



The logo features the word "eContent" in a stylized, cursive font. The "e" is yellow, and "Content" is blue. To the right, the word "LIRICS" is written in a bold, black, sans-serif font.

Deliverable D 6.4

Demonstration Workshop Report

| | |
|--|--|
| Project reference number | e-Content-22236-LIRICS |
| Project acronym | LIRICS |
| Project full title | Linguistic Infrastructure for Interoperable Resource and Systems |
| Project contact point | Laurent Romary, INRIA-Loria 615, rue du jardin botanique BP101. 54602 Villers lès Nancy (France) romary@loria.fr |
| Project web site | http://lirics.loria.fr |
| EC project officer | Erwin Valentini |
| Document title | Demonstration Workshop Report |
| Deliverable ID | D 6.4 |
| Document type | Report |
| Dissemination level | Public |
| Contractual date of delivery | M28 |
| Actual date of delivery | M30 |
| Status & version | 1.0 |
| Work package, task & deliverable responsible | University of Vienna |
| Author(s) & affiliation(s) | Gerhard Budin |
| Additional contributor(s) | <representatives of all consortium partners in particular Laurent Romary, Gil Francopoulo, Thierry Leclerck, Harry Bunt, Nicoletta Calzolari, Nuria Bel, Lee Gillam, et al.> |
| Keywords | <demonstration session, industrial applications, dissemination, promotion, ISO> |

Overview and Structure of this Deliverable

The meeting of the Industrial Advisory Board that was held at AFNOR on the 10th of May 2007 also served as the demonstration workshop, since the right target audience for this purpose was present. Since then, additional presentations were given in different locations and to different audiences with demonstrations of LIRICS results as a main purpose of such events (also as part of our dissemination efforts) .

First, the agenda of the demonstration workshop is presented, then a short summary is given for each presentation. Conclusions finally describe the result of this work.

This report is closely linked to D6.3A, the industrial recommendation report.

1. Agenda of the demonstration workshop



Linguistic **I**nfrastructure for **I**nteroperable
Resour**C**es and **S**ystems

<http://lirics.loria.fr>

Industry Advisory Board Meeting

Paris, 10 May 2007

LIRICS addresses the needs of today's information and communication society where globalisation and localization necessitate multilingual communication creating an increasing need for new standardization as well as urgent recognition of existing de facto standards and their transformation into 'de jure' International Standards. LIRICS thus aims to:

- ‡ Provide ISO ratified standards for language technology to enable the exchange and reuse of multilingual language resources;
- ‡ Facilitate the implementation of these standards for end-users by providing an open-source implementation platform, related web services and test suites building on legacy formats, tools and data;
- ‡ Gain full industry support and input to the standards development via the Industry Advisory group and demonstration workshops
- 🔗 Provide a pay-per-use business model for use by Industry and in particular SMEs validated during the project for the benefit of all actors in the content and language industries

Agenda 2007-03-05

1 Opening of the meeting and welcome of participants

2 Presentation of the state-of-the-art in the LIRICS project – demonstration of results:

2.1 INTRO

Opening session: outlines of the project (Laurent Romary)

LIRICS application contexts - overview and history of TC37 activities (Gerhard Budin)

2.2 LANGUAGE CODES & DCR

639 family of standards (Lee Gillam)

Morpho-syntactic profile (Gil Francopoulo)

Syntactic profile (Thierry Declerck)

Semantic profile and TimeML (Harry Bunt)

2.3 LEXICONS & ANNOTATION

LMF (Monica Monachini, Gil Francopoulo)

MAF & SynAF (Nuria Bel, Thierry Declerck)

2.4 MULTIMEDIA ENVIRONMENT

MLIF (Samuel Cruz-Lara)

Controlled Authoring and Video Annotation (Lee Gillam)

2.5 SOFTWARE Demonstrations

Reference implementation for DCR management, lexicon (Adam Funk, Marc Kemps)

Management, annotation

3 Discussion with IAG members on