quality of the used system. On any site with significant translation requirements, translation costs are likely to dwarf all other running costs.

Localization - depending on the addressed content domain is a more or less complex process. Usually multilingual object (website, CD…) is a mixture of global and local content. Local content presents no particular content management issues while global content (which has to be translated across all language supported) does. Deciding where multiple language versions of content are going to be required and where content can be maintained separately for different languages is a critical decision that will affect how a the products should be maintained and what it will cost.

Culture - Differences in language are only part of what distinguishes different content instances. Graphical conventions, matters of taste, sense of humour, socially acceptable forms of address and issues of privacy all vary from place to place. Moreover it has to be taken into account that whenever a word is missing in a specific language it is often because the expressed concept is missing and therefore some important concepts may have no useful meaning if translated literally, or can even turn into tricky or sense-less piece of content (think of *"natura morta"* that should be translated as "*still life*" and not as "*dead nature*"). This may also become trickier when you take into account sub-cultures. For instance, in France, the "*stop*" sign is "*stop*" while it is "*arrêt*" in Québec, as "*Rollerblade*" is "*Rollerblade*" in France and "*Patins à roulettes à roués alignées*" in Québec.

Feedback - is another crucial issue, as responses to any feedback will need to be addressed in the language of the initial communication. User feedback should not be solicited in a language if it cannot be routed to a suitably qualified person/system/content that can answer/be accessed in the appropriate language. Scripts that handle interactivity, such as discussion forums, search results and feedback forms, will also need to be configured appropriately.

Design - is perhaps the most common, and an easily overlooked, difficulty encountered in developing multi-lingual products. This applies both at production and maintenance as it implies a consistent design across different language versions, and in particular of navigation layout. Text or graphic labels that fit the design constraints in one language may not work well in translation. The only sensible way to tackle this issue is to ensure that the initial design brief for a specific content already includes all language variations of branding and of the major navigational elements that could be of relevance or expected to happen (if the original content is relevant for the German market it would be sensible to design it in a way to host a German translation even if the original content is English or French…). Also, links between pages/components should not lead unsuspecting users from one language locale into another.

Workflow - even though this aspect may appear quite out of scope at first it is essential reminding that in a digital world often operations performed by users may be integrated either in a workflow as part of a work process (production, aggregation, distribution…) or as a part of a controlled value chain (acquisi-

7

tion and fruition of content on a pay per use basis). Simple workflow mechanisms usually offer some kind of notification when some action is performed on a content/page or when the content/page moves from one state to another. Translation workflow, on the other hand, requires that changes to a content/page trigger appropriate notification of required changes to the other language versions of that content/page. In addition, it is usually helpful to have some mechanism for identifying which elements, within the content/page, have changed.

Non-Latin character sets - Given the present structuring of the e-market and the inherent globalization the multilingual approach is impossible to avoid facing the need to address those markets where numbers are so relevant that can make the difference between a profitable and a just sustainable business (China, Russia, Arab countries). There are some interesting challenges associated with the creation and rendering of non-Latin alphabets, although modern browsers have better support for them than in the past thanks to Unicode as the recognized (and growing) international standard that includes most non-Latin characters and makes storage and retrieval of non-Latin characters in distributed environments (such as the web) much easier. Still Unicode support for some character sets (such as Bengali) is still not universal, so the use of legacy character sets may occasionally be necessary (at least in the short term), but, ideally, content should be stored and edited as Unicode. In addition, content/website pages should be published with an appropriate character set (either UTF-8 or a language-specific character set, such as windows-1256 an Arabic subset of Unicode) and language META tags; any characters that are not in the publishing character set should be published as html entities; and direction tags should be specified, where appropriate. It should also be possible to mix languages on a single page, allowing links to other language versions of the content/page to be handled simply.

3.1 Presenting Multilanguage content - practical issues

Further to what already stated is worth mentioning that in the current global environment most organizations (commercial, governmental, nonprofit…) have the need to present their content in multiple languages. This is not an issue or problem free effort and what follows presents shortly a few practical issues and experiences related to this topic, in order to help achieving the best possible result within the given resources.

3.1.1 Benefits

For a commercial company having multilingual content and web site presents two main benefits. Obviously it allows access to customers and partners in many countries. But most importantly is a very effective tool for finding new customers. It is rapidly becoming routine for people to search for potential content and suppliers over the Internet. In our experience (and the one of our partners) between 60% and 80% of traffic on companies Web site is generated by searches for content/products, and is performed using search engines. Such searches will lead customers to the target only if the information provided/available presents the content in the language used for the potential customer's search.

3.1.2 Logistical issue

Maintaining a medium sized content collection (and the related Web site) in a single language used to involve a cumbersome process of preparing the content by authorized people and then having technically skilled personnel to handle this content either for storage or for the Web site. Since this process involved 2 (groups of) people it was relatively costly, prone to delays, errors in communication etc. The advent of various authoring and mainly content management tools made it possible to often reduce the above 2 stage process to just one stage one with direct editing and content management performed by the people responsible for the content production. Obviously this is a desired outcome yet is referring to a mono-language situation. We omit here the (non-trivial) issues of multistage authorization of content changes and the issue of restricted domains of authority over various kinds of content or parts of the related storage or Web site. These issues are usually resolved by a decent content management system, but they do not necessarily arise in the case of smaller organizations. On the other side when the content (and its related Web site) starts to be multilingual, another step in the content production process need to be taken into account and handled: **translation**. Even for a small content collection or a small web-site, having to manually handle the submission of text for translation, receive the translated versions back and upload the translations to the DB/Web site is a tedious, error prone and slow process. As the content has to or should be updated/expanded often, the delays/complexity introduced by the translations becomes a serious problem. In this view is highly recom-

mended that owners of multilingual content consider some sort of content management system which also organizes, speeds up and makes simpler the management of the translation process.

3.1.3   Content consistency issue

For a multilingual content collection (and the related Web site) a problem that is often encountered is the issue of how to handle and manage the structure similarities/differences among the various language versions. In most cases content owners would like to have, ideally, the same structure for every language version, as this will greatly simplify the production and maintenance process.

Surprisingly, what just stated applies despite the content format/media as it can be easily verified by checking the various available guidelines for publishing, desktop publishing, web-publishing, media editing, etc. This basically means that any language version should have approximately the same amount of content/Web pages; it should be cross-linked basically in the same way, with the same, or comparable, navigation systems (buttons, menus) across various language versions. Additionally, the corresponding content/Web pages in different languages should be cross-linked to each other.

This requirement implies that any extension/update to the content/Web site must be done simultaneously in all languages. This is often not practical. Most often the content is created in one 'basic' language, then translated (possibly using computer-assisted procedures). In most cases the translation process takes time, delaying the publication of all language versions of the update, including the 'basic language' version and often this delay is not acceptable (at least in present days tight "time to market" constrained environment). It has to be accounted that often publishers have an alternative approach to the abovementioned issue to cope with the delays due to the translation process (other options are listed later) that is to allow different language versions of the content to have somehow different content and structure; this is often done with websites and in certain case also with content. In such case the primary language will provide a full fledge content while other languages will provide only either basic content or a summary. While this is one of the approaches that could be recommended, some related potential dangers should be briefly examined. Obviously one cannot blindly translate and publish in 'foreign' languages random content/pages from the 'base language' version of the content/site. Most content/pages contain links, buttons etc. leading to other content/pages. It is unacceptable to translate just on content/page of a set and have it contain 'dead' links and buttons to content/pages which were present in the 'basic language' version, but are not yet translated (and published). When the deferred progressive translation approach is followed the risk that the user could found similar harmful manifestations of the reported problem is likely to happen especially as far as sites are concerned as their complexity (as an object) is usually high, yet something like this can happen also in CD or products. Sometimes it happens that given the way translation process is carried out the output is then used either replacing / adding portions of the translated content in several sections of the object – it could be in the text, interface, graphics (buttons, tabs…), dialogues, menus… – and this could be done either working with traditional tools working on text/image/audio/video (Word, Photoshop, Soundforge, Avid…) and some with development tools that may vary quite greatly on the basis of object nature (director, visual studio, aftereffect…) this may cause that text produced in the translation process may need to be processed in complex environment (often via cut and paste) leading to a potentially error-prone process with a probability that is increasing with the complexity of the object being localized. Having said this is necessary to point out that links leading to not-yet-translated content 'hanging' ('dead') will cause a lot of user frustration. If in sites is often used to have a process that make the links which should point to not-yet-translated pages point to some other language, already translated version or to pages that held a "work in progress" message. In most cases this results in bewildering. This issue of 'dead links' (in web-sites) can be resolved using 2 techniques quickly reported hereafter.

1) If available and possible, it should be the multi-language content management system to control generation of the links. To put it simply, it is the multi-language CMS, which will automatically put a link or a button on a Web page ONLY WHEN the page to which the link/button points to translated and published is available. This way, the Web site will never contain any 'dead' links. Using the above mentioned technique, one could manage multi-language Web site simply by writing the content in the 'base' language, mark pages (but not necessarily all of them) for translation and let the multi-language content management system handle the translation process and publish the translated text automatically as they becomes available. The drawback of using exclusively the above approach is that the Web site has some structure and the 'foreign' language version will not be useful if only random pages get translated.

2) There may be also a combined solution: first, the Web site should be designed in such a way as to allow partial content translation - with reduced or simplified structure but maintaining the consistency. Practice shows that it is rather easy to design. Second, the content management system can track the logical dependencies between pages and only allow partial publishing of the foreign language version when the resulting partial version will maintain a consistent structure.

It would be simple to argue that all depends on what one would like to achieve, and therefore demand the multi-lingual content handling to a case-by-case situation. For instance, Wikipedia (www.wikipedia.org) consider each language as independent but every topic may link to the same topic in other languages. Yet their approach rises two main issues: on a side the open community supporting it and the freedom of contribution have been the driving engine and helped the system to develop and grow to a stage where its relevance could be no more ignored (becoming a de facto reference), but on the other side there are strong controversies on the quality of the content especially in certain languages. This may be a serious problem as despite the indubitable interest that the adopted solution and in any case it is not a suitable format for any publishing group that is operating on a commercial basis.

As a totally different example we can take into account newspapers, they may be available in several countries, each country have its own content, as the news have to be targeted in visibility and format to the audience of the targeted country. As a matter of fact, it is very important to make a difference between language and culture, this is basically meaning that the content for newspapers should be multi-cultural, which is even more specific than multi-language, as the same article that might have 5 paragraphs and a picture in a country, and only one paragraph in another country in order to meet the local sensibility and relevance of the topic to the audience.

Taking into account this latter example, the kind of issue we are dealing with implies that the distributed editorial supporting tool has to be able to support multi-char-sets, different kind of indexing and retrieval and, what is more tricky, enable people located in different part of the world to share content and tools seamlessly. For example *The Herald Tribune* or *The Financial Times* have a very well defined "*image*" and this is conveyed in every country they are available regardless the local language (when they have a local edition). This is achieved thanks to a combination of factors: a shared publishing platform (and related tools), a decentralized editorial board with centralized publishing policies, distributed archives, and many other supportive tools and solutions. Moreover it is often possible to have both a paper and a e-version of the same "newspaper" and this adds a further level of complexity to the overall as far as multi-language is concerned as it requires to take into account of the potentially ubiquitous access that an e-based approach provides to content with all connected (and already mentioned) issues.

3.1.4 The distributed human resources & cost issue

These are too major issues in the process as often companies (especially SME) do not have an adequate resource level to afford and sustain in the long run the production of multi-language content unless they are providing translation services. Therefore technology supporting translation and multilingual content management is crucial for the small and medium enterprise in order to face all the challenges brought forth by globalization. What just stated may seem awkward but in reality is a matter of facts. Often automatic translation tools are used to achieve a fairly acceptable translated version of content to be used for e-based content in sectors that are traditionally not e-oriented, but this has been for a quite long period, even if less and less often, the initial transitory phase to achieve a basic result: have at hand something to address a specific interesting market. Companies will then move towards proper and more traditional translation services and they would grant the desired quality.

4 Solutions and approaches

Usually whenever there is content that needs to be localized is usual, at professional level, to go for proper translation and localization process. This is usually a quite complex and structured process that often presents problems and unexpected issues. Therefore as a starting point we would like to see how the process id rolled out and how it could be possible to smoothen it.

4.1 Translation/Localization Step Zero

There are several questions that a good translator or agency will ask prior to start the work. These questions are ones that should be considered and have thought out responses before embarking on a translation pro-

ject. Moreover there are several channels to consider when having a translation completed, including, but not limited to, the difference between software driven translation and that conducted by real translators. This again may seem superfluous, but is not as there is more to be taken into account than the simple cost saving coming form doing "in-house" SW supported translations rather than following the whole localization process. As we are presently dealing with content aggregation, adaptation and delivery for both B2B and B2C market encompassing and respecting all constraints coming form proper DRM support. We state this as often it is noted that one barrier to the localization toward a new market of some content will be originated by clauses on the quality of the achieved result that the content owner may explicitly or implicitly place at the time of contract signature. Having said so let's start with examining available possibilities. There are many software driven translation services. Many of the common ones are available in a limited way online as a free service. However, they are anything but accurate. Some agencies/translators use software to support their real people translators. This involves the software developing a translation memory. A translation memory is a catalogue of recurring terms, phrases and sentences in the source document and their equivalent in the target language. The translation memory comes into play as the translator works through the project, allowing speedy insertion of stock standard items from the translation memory. This can be very useful for large-scale projects such as product operating manuals etc. but it can also lead to stilted, repetitive copy when used on other kind of contents and especially with marketing materials.

4.2   Translation/Localization Step One

If what just presented is a preliminary step, the first real step in the content localization process is the definition of the goal of the process. In other terms, from the client perspective the first step in any translation is to define the end goal. What just stated has a simple rational as translation outcomes may be different according to what does the client want the translated content to achieve. For example if there is the requirement for achieving effective and efficient marketing content that draws the reader in, leading further along a pathway to generate a conversion (an enquiry, a purchase etc.) the translator should use the proper expressive style and format suitable to target audience language, culture background (this in same case may require also special product design). On the other hand if the content is a product description that requires attention to detail, knowledge of technical terminology (jargon) in a specific industry and a sense of the underlying logic of how the end user will interpret the description or instructions the translator will have to strictly respect the inherent constraints. Therefore, as mentioned before, in order for the translator or translation agency to produce the desired outcome, it is needed to clearly explain what it is expected the translation to do. In what just stated is necessary to add also, not to neglect dialects. Within some languages there are distinct variations in grammar, syntax and vocabulary. Therefore if a specific geographical market in addressed, it will be necessary to specify this when looking for a translation.

4.3   Translation/Localization Step Two

Having defined the scope and target audience of the content/application to be translated/localized, it will be necessary to define the document type and volume. In other words what context is this translation to be used in, if it will be required to have the same content across multiple formats (html, pdf etc. etc.); it will also be necessary to verify if the access/distribution website is supported via a unique content management system (CMS) and if so, if this is multilingual capable, as it is very important that to know how you want to use the content before you send it out for translation. In particular website/html presentation can create difficulties. Some translators do not work within the source HTML code. Instead, they provide a translation in text format (e.g. a Word document) for the webmaster to cut and paste into the HTML. This process is not recommended, as it can create several issues:

- operator error (cutting & pasting into wrong location, dropping characters during the process, etc…);
- document incompatibility (when working across different versions of software, and operating systems, data can become corrupted or misinterpreted).

It is far more efficient and effective for translators to always work in the source code of the source document, whether that be a Word file, a website or an Illustrator EPS. Furthermore it will also be necessary to define the volume. Typically translations are quoted based on the number of words in the source document (or characters in the case of ideographic scripts such as Chinese or Japanese) and the document type. An accurate word count will enable to measure quotes against each other as well as give an estimate for time frame.

11

As a rough guide, a good translator, paying attention to detail and producing quality copy can translate 2000 words/day.
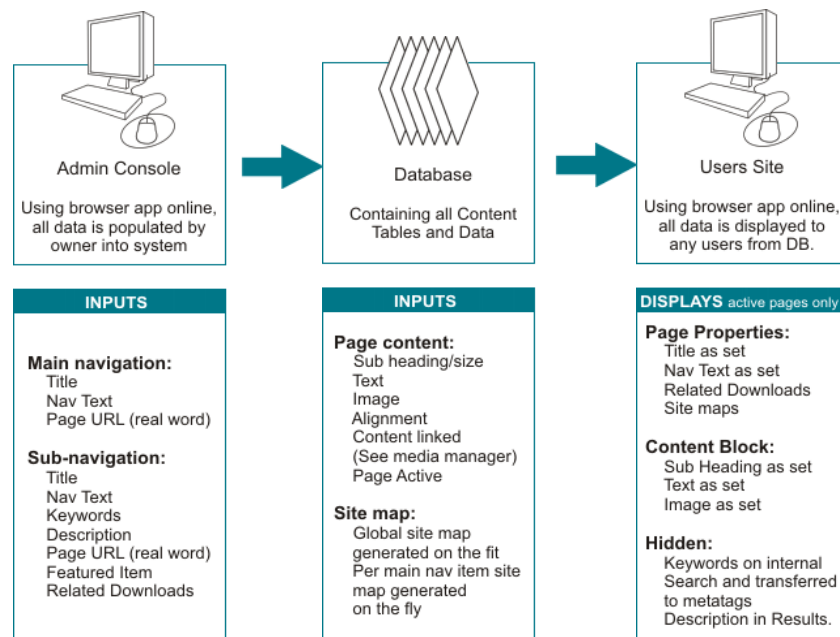
4.4   Translation/Localization Step Three

At this stage it will be necessary to identify the most appropriate translator. In other words, given a defined translation outcome, it now comes down to whose skills fit that translation outcome the best, typically there is one of three options:
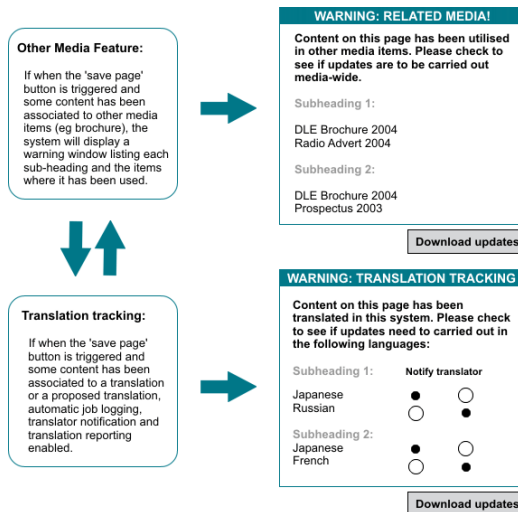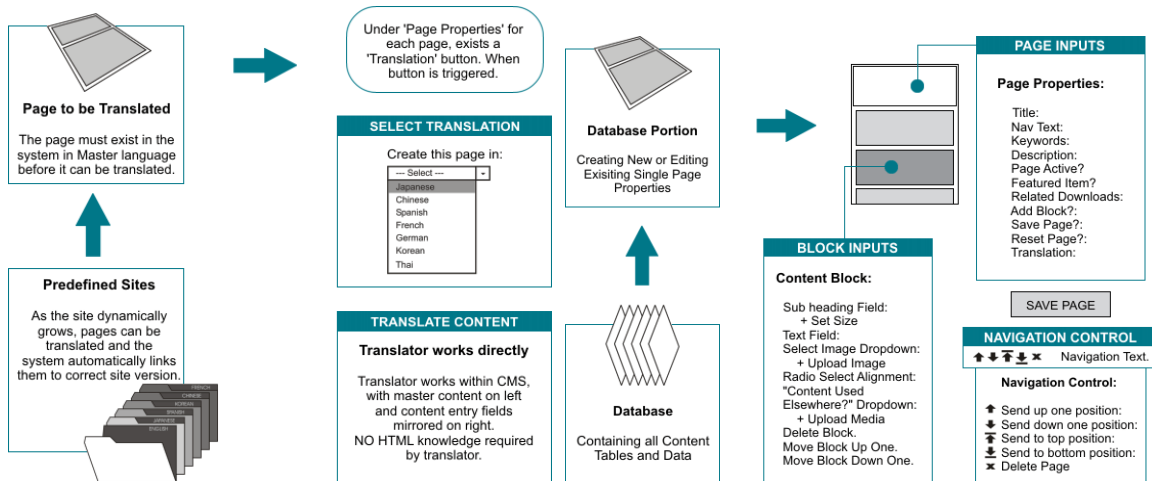
1. Use in-house staff. If there is in-house staff with the translation skills, it is possible to use them to do the translation. Of course, this has the benefit of not having to incur external expenses, but also has some embedded risks, nominally:
   - selected staff may not have enough high-level language skills as the one required,
   - staff member may not have the time to dedicate to the job, or may need to squeezed it in between all the regular, routine tasks resulting in lover than expected quality.
2. Use a professional translator. Preferable to Option 1 as should be able to satisfy all of the above questions, yet there is still the issue of how to find the translator, and verify that is up to the level.
3. Use a Translation agency, a one-stop shop for translation requirements. Agencies typically work across a range of languages and document formats, and verify the skills of the translator(s) and the quality of the output.

4.5   Translation/Localization Step Four

The final word in all this process is: Proofreading. We operate in the real world where, unfortunately, no one is perfect and mistakes do happen, yet is possible to minimize or eliminate some of them by ensuring that more than one set of eyes looks at the translation. Don't skimp on the proofreading (this is also why automated translation is often weak and error-prone). If you are working with an agency, always get them to specify that the translation will be proofread by an independent (i.e. NOT the original translator) professional. For marketing copy it should also be edited by a professional copywriter. If you are using freelance translators or in-house staff, ensure that at least one more person works through the translation.

In what follows is reported a set of schematics related to the abovementioned process and to solutions commonly adopted for supporting it (they will also be detailed later on) encompassing form simple DB-based solutions to multi-lingual content management systems (mCMS).

Page to be Translated

The page must exist in the system in Master language before it can be translated.

Predefined Sites

As the site dynamically grows, pages can be translated and the system automatically links them to correct site version.

Under 'Page Properties' for each page, exists a 'Translation' button. When button is triggered.

SELECT TRANSLATION

Create this page in:

--- Select ---
Japanese
Chinese
Spanish
French
German
Korean
Thai

TRANSLATE CONTENT

Translator works directly

Translator works within CMS, with master content on left and content entry fields mirrored on right.
NO HTML knowledge required by translator.

Database Portion

Creating New or Editing Exisiting Single Page Properties

Database

Containing all Content Tables and Data

BLOCK INPUTS

Content Block:

Sub heading Field:
+ Set Size
Text Field:
Select Image Dropdown:
+ Upload Image
Radio Select Alignment:
"Content Used Elsewhere?" Dropdown:
+ Upload Media
Delete Block.
Move Block Up One.
Move Block Down One.

PAGE INPUTS

Page Properties:

Title:
Nav Text:
Keywords:
Description:
Page Active?
Featured Item?
Related Downloads:
Add Block?:
Save Page?:
Reset Page?:
Translation:

SAVE PAGE

NAVIGATION CONTROL

Navigation Text.

Navigation Control:

Send up one position:
Send down one position:
Send to top position:
Send to bottom position:
Delete Page

Other Media Feature:

If when the 'save page' button is triggered and some content has been associated to other media items (eg brochure), the system will display a warning window listing each sub-heading and the items where it has been used.

WARNING: RELATED MEDIA!

Content on this page has been utilised in other media items. Please check to see if updates are to be carried out media-wide.

Subheading 1:

DLE Brochure 2004
Radio Advert 2004

Subheading 2:

DLE Brochure 2004
Prospectus 2003

Download updates

Translation tracking:

If when the 'save page' button is triggered and some content has been associated to a translation or a proposed translation, automatic job logging, translator notification and translation reporting enabled.

WARNING: TRANSLATION TRACKING

Content on this page has been translated in this system. Please check to see if updates need to carried out in the following languages:

Subheading 1:	Notify translator	
Japanese	●	○
Russian	○	●
Subheading 2:		
Japanese	●	○
French	○	●

Download updates

TRANSLATE

Japanese Site: Home \\Page Welcome

Title:	Welcome to the site!	ホームページへようこそ。
Nav Text:	Welcome	ようこそ
Keywords:	Home, Welcome, Education, Introduction, General Info	ホーム, ようこそ, 教育, 紹介, 一般 情報
Description:	This site has been set up with the intention of assisting you, the consumer find the best possible solution within the industry.	このサイトはその業界内で, 消費者が最適可能な解決方 法を見つけるための支援を 意図とし設立されています。

Save Translation

Subheading: The History of Us

Text: In order to be useful, the text must convey the same function or purpose as the image. For example, consider a text equivalent for a photographic image of the Earth as seen from outer space. If the purpose of the image is mostly that of decoration, then the text "Photograph of the Earth as seen from outer space" might fulfill the necessary function.

Subheading: Our Mission Statement

Text: Content developers must consider these different situations during page design. While there are several situations to consider, each accessible design choice generally benefits several disability groups at once and the Web community as a whole.

Save Translation

RELATED DOWNLOADS

Manage Downloads - Page 1
Current Downloads

Brochure1.pdf
Calendar1.pdf

Edit Delete

Add Download:

Select the "Browse" button below to choose an image on your local or network drive. Enter in a display name for the file and also a description, then select the "Upload" button to submit the file.

Display Name:

File Description:

Direct Download: Yes ◉ No ◉
New File: Browse

Upload
Done

IMAGES

Manage Images - Page 1
Current Images

photo_Jerry.jpg
corporateLogo.gif
advert_banner.gif
main_campus.jpg
winterphoto.jpg

Edit Delete

Add Image:

Select the "Browse" button below to choose an image on your local or network drive, then select the "Upload" button to submit the file.

Browse

Upload

Done

It is also necessary to examine a bit more in detail how the various kind of content are dealt with and which

are relevant inherent issues.

4.6   Text

Probably text is the widest category of content to be taken into account when dealing with multi-lingual issues as all other content will somehow include some test either as a specific part of the content or as metadata or as specific support to multi-lingual issue management (subtitling…). Language tags should be used for all text-based content. A simple example of the use of multilingual text and metadata is in the production of html, where keywords and page descriptions need to reflect the language of the page content. In these cases, for multilingual searchability and accessibility, the best user experience is achieved by declaring the primary language of each page using the <html lang="nn"> tag, where nn is an international language code defined under ISO-639 ('it' for Italian, 'en' for English, 'fr' for French, etc.). Once a page's language is correctly identified, keywords and descriptions will then be indexed in the relevant primary language. Multiple language versions of the same content are best presented as separate pages, each with an appropriate language identifier tag and with all keywords and descriptions as well as content translated. This ensures that the result will be returned in the same language as the query. The <link> tag of the header enables to link a version in a language to alternate versions in other language, as show in the example from http://www.w3.org/TR/html4/struct/links.html#h-12.3.3 presented below:

```
<HEAD>
<TITLE>The manual in English</TITLE>
<LINK title="The manual in Dutch"
      type="text/html"
      rel="alternate"
      hreflang="nl"
      href="http://someplace.com/manual/dutch.html">
<LINK title="The manual in Portuguese"
      type="text/html"
      rel="alternate"
      hreflang="pt"
      href="http://someplace.com/manual/portuguese.html">
<LINK title="The manual in Arabic"
      type="text/html"
      rel="alternate"
      charset="ISO-8859-6"
      hreflang="ar"
      href="http://someplace.com/manual/arabic.html">
<LINK lang="fr" title="La documentation en Fran&ccedil;ais"
      type="text/html"
      rel="alternate"
      hreflang="fr"
      href="http://someplace.com/manual/french.html">
</HEAD>
```

This standard approach can be generalised to all XML documents, using the 'lang=' metatag to describe the language of the content. Also to support text readers and search engines, secondary languages used within text should be encapsulated within a tag using 'lang=nn'.

4.7   Images

Apart from the textual content that may appear on an image (an image could also be: a sales trend graphic, a market share pie representation…), images always present a relevant set of textual accompanying information. Some of this information is also embedded inside the image file format (JPEG header…). What we are referring here is the heterogeneous set of language related aspects that are related to the specified kind of object. For example when translating a composite document, such as a poster, the disposition of the text is really important. As a translated text is not necessarily as long as the original, it might be needed to adapt the layout or the styles, such as increasing or decreasing a font size. A similar problem exists with maps, where the tag over a location might need to be adjusted as a change in the width of the text can make two labels to overlap. These are just examples of content that are usually composed basically by images or transformed in images either for print or for usage as digital objects. What just stated has a even more relevant impact onto products where in the graphical user interface (GUI) there are graphic components holding text that is language dependent, like buttons, etc. this is particularly relevant in the case of packed interfaces (for example PDAs or mobiles) where even a small change in dimensions of a specific object may alter the overall balanced aspect of the GUI. This applies more and more frequently also to web-based applications where is

becoming quite common to have GIF objects used to achieve a better overall graphical result for the interface. To these problems there is no unique solution, what is advisable is to take into account such aspects at design time and organise graphical components of images / GUI to accommodate the most probable set of languages that may be addressed at distribution time (whenever possible).

4.8   Audio / Video

We have already mentioned how relevant it is for this kind of content the linguistic issue and how subtle it could be the difference between a good quality and a poor quality multi-lingual management for this kind of content. Nevertheless it is important to see how this issue has been, so far, tackled. At this point is essential to place a clear distinction between audio and video content. As far as audio is concerned we have to take into account the following issue: a song either is available in a specific language or is not. For example there used to be the so-called "*cover*" versions of songs having had a relevant success. This phenomenon was quite common during the '50s and '60s and still applies to some countries (Italy, Spain, South America, China…) both with a localised version of the song or with a re-made one. Just to give some examples of the various possibilities is interesting to take into account that during their long staying in Hamburg, The Beatles, made a German version of Michelle, or that the Italian group Dick Dick made an Italian version of California Dreaming (Sognando California), or singers like Nec, Lura Pausini, Eros Ramazzotti, Rita Pavone, Raffaella Carra and others that have made Spanish versions of their best products. On the opposite side there is "*opera*" where usually no matter where it is performed, it is always performed in the original language (Italian, French or German) even if the public can access to both original and translated version of the text. As far as re-make of songs is concerned is sufficient to think to "Summer in the city" that has been sung at least from Lovin' Spoonfool, Joe Cocker, Joe Jackson each with his own stile and finally by Nelly in an Hip Hop version. As far as video is concerned the issue is far more complex. In this case there are a set of possible solutions, spanning from dubbing up to subtitling. One could try to limit the burden to provide several audio tracks to allow the user to select the desired one, this would be a very coarse solution as in case of absence of synchronisation the result may be quite distant from the user expectation, on in some cases even disturbing or annoying. A totally different issue is the one related to the metadata associated to this kind of content as in this case it does not matter whether the content is audio or visual, it is just a matter of which languages are supported for the metadata management and how the info are stored. In this latter case (metadata) all that has been said for text and/or metadata applies.

4.9   Animations & Multimedia

This kind of content is basically a combination of the previously mentioned ones and therefore presents the combination of all problems reported so far. At the same time this kind of content has an advantage in respect to the previously mentioned ones, namely they can benefit of all achievement of synthetic graphics that span from the possibility to automatically tune lips-sync up to enable things that are unfeasible to humans. Moreover the implicit digital nature of the product, further enhance the chances to achieve real and effective multi-lingual support.

4.10 User Interfaces and applications

In what follows we will tackle the issue of multi-language support for User interfaces and applications. Providing examples of available solutions, ranging from the simpler one up to programming based ones. The case of news or articles is quite relevant as it applies to several environments and has as main output a web-oriented delivery that can profitably also accommodate a web-based input system. In the present section is reported a simple case of an application that has been developed and delivered within SEN-IST-NET project to manage news and content in 14 languages having as main front-end people content experts with no programming experience. For carrying out the translation job in a simple and effective way, it was devised a simple database based hosting the original master (English) version of content and all translations. To make process management simple and allow easy access to people involved in translation user IDs have been mapped onto language IDs. A set of macros would handle all issues related to temporal issues (newsletter or article publishing date and validity). As already mentioned it was identified English as the master language. All sources for the translation process would therefore be in English and the same would apply to contributions, news etc. Once registered onto the DB the user can start to operate. The interface devised was meant to provide basic functionality. Hereafter, are presented main implemented functionality and related pages.

15

For the sake of simplicity the users for the translation process are identified with the international code for the related language (**en** for English, **fr** for French, **es** for Spanish, etc.). It was established a special user (**Manager**) whose language code was the same used for English. This user would insert original data (the English source) using the master language and create the master records.

In the previous picture is presented the Home-page of the System Manager / English author or translator, where **Title** is the title of a text portion while **Subject** is the subject of the same text portion. The **Detail** button will cause a window to open where it is possible to examine text detail (shown later on) while the **Update** button will cause the opening of a window where it's possible to modify the previously inserted text. Finally there is an **Insert** button, which allows inserting a new text into the database. Insertion is allowed only for English text while for all other languages it is expected only to have either a translation (the 1st step in the process) or an update if the translated text needs modification. In the following image is shown the input interface for a generic text insertion. We have devised that each text shall present the following characteristics:

- **Title -** is the text title. It can be omitted if not necessary or it can be the only field in case it represents a section title.
- **Subtitle -** is the text subtitle. It can be omitted if not necessary or it can be the only field in case it represents a sub-section title.
- **Body -** it's the text body. The core information but can be omitted in case the inserted text has to be managed as a section / sub-section title. It can be left without title and / or subtitle in case it has to be treated as a plain paragraph
- **Subject -** this is a service field used to provide a support information to the user when accessing the main page as this description will provide information on the displayed record
- **Image description -** is the image caption and has to be loaded only if there is an image. It is possible that an image does not have a title/sub-title or associated text. If this field is present the next should hold a URL address.
- **Image -** is the URL address of the image referred by the caption.

In the following picture is presented the overall interface aspect in case the Detail button is pressed and the detail window has been opened. It is worth mentioning that the system Manager and the English translator will be able to see all inserted record one after the other (in this version) while a generic translator will always see English text on the left and the translated one on the right.

16
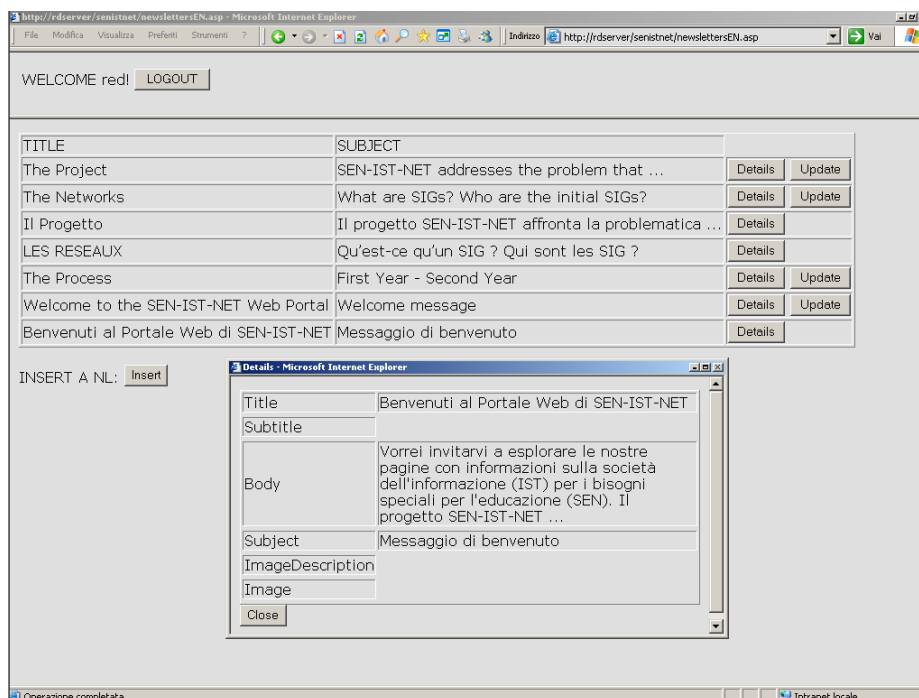
The update window has on the left side there is the original text, shaded as cannot be modified, and on the right the currently imputed text (the same of the left in case the work has not started yet).
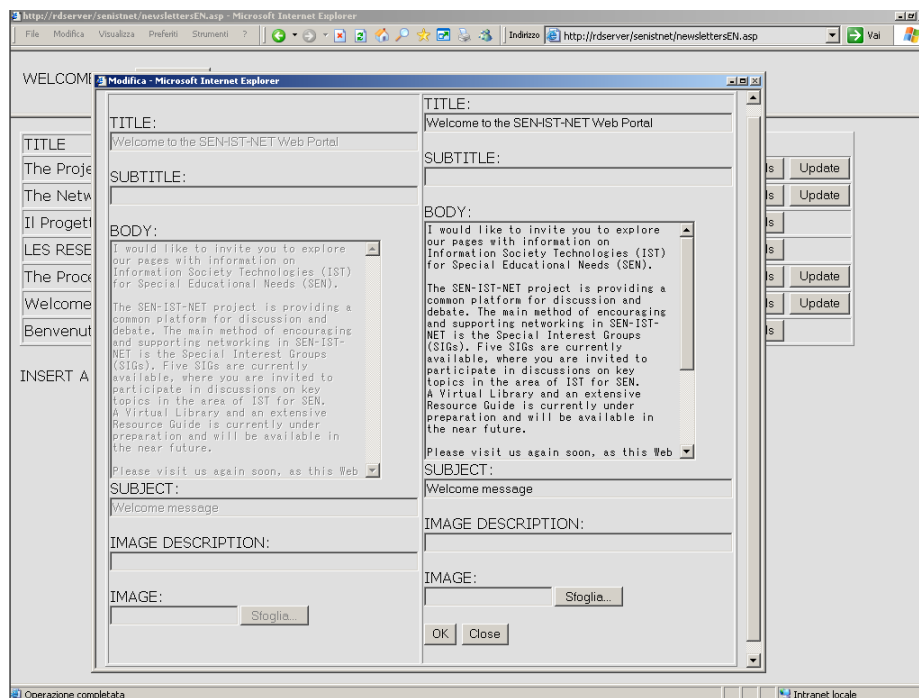
In the following picture is visible the main page for a generic translator.

In the following picture is evident the aspect of the generic item translation window, where on the left there is the original text and on the right the translated one.

What has been prepared was thought for an extremely simple usage requiring the minimum effort and giving the maximum flexibility at the same time. Therefore all it's required to install is summarised in the following list of working steps:

1) Create in the site reserved area a section/folder devoted to translation process where all involved people will be able to access once they log in.
2) Copy in such area the content provided along with such file (nominally all the *"asp"* pages and the directory holding the DB table file).
3) Prepare all that may be needed by the system administrator to retrieve data and publish it.
4) Connect the new pages to the web portal via a link in the restricted area

Due to the implemented structure for a generic item record it was suggested to manage data insertion following the rules sketched here after:

1) Records should be inserted initially in English following the structure that the system manager intends to give to the newsletter.
2) Main titles and section titles should be inserted as text without **sub-title, body …**
3) Secondary titles and sub-section titles should be inserted **title, body …**

18

4) Paragraphs should be inserted with all fields (the only optional one should be those related to images).
5) It would be better to avoid having records holding only bodies or pictures.
6) The translation process should be initiated only when all English records have been inserted.

4.11 Multi-language management and programming

In this paragraph we will just provide a tiny example of how the issue is tackled in the programming environment, where practices and policies are so different form place to place and even affected by application domain. The best example is the one of ERPs where BIPCS used to be the most prominent product (on the US market) but could not stand the comparison with SAP in Europe. This was basically due to a factor: despite BIPCS was localised in terms of linguistic support, it was not up to the level in terms of complexity management. What is intended here is that in Europe (for example in Italy) the complexity of the legal and administrative scenario was at least an order of magnitude more than in the US. Companies like SAP being based in Europe and having to deal with intra/extra border issues for goods exchange, different currencies and taxation systems even within the same country (e.g. old and new shillings or east/west deutsche mark…) were more suitable to accommodate for a wider and deeper set of customisation (at in this case localisation would basically comprise a good set of customisation). In other environments this was not the case (think to MatLab) and therefore localisation would be limited to the GUI. This process has to be carefully planned in any case to avoid the risk of translating also "keywords" or other components of the programming environment potentially leading to incompatibility among products of the same family but representing localised instances; just look to what happened in a certain period of time with Excel where functions names were localised along with the interface leading to the strange situation where the function SUM() of the USA/UK version would become SOMMA() in the Italian one, with all consequent problems of interoperability.

4.11.1 wxWidgets

In order to support multi lingual for the software designed, internationalization in *wxWidgets* has been investigated since *wxWidgets* is being utilised for various key components of the framework. A separate program called *gettext* needs to be installed in order to generate the language catalogues. A GUI wrapper for *gettext* called *poEdit* is available if preferred. The process is as follows:

- All translatable strings should be wrapped in a _() macro, e.g. *wxString s = _("this is a string")*;
- All menu items, labels, dialog messages, etc. should be converted.
- All non-translatable strings should be wrapped in a _T() macro so that character width conversion can be handled appropriately depending on whether Unicode is used on the system.
- *gettext* should be run on your source to generate a *.po* file containing all the translatable strings.
- These strings need manually translating.
- Finally, a catalogue file (*.mo*) can be generated for each supported language.
- The Locale can be set in the system initialization method to the system locale using *wxLANGUAGE_DEFAULT* and the correct translation of the strings will be used where available.

What just stated is based on the fact that *wxWidgets* uses a certain number of user-readable strings such as "*help*" or "*Load file*" which should be translated to the users language if it is different from English. *wxWidgets* has built in support for internationalization (*i18n*) which allows for this to happen automatically if the translations to the current language are available. Hereafter is reported the list of all existing translations. The columns of this table have an obvious meaning: in each row is reported the language and the status of the translations as of current CVS version of *wxWidgets* (regenerated once a day).

Language	Translated	Fuzzy	Un-translated	Graph
Afrikaans (af)	75.9 %	12.8 %	11.2 %	
Catalan (ca)	65.1 %	19.0 %	15.8 %	
Czech (cs)	68.0 %	15.1 %	16.9 %	
Danish (da)	73.0 %	18.7 %	8.3 %	
German (de)	84.4 %	8.4 %	7.3 %	
Greek (el)	88.2 %	5.8 %	6.0 %	
Spanish (es)	89.7 %	5.5 %	4.9 %	

Language	Translated	Fuzzy	Un-translated	Graph
Basque (eu)	14.0 %	6.8 %	79.2 %	
Finnish (fi)	71.1 %	14.6 %	14.4 %	
French (fr)	100.0 %	0.0 %	0.0 %	
Hindi (hi)	75.5 %	13.8 %	10.7 %	
Hungarian (hu)	92.2 %	4.6 %	3.2 %	
Indonesian (id)	76.9 %	11.8 %	11.3 %	
Italian (it)	100.0 %	0.0 %	0.0 %	
Japanese (ja)	89.1 %	6.0 %	4.9 %	
Latvian (lv)	19.1 %	6.0 %	74.8 %	
Norwegian Bokmal (nb)	92.2 %	4.6 %	3.2 %	
Dutch (nl)	64.5 %	19.4 %	16.1 %	
Polish (pl)	100.0 %	0.0 %	0.0 %	
Portuguese (pt_BR)	87.7 %	11.9 %	0.4 %	
Russian (ru)	92.2 %	4.6 %	3.2 %	
Slovenian (sl)	100.0 %	0.0 %	0.0 %	
Swedish (sv)	100.0 %	0.0 %	0.0 %	
Turkish (tr)	92.2 %	4.6 %	3.2 %	
Ukrainian (uk)	58.2 %	24.9 %	16.9 %	
Chinese (zh_CN)	89.2 %	7.1 %	3.7 %	
Chinese (zh_TW)	74.8 %	13.2 %	11.9 %	

As already stated *wxWidgets* uses the standard GNU *gettext* tools for *i18n* therefore it is possible to contribute to work completion. Here are the steps that should be followed to get the most current version and contribute:

1. Get the latest version of the file locale/wxstd.po from the wxWidgets source tree; otherwise retrieve it directly from the cvs repository via the Web interface.
2. Rename it to ll_CC.po where "ll" is the 2 letter ISO 639-1 language code for your language and "CC" is the 2 letter ISO 3166 country code. If the country is the default one for this language (e.g. Italy is default country for Italian), then the country part and underscore preceding it should be omitted -- this is, in fact, the most common case.
3. Translate the strings in this file using either your favourite text editor or a specialized tool.
4. Verify that your translations can at least be compiled (even if they are yet incomplete) by running msgfmt -v ll_CC.po command: please note that you *must* use the -v option. In particular, please fill the header fields because msgfmt doesn't accept the default values for them.
5. Send the finished translation that will be added to the next wxWidgets release or snapshot.

4.11.2 MFC Microsoft Foundation Class

The main problem of using dictionary to translate the user interface is that it lacks flexibility, because sometimes the form should be completely reorganized to meet the target language constrains. This is why, in the MFC, Microsoft used "resource DLL" to handle localization. A resource DLL only contains the resource of an application, like string tables, error messages, but also the complete layout of a dialog box. Only the resource of the language of the user is loaded in memory and, as the identifiers of the resource are language independent, the application just have, for instance, to load the dialog with the id IDD_ABOUTBOX, and the English version will be loaded for the English, French version for the French..

4.11.3 Microsoft .Net

In .Net, this approach as evolved a little. First of all, there is a fallback mechanism, so that, for instance, if the resource for de-CH is not found, then the system search in the de resources, and if it does not found it either, it searches in the neutral language. Also, when you design a form, you can design it for several cultures. In the designer of the forms as also evolved so that some properties can be localized, but not all of them. The form presented is designed in the neutral language as apparent from the first picture. When translating it in

20

another language, the picture may change, the labels might be widen and reposition. The accelerator keys also changed.
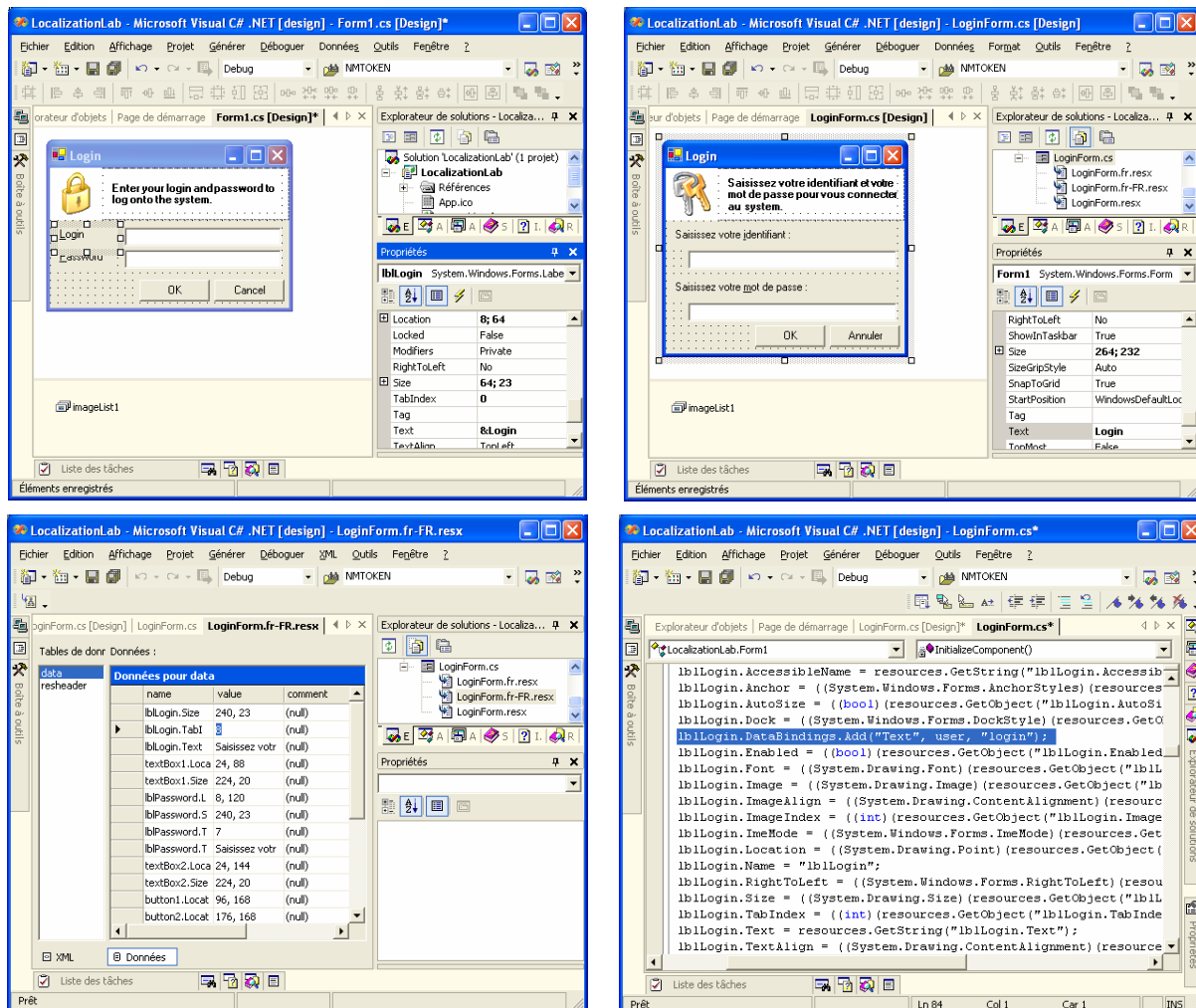
In the solution explorer panel, you see that the LoginForm has several resource files, one for the neutral language, one for the language (fr) and also, if needed, one for a sublanguage (fr-FR). The resource file contains values for the all the localizable properties of the form and the controls. But, as you may see below, not all the properties are localizable. For instance, the DataBinding property, used to bind a control and the property of an object, is not used for layout but for managing the form. This approach to handle multilingual user interface is very flexible because it is possible to redesign the form, an issue that can be important when dealing with right to left text or top to bottom text. But the drawback is that it may make the translation process more difficult if the editor is not good enough to leverage the difficulty.

4.12 Multilingual Content Management Systems (m-CMS)

To most of the previously mentioned issues (both in terms of content production and support tools) is worth taking into account the most complete solution that covers almost all the process but has also a non marginal drawback, namely the cost. The issue is usually raised in international organizations where multilingual marketing is part of global marketing needs (that is marketing across regions, cultures and languages) or distributed production and support causing different teams and work-groups to operate together on a 7 days a week 24 hours a day operation coverage. On a side this is an organizational challenge on the other side it is an opportunity. The potential that the internet brings means that it is possible to effectively communicate with anyone, but at the same time for this to happen, it is needed to communicate in the first or native language of the counterpart. As Willy Brandt, the former German chancellor, put it: *"If I'm selling to you, I speak your language. If I'm buying, dann muessen Sie Deutsch sprechen [then you must speak German] "*. The issues

21

that arise when developing and deploying multilingual solutions is not a technologically challenging exercise, but it does pose unique coordination issues, especially when working in multiple languages across multiple media. Here we will discuss some of the issues surrounding multilingual content management and how they can be solved within content management system. Each content management system needs to be customized to the client's individual needs and specifications, as on the web, unfortunately, there is NO "one size fits all"! The starting point is that to deploy contents in a timely and cost effective manner, across multiple languages and multiple media, whilst maintaining consistency of content across languages and media is needed to:

- draw B2C business direct from consumer,
- recruit distributors/resellers globally,
- support distributors/resellers by providing customized content for their language,
- maintain control of brand and market image and ensure consistency of brand image globally.

On the other side there are several issues to be tacked, starting from the cost of translation into multiple languages. Translation costs rise as specialist skills are required (e.g. HTML knowledge), and at the same time it is problematic to use in-house staff if they do not have the required skills. Where the same content is used across multiple media there is a waste of resources if it is re-translated instead of recycled. This can be tacked as follows, first some goals are set and then a set of solutions are adopted to achieve them, namely:

Goal
- Reduce cost by removing skill set required of translator
- Recycle content translations across multiple media (web, print. etc.)

How
- A master language is set in the CMS. All screens can then be translated within the CMS from anywhere, by anyone with internet access, with no special HTML knowledge. Where unique content is required for specific language/market(s), a master in the default language is created first, guaranteeing a central, monolingual master catalogue of all materials. This default language version can be disabled to prevent viewing from outside the administration system, presenting only the translated unique content to the target market.

But this is not enough as consistency of content across all media and languages should be granted. As a matter of fact products, contents and marketing media requires frequent updating. When working across multiple languages and media it is a coordination challenge to ensure that all media remain consistent with one another, and updates are applied universally and in a timely fashion.

Goal
- To provide a system whereby multiple media releases can be linked, so that when one is updated/altered the administrator is alerted where same/similar content is to be found and updated.

How
- Set the website as the master marketing platform; this means web is always updated first as procedure. Where content is re-used in alternative media, link specific blocks of content to specific media. As web is updated, administrator is alerted to where same/similar content exists. Updates to other media can be applied immediately or catalogued for batch updating at a later date.

There is undoubtedly a cost of updating content (especially web-based) and this has to be carefully taken into account. Need to maintain, update and expand content (especially web-based one) is always there, but cost of programmers restricts localization of new content, impacting on time to market. Maintaining content means task is performed at least one step removed from source e.g. marketing manager writes copy, delivers to web developer for loading, creating choke points in contents/products release.

Goal
- Website should be no more difficult to update than a Word document
- Content creator enabled to be content loader

How
- Create CMS that operates very much like a word processor, with NO HTML knowledge required. Allow content creation source to edit website directly.

Web promotion/Search Engine Optimization is a must. Web marketing usually requires some level of SEO. This can be performed in-house or contracted out.

Goal

22

- Fully optimizable website.

How
- Manually editable title, description and keywords meta statements
- Manually editable URLs (insert keywords to URL)
- Create multiple landing pages for same product/service (especially relevant to PPC campaigns)

Hosting/maintenance can be far from trivial as different organizations have different capabilities; therefore cost effectiveness and efficiency of internal hosting depends on organization size and IT strengths.

Goal
- Flexible hosting solutions on offer: hosted solution, in-house solution, contracted solution to give client maximum flexibility

How
- CMS is available as a hosted service, as a stand-alone application on the client's server, or as hosted by a server of the client's choice.

In most of the case CMS are used to support company content production for marketing and customer support services. This is certainly not the only usage for CMS, but I soften the more frequent and therefore it is worth focusing onto them as other application (where multi-lingual support may be needed) can be adopting solution originated from the ones adopted here. The question at this point could be: what to look for in a multilingual content management system?

As already stated previously M-CMS (multilingual CMS) creation and deployment is not a technologically challenging exercise in and of itself. But it does pose unique coordination issues, especially when trying to maintain consistency of content across multiple media. The crucial points are summarized hereafter:

4.12.1 Time to market

In commerce and in particular in an online world, time is crucial (both in terms of time to market and in terms of up-time). Web users expect immediate results (it is in the nature of the end-user). Traditional web deployment had three key stages: the content developer (e.g. marketing manager) writes a product description update. This is then passed to the webmaster for updating the website in the master language. Once the master update has been applied and checked, the content is then translated into the target languages, passed back to the webmaster, and then the foreign language content is updated. This creates four potential choke points, where communication between parties can break down or conflicting priorities and sheer workload can lead to delays:

- Development point (marketing manager)
- Deployment point (e.g. webmaster)
- Localization point (e.g. translator)
- Deployment point again for localized content

In a multilingual situation, an effective CMS should allow the Content developer to deploy the master set of content. In other words the CMS should require no specialized HTML knowledge. The developer should be able to enter content directly into the CMS without assistance from the webmaster. Likewise, when moving from deployment to localization, the localizer/translator should not require any specific HTML knowledge. He/she should also be able to work directly in the CMS. So it is possible to move from four potential choke points to two or even one. The key here is to have a CMS that allows direct, agile access to the person that writes the content, in whatever language it needs to be deployed; at the same time, access needs to be limited to the areas the particular operator requires to complete the job at hand. The use of such systems may reduce time to market (TTM) of online content by over 50%.

4.12.2 Operator error reduction & Translation cost minimization

Every time data changes hands, there is the chance of an error occurring. Different platforms, different application versions and plain operator error are significant problems that need to be reduced if not eliminated. Using a CMS that reduces the number of people that have to handle content reduces the chances of errors creeping in. The cost savings of direct CMS access cannot be ignored. The ongoing requirement for a webmaster for updates/maintenance can be eliminated (aside from structural issues, the site should run without webmaster input). At the same time the technical knowledge requirement of the translator is reduced. This can be of benefit in two ways; outsourced translation expenses can be reduced and, where available, you can make use of in-house language skills so that non-critical edits/updates to content can be resolved without recourse to external translation services.

4.12.3 Coordinating between web and other marketing media

Perhaps the next biggest issue is one of coordination and "logistics" as when working across multiple languages, coordinating multiple media updates can be problematic, and inconsistency of content across media and languages is often a common issue. For example, an international education provider has multilingual marketing collateral in the form of web, printed brochures, DVD, PowerPoint presentations etc. that are distributed through resellers globally. The provider needs to release updates to course outlines; web content is updated and localized via the CMS, but there is still the issue of what has to be done with the other collateral. A well-designed M-CMS should provide a method for the linking of content between media and languages. One structure is to have within the M-CMS a media catalogue, where a list of all media is kept. It is then a simple matter of linking content from the M-CMS to content in other media. Whenever an update is applied to web content, the relevant coordinators can then be alerted automatically and make the decision as to whether the other collateral needs updating. All materials can be kept consistent, or updates can be collated for bulk application within a cycle. The important thing is that everyone within the cycle may know what changes have been made to what content and in what languages.

Ideally the media management system described above should be done at the lowest level (individual content blocks, rather than at the screen or section level) as there is a cost implication here again. If you can link individual blocks of content to individual media items, you can recycle translations.

4.12.4 Website localization and customization

As already mentioned true localization requires customized presentation of content tailored for a particular market. For example it is worth considering if the colors of a German site would work for the Thai version or if it would be needed to change the whole look and feel. The same issue was addressed also earlier on, yet it is worth recalling here that the lead article/content on a French news site could not be appropriate for the Vietnamese site of the same press publisher, but at this point there would be the issue of whether the person in charge of the content can change it or not. For some organizations this may be beyond their budget and capabilities, but still should still consider this in CMS purchasing decisions, as there are options available. A true multilingual CMS will allow presenting a unique image for a specific market (something it is called "re-skinning". Therefore, whenever possible it would be far better to look for a CMS that allows presenting a fully localized (i.e. unique look and feel) web presence for each target market.

Whatever CMS is selected it should work on Mac, Linux and Windows and be cross-browser compliant (v4 or better, although v5 is acceptable, in IE, Netscape, Safari, Mozilla, Opera, Firefox etc.) as it is no more difficult to code for cross platform than mono-platform. Actually there is nothing worse than looking to localize for Japan and finding out that the CMS does not display well (or worse doesn't work) on the target delivery platform. Moreover it is crucial to remember that it is always needed to have good search engine optimization capabilities. Key factors to consider here are: full manual editing of all meta-statements, on a page-by-page, language-by-language basis. Therefore the body content of every page should be index able by the main search engines. All navigation text (i.e. the links to the various pages within the site) should be text, not graphics, and should be editable.

If the website is part of a commercial or marketing strategy, or even just in case of a multi-national structuring of company business, it is usually needed to have supportive statistics. One of the key data sets in need to be tracked is search key phrases; these statistics need to be multilingual capable. Only a few statistics packages can handle double-byte characters, as a matter of facts, many statistics packages just treat all non-Roman characters as Roman characters and produce unintelligible results.

4.12.5 Multilingual CMS hosting

Paraphrasing Hamlet one could say: *To host or to be hosted?* As a matter of fact it may seem a trivial issue (as one could think it affects primarily economical aspects) but actually it is not. To achieve best possible results (including those related to multi-language management) it is necessary to look for hosting options that suit company's current scale of operations, but have the capability to expand in the future. Vendor hosting generally comes with some significant benefits. Most hosted CMS solutions will provide ongoing updates to ensure the system is current; also, as server configurations change and are updated, the host is responsible for testing and ensuring CMS compliance with server updates. A good-hosted solution will also provide back up of all your data (although you should still back it up yourself as well) and 24hr service. One issue to consider seriously with hosted solutions is database integrity. Some hosted solutions have all client content stored in a single database; it is preferable if the content is stored in separate independent databases to ensure that there is no chance of data corruption.

24

5 Media & multi-language support (XIM, OD2, ILABS, SEJER, ALL)

When dealing with multi-language support, it is interesting to allow the user to define an ordered list of preferred languages, so to ensure that it will be possible to fallback to a secondary language when some content would not available in the favourite one. For instance in case the user prefers reading in French, but does not mind read in English, such two languages could be selected as primary and secondary preferred for content and/or GUI. This could therefore provide the system with useful information to exploit whenever content and/or components of a site are missing in some specific language while is available in other (it is worth taking into account that this is a good solution especially whenever an on-line approach is followed). What just stated is relevant also in light of the fact that it is not only useless but also annoying to show the user content in Chinese in case this language is not understood or known. The same applies to audio-visual content and, of course, to multimedia. From what just stated is apparent that the most relevant aspect to focus our attention on in terms of media and multi-language support is concerned is the design level. It has already been shown how to deal with mCMS and what it is actually under the translation/localization process, therefore here it will not be necessary to re-instantiate what already presented, yet it is worth mentioning that in the media environment the multi-lingual issue has been tackled since a long time. In the movie and TV industries is usual to adopt either dubbing or subtitling policies for content to be delivered abroad.

6 Authoring, Packaging & multi-language support (ILABS, SEJER, XIM)

In this section we will focus our attention on how the media market usually deals with the multi-lingual issue mainly at authoring and production level.

6.1 What is presently achievable in multimedia e-learning

In the content production process, what is of the utmost relevance is the possibility to deliver/package content that is multi-language. This can be achieved regardless of the original authoring and packaging environment at least up to when the selected/used environment is able to deal with multi-language content. This basically can be translated as follows: GUI is in English or any other language the user cha use, the system supports UNICODE or any other suitable multi-byte oriented encoding suitable for the intended delivery. As already mentioned there are standardised ways to represent and handle language encoding related issues, but this is just an aspect of the whole issue; as multi-language support is much more than this, as matter of fact one of the most advanced field in terms of multi-language support is the one for eLearning where this issue has been tackled since long time, also in relation to the inherently multi-language nature of teaching and learning. The following in an example of XML code representing what presently possible and implemented in authoring and production level for eLearning environment following IMS standard and that could represent a good starting point for future developments in the field of multi language management at authoring and production level.

```xml
<?xml version="1.0" encoding="utf-8"?>
<!-- XML file generated by eXact Packager -->
<!-- eXact Packager licensed to: monbile2.giuntilabs.com -->
<LOMobile xp:deliverytype="SCORM" xlink:label="L5414" xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:xp="http://www.giuntilabs.com/exact/xp_v1d0" href="LOMobile.html" xp:description="L06_Madonna_pomegranate" Tracking="AICC-
HACP" id="L06_Madonna_pomegranate">
    <style xmlns="http://www.giuntilabs.com/exact/xp_v1d0">
        <colors c1="#000000" c2="#FFFFFF" c3="#D4D4D6" c4="#005FAE" c5="#003060" c6="#01D4FA"/>
        <linkcolors normal="#0000FF" visited="#800080" active="#FF0000"/>
        <tracking>AICC-HACP</tracking>
        <pagesize width="760" height="640" type="1"/>
        <textstyles>
            <textstyle type="Title">
                <normal fontface="Arial" fontsize="28" color="#000000" bold="false" italic="false" underline="false"/>
                <emph fontface="Arial" fontsize="28" color="#000000" bold="false" italic="true" underline="false"/>
                <moreemph fontface="Arial" fontsize="28" color="#000000" bold="true" italic="false" underline="false"/>
            </textstyle>
            <textstyle type="Subtitle">
                <normal fontface="Arial" fontsize="24" color="#000000" bold="false" italic="false" underline="false"/>
                <emph fontface="Arial" fontsize="24" color="#000000" bold="false" italic="true" underline="false"/>
                <moreemph fontface="Arial" fontsize="24" color="#000000" bold="true" italic="false" underline="false"/>
            </textstyle>
            <textstyle type="Subsubtitle">
                <normal fontface="Arial" fontsize="20" color="#000000" bold="false" italic="false" underline="false"/>
```

```xml
                    <emph fontface="Arial" fontsize="20" color="#000000" bold="false" italic="true" underline="false"/>
                    <moreemph fontface="Arial" fontsize="20" color="#000000" bold="true" italic="false" underline="false"/>
                </textstyle>
                <textstyle type="TextIntro">
                    <normal fontface="Arial" fontsize="10" color="#000000" bold="false" italic="false" underline="false"/>
                    <emph fontface="Arial" fontsize="10" color="#000000" bold="false" italic="true" underline="false"/>
                    <moreemph fontface="Arial" fontsize="10" color="#000000" bold="true" italic="false" underline="false"/>
                </textstyle>
                <textstyle type="TextBody">
                    <normal fontface="Arial" fontsize="12" color="#000000" bold="false" italic="false" underline="false"/>
                    <emph fontface="Arial" fontsize="12" color="#000000" bold="false" italic="true" underline="false"/>
                    <moreemph fontface="Arial" fontsize="12" color="#000000" bold="true" italic="false" underline="false"/>
                </textstyle>
                <textstyle type="TextEnd">
                    <normal fontface="Arial" fontsize="10" color="#000000" bold="false" italic="false" underline="false"/>
                    <emph fontface="Arial" fontsize="10" color="#000000" bold="false" italic="true" underline="false"/>
                    <moreemph fontface="Arial" fontsize="10" color="#000000" bold="true" italic="false" underline="false"/>
                </textstyle>
                <textstyle type="References">
                    <normal fontface="Arial" fontsize="8" color="#000000" bold="false" italic="false" underline="false"/>
                    <emph fontface="Arial" fontsize="8" color="#000000" bold="false" italic="true" underline="false"/>
                    <moreemph fontface="Arial" fontsize="8" color="#000000" bold="true" italic="false" underline="false"/>
                </textstyle>
                <textstyle type="Fig-caption">
                    <normal fontface="Arial" fontsize="8" color="#000000" bold="false" italic="false" underline="false"/>
                    <emph fontface="Arial" fontsize="8" color="#000000" bold="false" italic="true" underline="false"/>
                    <moreemph fontface="Arial" fontsize="8" color="#000000" bold="true" italic="false" underline="false"/>
                </textstyle>
                <textstyle type="Footnote">
                    <normal fontface="Arial" fontsize="8" color="#000000" bold="false" italic="false" underline="false"/>
                    <emph fontface="Arial" fontsize="8" color="#000000" bold="false" italic="true" underline="false"/>
                    <moreemph fontface="Arial" fontsize="8" color="#000000" bold="true" italic="false" underline="false"/>
                </textstyle>
                <textstyle type="Button Text">
                    <normal fontface="Arial" fontsize="8" color="#000000" bold="false" italic="false" underline="false"/>
                    <emph fontface="Arial" fontsize="8" color="#000000" bold="false" italic="true" underline="false"/>
                    <moreemph fontface="Arial" fontsize="8" color="#000000" bold="true" italic="false" underline="false"/>
                </textstyle>
            </textstyles>
        </style>
    <Page xp:deliverytype="SCORM" xlink:label="L5415" xp:description="S1" id="Page01">
        <SubPages xp:deliverytype="SCORM" xlink:label="L5416" xp:description="Introduction" kind="R">
            <SubPage xp:deliverytype="SCORM" xlink:label="L5417" xp:description="Introduction_text">
                <ElementGroup xp:deliverytype="SCORM" xlink:label="L5418" xp:description="introduction_text_en" language="en">
                    <ElementGroup xp:deliverytype="SCORM" xlink:label="L5419" xp:description="Introduction_text_a" type="Tourist">
                        <Text xlink:label="L5420">Welcome. In this section you will discover one of Botticelli most beautiful masterpieces
shown in the Galleria degli Uffizi, the painting Madonna of the Pomegranate.</Text>
                        <audio src="L06_Madonna_Pomegranate_high/L06_Madonna_Pomegranate_high_Page01_en_tourist.wav"/>
                    </ElementGroup>
                    <ElementGroup xp:deliverytype="SCORM" xlink:label="L5421" xp:description="Introduction_text_b" type="Student">
                        <Text xlink:label="L5422">Welcome.  In this section you will discover one of Botticelli most beautiful masterpieces
shown in the Galleria degli Uffizi, the painting Madonna of the Pomegranate. Though it you will learn more about the painting.</Text>
                        <audio src="L06_Madonna_Pomegranate_high/L06_Madonna_Pomegranate_high_Page01_en_student.wav"/>
                    </ElementGroup>
                    <ElementGroup xp:deliverytype="SCORM" xlink:label="L5423" xp:description="Introduction_text_c" type="Teacher">
                        <Text xlink:label="L5424">Welcome. In this section you will discover one of Botticelli most beautiful masterpieces
shown in the Galleria degli Uffizi, the painting Madonna of the Pomegranate. Though it you will learn more about the painting and its mythological,
philosophical and literary background.</Text>
                        <audio src="L06_Madonna_Pomegranate_high/L06_Madonna_Pomegranate_high_Page01_en_teacher.wav"/>
                    </ElementGroup>
                </ElementGroup>
                <ElementGroup xp:deliverytype="SCORM" xlink:label="L5425" xp:description="introduction_text_it" language="it">
                    <ElementGroup xp:deliverytype="SCORM" xlink:label="L5426" xp:description="Introduction_text_a" type="Tourist">
                        <Text xlink:label="L5427">Benvenuti. In questa sezione scoprirete uno dei capolavori del Botticelli presente nella Galleria
degli Uffizi la Madonna della melagrana.<br/>
                        </Text>
                        <audio src="L06_Madonna_Pomegranate_high/L06_Madonna_Pomegranate_high_Page01_it_tourist.wav"/>
                    </ElementGroup>
                    <ElementGroup xp:deliverytype="SCORM" xlink:label="L5428" xp:description="Introduction_text_b" type="Student">
                        <Text xlink:label="L5429">Benvenuti. In questa sezione scoprirete uno dei capolavori del Botticelli presente nella Galleria
degli Uffizi la Madonna della melagrana.<b> </b>Potrete scoprire i segreti del dipinto.<br/>
                        </Text>
                        <audio src="L06_Madonna_Pomegranate_high/L06_Madonna_Pomegranate_high_Page01_it_student.wav"/>
                    </ElementGroup>
                    <ElementGroup xp:deliverytype="SCORM" xlink:label="L5430" xp:description="Introduction_text_c" type="Teacher">
```

```
                                <Text xlink:label="L5431">Benvenuti. In questa sezione scoprirete uno dei capolavori del Botticelli presente nella Galleria
degli Uffizi la Madonna della melagrana.<b> </b>Potrete scoprire i segreti del dipinto e i suoi legami con la storia, la filosofia e la letteratura del pe-
riodo.<br/>
                            </Text>
                            <audio src="L06_Madonna_Pomegranate_high/L06_Madonna_Pomegranate_high_Page01_it_teacher.wav"/>
                        </ElementGroup>
                    </ElementGroup>
                </SubPage>
            </SubPages>
        </Page>
</LOMobile>
<!-- End of XML file generated by eXact Packager -->
```

6.1.1   Authoring & Language related design

Creation and usage of the same content from different groups of people that may also have different sets of languages is a real problem and has been taken into account since long in the publishing environment yet is still at its dawn in the e-world. Traditional publishing tended to solve this issue through co-editions in which publishers belonging to the different linguistic environment would edit the same content in their own language adapting images and any other multimedia content to the needs. This would result in a set of titles bearing in essence the same original copyright (for the original version) and a set of new ones for each language. If the passage from language "A" to language "J" would require a passage from language "C" this would be addressed with specific contracts and in the edition "J" will be mentioned both "A" and "C" ©. In the context of e-learning we have addressed the problem exploiting the possibilities offered by the IMS standard format for content package description (the manifest) and in AXMEDIS we will be further exploiting this approach possibly extending it to a new level of completeness. What follows is an example of how to handle Multilanguage content in the same object using a simple structure.

```
    <Page xp:deliverytype="SCORM" xlink:label="L5415" xp:description="S1" id="Page01">
        <SubPages xp:deliverytype="SCORM" xlink:label="L5416" xp:description="Introduction" kind="R">
            <SubPage xp:deliverytype="SCORM" xlink:label="L5417" xp:description="Introduction_text">
                <ElementGroup xp:deliverytype="SCORM" xlink:label="L5418" xp:description="introduction_text_en" language="en">
                    <ElementGroup xp:deliverytype="SCORM" xlink:label="L5419" xp:description="Introduction_text_a" type="Tourist">
                        <Text xlink:label="L5420">Welcome. In this section you will discover one of Botticelli most beautiful masterpieces
shown in the Galleria degli Uffizi, the painting Madonna of the Pomegranate.</Text>
                        <audio src="L06_Madonna_Pomegranate_high/L06_Madonna_Pomegranate_high_Page01_en_tourist.wav"/>
                    </ElementGroup>
                </ElementGroup>
            </SubPage>
        </SubPages>
    </Page>
```

In fact, to add a new language version is enough to add an element with the same structure but a different language identification this will result in what follows.

```
    <Page xp:deliverytype="SCORM" xlink:label="L5415" xp:description="S1" id="Page01">
        <SubPages xp:deliverytype="SCORM" xlink:label="L5416" xp:description="Introduction" kind="R">
            <SubPage xp:deliverytype="SCORM" xlink:label="L5417" xp:description="Introduction_text">
                <ElementGroup xp:deliverytype="SCORM" xlink:label="L5418" xp:description="introduction_text_en" language="en">
                    <ElementGroup xp:deliverytype="SCORM" xlink:label="L5419" xp:description="Introduction_text_a" type="Tourist">
                        <Text xlink:label="L5420">Welcome. In this section you will discover one of Botticelli most beautiful masterpieces
shown in the Galleria degli Uffizi, the painting Madonna of the Pomegranate.</Text>
                        <audio src="L06_Madonna_Pomegranate_high/L06_Madonna_Pomegranate_high_Page01_en_tourist.wav"/>
                    </ElementGroup>
                </ElementGroup>
                <ElementGroup xp:deliverytype="SCORM" xlink:label="L5425" xp:description="introduction_text_it" language="it">
                    <ElementGroup xp:deliverytype="SCORM" xlink:label="L5426" xp:description="Introduction_text_a" type="Tourist">
                        <Text xlink:label="L5427">Benvenuti. In questa sezione scoprirete uno dei capolavori del Botticelli presente nella Galleria
degli Uffizi la Madonna della melagrana.<br/>
                        </Text>
                        <audio src="L06_Madonna_Pomegranate_high/L06_Madonna_Pomegranate_high_Page01_it_tourist.wav"/>
                    </ElementGroup>
                </ElementGroup>
            </SubPage>
        </SubPages>
    </Page>
```

It is worth noting that also the newly released XML MAG schema (v2.0) for CH description deals with languages, holding specific fields for this.

AXMEDIS Project

6.1.2 Specifying languages

The language aspect is to be handled also in the metadata part of the content, luckily in that respect standards provide some support as they state that wherever it is necessary to specify a language such as in data element 'General.Language' or in any language string the following coding can be used:

1. use a 2 letter code from ISO 639-1
2. use a 3 letter code from ISO 639-2 (see: http://www.loc.gov/standards/iso639-2/normtext.html, it does not matter between bibliographic & terminology since they only differ for languages that have 2-letter codes)
3. add the ISO Country code [ISO3166] when necessary, separated by a dash.
4. use IANA registered language tags, prefixed with **i-**
5. use SIL Ethnologue 3-letter codes, prefixed with **x-E-**
6. make up a name for token languages prefixed with **x-T-**
7. make up a name, prefixed with **x-** for user defined languages

All of the above are acceptable but partners should prefer 1, 2, or 3; in the following table some examples are provided:

Coding	Description	Type
nl	Dutch	ISO 639-1
aus	Australian Languages	ISO 639-2
IT	Italian	ISO 3166
i-xxxxx	IANA registered xxxxx	IANA
x-E-pcd	Picard	SIL Ethnologue
x-T-ELR	The ELR Token Language	SIL Ethnologue
x-none	Not possible to identify a language	SIL Ethnologue

It is worth mentioning that in several environments is currently being used the xml:lang attribute based on [IETF RFC 3066] (http://www.ietf.org/rfc/rfc3066.txt)

Delivery is not really language dependent as the language selection for content would have been performed at the initial step (access….) and therefore handled via fruition interface relying onto production based efforts (which ends up in the packaging).

6.1.3 Licenses and rights

Even though this may seem as a non-crucial problem yet it is a very important issue to be tackled, as end users will inevitably rely on localised versions of the license also because the license has to comply with the local rules and laws. We should differentiate here between contracts and licenses, and also visual representation from storage format. Further details will be given in the Metadata section of the present document. Contracts have to be expressed on the language of the party offering the service, content, etc. but licenses will be expressed using standard Rights Expression Languages (like MPEG-21 REL or ODRL), that are related to rights data dictionaries mainly expressed in English terms, although other languages could be possible, and even relationships between terms in different languages. A localized version of a license should comprise the expression of rights and conditions available for that county, but using the standard rights expression language. On the other hand, the visual representation of the license could be translated, but only at GUI level, as the license format will be the same for any country and language.

7 Metadata & multi-language management (ILABS, SEJER, OD2, ALL)

We have mentioned in several documents and also in other sections of this document the relevance of metadata as far as multi-lingual issue management is concerned. It is important, nevertheless, to make clear that there are two very different aspects to be taken into account in this respect:

- Multi-lingual aspects of the content and
- Multi-lingual aspects of metadata itself.

So far we have primarily focused on content related aspects, now our attention will be moved to metadata itself. This will be done in steps staring form underlying principles to then achieve a better detail.

7.1 Principles

Given the fact that AXMEDIS has its own metadata it is necessary to see where superposition allows automatic replication/transferring of data and where it will be necessary to specifically insert relevant metadata. In practice we have:

- **Data common between AXInfo and LOM** – this data will be possibly automatically filled in at creation time.
- **Data common between Dublincore and LOM** – this data will be possibly automatically filled in at creation time.
- **Data specific to AXInfo** – this data will be filled in at authoring time.
- **Data specific to LOM** – this data should have been possibly filled in at authoring time.
- **Data specific to Dublincore** – this data will be possibly filled in at authoring time for classification and management purposes.

There are two types of element subsets defined here: the elements that should be filled in every metadata instance (mandatory elements) and the elements that would be very useful to be filled (recommended elements). All other elements of our full element set are considered as optional and there is also information about some optional elements. In addition to element explanations this section contains full listings of the vocabularies defined by the project and the data types to be used as value spaces of metadata elements. The purpose of the Metadata Application Profile is to support the exchange of information about online digital resources (Learning Objects) between partners. The metadata described in this application profile supports a variety of LO uses including management, searching, finding, technical interoperability and description of properties of individual LOs including: educational attributes, Digital rights and technical features.

IEEE Learning Object Metadata standard (LOM) has been selected as basis for the set of adopted metadata and to support interoperability with other metadata schemes. The information model for the metadata is similar to that of LOM where metadata for a described LO is stored in a metadata element and actual content of an element is called a value. Values can be entered as free text, inserted in predefined format or they are selected from set lists, which are called vocabularies. There are five data types in the LOM information model briefly, LOM types are:

- **CharacterString**: text can be entered in the element directly.
- **LangString**: the text must identify its language and there can be one or more character strings in the element.
- **DateTime**: the element contains date and time information and there can also be textual information about this point in time.
- **Duration**: the element contains information about an interval in time and there can also be textual information about the duration.
- **Vocabulary**: the element contains source and value where source is a reference to publicly sourced and maintained value set and value is a value from that set.

In the remainder of the section are reported the principal metadata associated to a LO and their partitioning in are of major relevance. There are several areas and for each there is a set of values. It is nevertheless possible to modify this structure adding or removing data. Both operations have to be performed following IMS standards in order to avoid ending up with a metadata structure no more compliant. In the packager interface most relevant metadata have been grouped for easier input. In the advanced view is possible to insert, modify or delete data following the regular structure foreseen in IMS. It is worth noting that ISO-639 is included in the MPEG21 standards for: original primary spoken language and primary language of text element. Therefore it is wise to suggest that it should be applied the principle stated from W3C in respect to accessibility best practice in assigning language tags (within html) to the domain of MPEG21 objects, i.e. to identify the 'primary language' of spoken and written content wherever possible within an object.

7.2 Mandatory fields (LOM and Dublin core)

In the following paragraph are reported the metadata field that should be considered mandatory when compiling metadata. These are basically the fields that will enable efficient search and retrieval of LO.

- *General.Identifier.*
- *General.Title.*
- *General.Language.*
- *General.Description.*
- *Technical.Location.*
- *Educational.Intended End User Role.*
- *Educational.Typical Age Range.*

29

- ***Rights.Copyright and Other Restrictions.***
- ***Rights.Description.***
- ***Classification.Keyword.***

Dublin core set comprises elements that are considered optional in the more general sense, but here have been considered as being part of the mandatory part of object metadata. This is basically due to the fact that they provide the only set of metadata that could generically fit to a wide set of content. The Dublin Core Metadata Element Set is composed by the following elements:

- contributor
- coverage
- creator
- date
- description
- format
- identifier
- language
- publisher
- relation
- rights
- source
- subject
- title
- type

To these elements it is worth taking into account also other ones (detailed hereafter) as they represent Other Elements and Element Refinements.

7.3   Optional fields (IMS, IEEE-LOM, Dublin core Sets)

Recommended elements are those that would be very useful to have filled in for every metadata instance that is exposed, but they could be left unfilled.

- ***General.Keyword.***
- ***General.Structure.***
- ***Life Cycle.Contribute.***
- ***Life Cycle.Contribute.Role.***
- ***Life Cycle.Contribute.Entity.***
- ***Life Cycle.Contribute.Date.***
- ***Meta-Metadata.***
- ***Meta-Metadata.Contribute.Role.***
- ***Meta-Metadata.Contribute.Entity.***
- ***Meta-Metadata.Date.***
- ***Meta-Metadata.Language.***
- ***Technical.Format.***
- ***Technical.Size.***
- ***Technical.Facet.***
- ***Technical.Facet.Value.***
- ***Educational.Learning Resource Type.***
- ***Educational.Learning Context.***
- ***Educational.Typical Learning Time.***
- ***Educational.Description.***

Dublin core set comprises the following additional elements:

- ***abstract***
- ***accessRights***
- ***accrualMethod***
- ***accrualPeriodicity***
- ***accrualPolicy***
- ***alternative***
- ***audience***
- ***available***
- ***bibliographicCitation***
- ***conformsTo***
- ***created***
- ***dateAccepted***
- ***dateCopyrighted***
- ***dateSubmitted***
- ***educationLevel***
- ***extent***
- ***hasFormat.***
- ***hasPart***
- ***hasVersion***
- ***instructionalMethod***
- ***isFormatOf***
- ***isPartOf***
- ***isReferencedBy***
- ***isReplacedBy***
- ***isRequiredBy***
- ***issued***
- ***isVersionOf***
- ***license***
- ***mediator***
- ***medium***
- ***modified***
- ***provenance***
- ***references***
- ***replaces***
- ***requires***
- ***rightsHolder***
- ***spatial***
- ***tableOfContents***

- *temporal*
- *valid*

Whenever recommended best practice is to use a value from a controlled vocabulary, an important point is achieved as this could be easily make comply with multi-language issues as it would be sufficient to localize the related vocabularies to have localized also the management of fields referring to those vocabularies. It is worth mentioning that this is exactly the approach followed in managing similar elements in the LOM.

As far as the audiences for a resource is concerned is worth noting that there are of two basic classes: (1) an ultimate beneficiary of the resource, and (2) frequently, an entity that mediates access to the resource. The mediator element refinement represents the second of these two classes. As far as resource references are concerned, recommended practice is to include sufficient bibliographic detail to identify the resource as un-ambiguously as possible, whether or not the citation is in a standard form.

7.4 IPR related fields

Information about who can access the resource or an indication of its security status. Access Rights may include information regarding access or restrictions based on privacy, security or other regulations. Recommended best practice is to identify the license using a URI, the same applies to elements used to indicate the entity. Examples of such licenses can be found at http://creativecommons.org/licenses/. Given the specific context is worth taking into account that MPEG21 offers a specific solution and support for this isse as detailed hereafter.

7.5 DRM related fields (MPEG21 set)

DRM information in the AXMEDIS project will be expressed by means of rights expressions (licenses), described with MPEG-21 Rights Expression Language (REL). Other RELs will be considered during the development of the project. The vocabulary used in these licenses for the description of actions is described in MPEG-21 Rights Data Dictionary (RDD). Terms on MPEG-21 RDD are written in English, but the content of the fields expressed inside licenses are not bounded to any specific language. For instance, the title of the license could be written in Spanish or Italian and it will still be the title element, as expressed in the MPEG-21 REL XML Schema.

```xml
<?xml version="1.0" encoding="UTF-8"?>
<r:license xmlns:r="urn:mpeg:mpeg21:2003:01-REL-R-NS" xmlns:sx="urn:mpeg:mpeg21:2003:01-REL-SX-NS"
xmlns:mx="urn:mpeg:mpeg21:2003:01-REL-MX-NS" xmlns:dsig="http://www.w3.org/2000/09/xmldsig#"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="urn:mpeg:mpeg21:2003:01-REL-R-NS
../schemas/rel-r.xsd
                urn:mpeg:mpeg21:2003:01-REL-SX-NS ../schemas/rel-sx.xsd
                urn:mpeg:mpeg21:2003:01-REL-MX-NS ../schemas/rel-mx.xsd">
    <r:grantGroup>
        <r:grant>
            <r:keyHolder>
                <r:info>
                    <dsig:KeyValue>
                        <dsig:RSAKeyValue>
                            <dsig:Modulus>KtdToQQyzA==</dsig:Modulus>
                            <dsig:Exponent>AQABAA==</dsig:Exponent>
                        </dsig:RSAKeyValue>
                    </dsig:KeyValue>
                </r:info>
            </r:keyHolder>
            <mx:play/>
            <r:digitalResource>
                <r:nonSecureIndirect URI="http://www.onlinemusic.com/mySong.mp3"/>
            </r:digitalResource>
            <sx:exerciseLimit>
                <r:serviceReference>
                    <sx:uddi>
                        <sx:serviceKey>
                            <sx:uuid>ee1398c0-8abe-11d7-a735-b8a03c50a862</sx:uuid>
                        </sx:serviceKey>
                    </sx:uddi>
                </r:serviceReference>
```

```
        </sx:exerciseLimit>
      </r:grant>
   </r:grantGroup>
   <r:issuer>
      <r:keyHolder>
         <r:info>
            <dsig:KeyValue>
               <dsig:RSAKeyValue>
                  <dsig:Modulus>X0j9q99yzA==</dsig:Modulus>
                  <dsig:Exponent>AQABAA==</dsig:Exponent>
               </dsig:RSAKeyValue>
            </dsig:KeyValue>
         </r:info>
      </r:keyHolder>
   </r:issuer>
</r:license>
```

Example of a license expressed in MPEG-21 REL

As it can be seen in the previous sample license, XML elements are expressed in English language, like grant, exerciseLimit or play. Elements coming from other XML schemas, like the ones related to digital signatures (those starting with "dsig:") are also written in English.

The content of the elements can be expressed in any language, but there is no way to indicate which language is being used. For most of the cases, this does not represent a problem, as we deal with URIs or identifiers. For the case of information whose visual representation may change from country to country, like dates or numbers, the application in charge of showing DRM information should solve this problem, not the rights expression language.

8 References & Bibliography (ALL)

In this sections are reported the more relevant reference and standards that have to be taken into account when selecting content either for re-editing / publishing or distribution purposes. The reader will find here quick reference information and links to more in dept info.

[W3C] Guidelines on Language Tags usage: http://www.w3.org/International/articles/language-tags/
[TGN] Thesaurus of Geographic Names http://www.getty.edu/research/tools/vocabulary/tgn/index.html
[W3CDTF] ISO 8601 date encoding http://www.w3.org/TR/NOTE-datetime
[MIME] Internet Media Types http://www.iana.org/assignments/media-types/
[RFC3066] Language specifications http://www.ietf.org/rfc/rfc3066.txt
[ISO639] Language specifications http://www.loc.gov/standards/iso639-2/
[DCMITYPE] DCMI Type Vocabulary http://dublincore.org/documents/dcmi-type-vocabulary/
[XML MAG] schema (v2.0) http://www.iccu.sbn.it/MAG/MAG_2.0/MAG_sito_Schema/mag_2_0_ec.html

[1]   Adeline Yen Mah, "A thousand pieces of gold – A memoir of China's past through its proverbs", Chapter 16, pp 293 Harper Collins Publishers 2003
[2]   "Language Review & Guidelines", APLAWS project website, 2002, www.aplaws.org.uk/download_content/documents/translation_is_not_enough.pdf
[3]   "Accessibility and Usability Review & Guidelines", APLAWS project website, 2002, www.aplaws.org.uk/download_content/documents/usability_study.pdf
[4]   "APLAWS Pathfinder Final Report", APLAWS project website, 2002, http://www.aplaws.org.uk/download_content/documents/Final_APLAWS_Pathfinder_Report.doc
[5]   "Centralizing The Localization Process", Transware White papers, 2005, www.transware.com/services/
[6]   "Developing Localization Friendly E-Learning", TRANSWARE White papers, 2004, www.transware.com/services/
[7]   "Multilingual E-Learning – Getting It Right From The Start", TRANSWARE White papers, 2004, www.transware.com/services/

[8]   "TRANSWARE's Solutions For Audio Localization", TRANSWARE White papers, 2004, www.transware.com/services/

[9]   "Internationalization", TRANSWARE White papers, 2005, www.transware.com/services/

[10]   "Web Globalization Best Practice", TRANSWARE White papers, 2005, www.transware.com/services/

[11]   John Harris and Ryan McCormack, "Translation is not enough. Considerations for global Internet development", Sapient Corporation, 2000, www.sapient.com

[12]   Robin Lloyd, "Three Approaches to Multilingual Content Management - Using a Content Management System (CMS)", KM World (Special Supplement), 2002

[13]   Willem Stoeller, "The Hub of the Wheel", PROJECT MANAGEMENT - The Guide from MultiLingual Computing & Technology - #63 Supplement April/May 2004

[14]   Claire Ingram, "Working With the Client to Ensure a Successful Project", PROJECT MANAGEMENT - The Guide from MultiLingual Computing & Technology - #63 Supplement April/May 2004

[15]   Thomas Waßmer, "Four Translation Project Management Tools", PROJECT MANAGEMENT - The Guide from MultiLingual Computing & Technology - #63 Supplement April/May 2004

[16]   Kenneth McKethan, Jr., "Project Churn: Counting the Cost", PROJECT MANAGEMENT - The Guide from MultiLingual Computing & Technology - #63 Supplement April/May 2004

[17]   Jeff Allen, "Project Management Checklist", PROJECT MANAGEMENT - The Guide from MultiLingual Computing & Technology - #63 Supplement April/May 2004

[18]   "LPM Best Practices Survey", PROJECT MANAGEMENT - The Guide from MultiLingual Computing & Technology - #63 Supplement April/May 2004

[19]   MLC Research, "Best Practices Build Solid Projects", PROJECT MANAGEMENT - The Guide from MultiLingual Computing & Technology - #63 Supplement April/May 2004

9   Terminology (ALL)

Term	Explanation (including source if available)
CMS	Content Management Systems
DCMIType	DCMI Type Vocabulary. A list of types used to categorize the nature or genre of the content of the resource
DDC	Dewey Decimal Classification. See: http://www.oclc.org/dewey/index.htm
IETF	Tags for the Identification of Languages see RFC 3066
IMT	The Internet media type of the resource. See: http://www.iana.org/assignments/media-types/
ISO3166	ISO 3166 Codes for the representation of names of countries. See: http://www.iso.org/iso/en/prods-services/iso3166ma/02iso-3166-code-lists/list-en1.html
ISO-639	International Standard naming system for languages. ISO-639 codes now include subtags to identify dialects such as lang='en-GB' indicates British-English content rather than US-English.
ISO639-2	ISO 639-2: Codes for the representation of names of languages. See: http://lcweb.loc.gov/standards/iso639-2/langhome.html
KMS	Knowledge Management Systems
LCC	Library of Congress Classification. See: http://lcweb.loc.gov/catdir/cpso/lcco/lcco.html
LCSH	Library of Congress Subject Headings
LCMS	Learning Content Management Systems
MESH	Medical Subject Headings. See: http://www.nlm.nih.gov/mesh/meshhome.html
NLM	National Library of Medicine Classification: See: http://wwwcf.nlm.nih.gov/class/

Term	Explanation (including source if available)
ODL	Open and Distance Learning
RFC1766	Internet RFC 1766 'Tags for the identification of Language' specifies a two letter code taken from ISO 639, followed optionally by a two letter country code taken from ISO 3166. See: http://www.ietf.org/rfc/rfc1766.txt
RFC3066	Internet RFC 3066 'Tags for the Identification of Languages' specifies a primary subtag which is a two-letter code taken from ISO 639 part 1 or a three-letter code taken from ISO 639 part 2, followed optionally by a two-letter country code taken from ISO 3166. When a language in ISO 639 has both a two-letter and three-letter code, use the two-letter code; when it has only a three-letter code, use the three-letter code. This RFC replaces RFC 1766. See: http://www.ietf.org/rfc/rfc3066.txt
TGN	The Getty Thesaurus of Geographic Names. See: http://www.getty.edu/research/tools/vocabulary/tgn/index.html
UDC	Universal Decimal Classification. See: http://www.udcc.org/
URI	A URI Uniform Resource Identifier. See: http://www.ietf.org/rfc/rfc2396.txt
W3CDTF	W3C Encoding rules for dates and times - a profile based on ISO 8601. See: http://www.w3.org/TR/NOTE-datetime

9.1 The DCMI Type Vocabulary

Term	Explanation (including source if available)
Collection	A collection is an aggregation of items. The term collection means that the resource is described as a group; its parts may be separately described and navigated.
Dataset	A dataset is information encoded in a defined structure (for example, lists, tables, and databases), intended to be useful for direct machine processing.
Event	An event is a non-persistent, time-based occurrence. Metadata for an event provides descriptive information that is the basis for discovery of the purpose, location, duration, responsible agents, and links to related events and resources. The resource of type event may not be retrievable if the described instantiation has expired or is yet to occur. Examples - exhibition, web-cast, conference, workshop, open-day, performance, battle, trial, wedding, tea-party, conflagration.
Image	An image is a primarily symbolic visual representation other than text. For example - images and photographs of physical objects, paintings, prints, drawings, other images and graphics, animations and moving pictures, film, diagrams, maps, musical notation. Note that image may include both electronic and physical representations. (Broader than: StillImage/MovingImage)
InteractiveResource	An interactive resource is a resource which requires interaction from the user to be understood, executed, or experienced. For example - forms on web pages, applets, multimedia learning objects, chat services, virtual reality.
MovingImage	A series of visual representations that, when shown in succession, impart an impression of motion. Examples of moving images are: animations, movies, television programs, videos, zoetropes, or visual output from a simulation. Instances of the type "Moving Image" must also be describable as instances of the broader type "Image".
PhysicalObject	An inanimate, three-dimensional object or substance. For example -- a computer, the great pyramid, a sculpture. Note that digital representations of, or surrogates for, these things should use Image, Text or one of the other types.
Service	A service is a system that provides one or more functions of value to the end-user. Examples include: a photocopying service, a banking service, an au-

Term	Explanation (including source if available)
	thentication service, interlibrary loans, a Z39.50 or Web server.
Software	Software is a computer program in source or compiled form which may be available for installation non-transiently on another machine. For software which exists only to create an interactive environment, use interactive instead.
Sound	A sound is a resource whose content is primarily intended to be rendered as audio. For example - a music playback file format, an audio compact disc, and recorded speech or sounds.
StillImage	A static visual representation. Examples of still images are: paintings, drawings, graphic designs, plans and maps. Recommended best practice is to assign the type "text" to images of textual materials. Instances of the type "Still Image" must also be describable as instances of the broader type "Image".
Text	A text is a resource whose content is primarily words for reading. For example - books, letters, dissertations, poems, newspapers, articles, archives of mailing lists. Note that facsimiles or images of texts are still of the genre text.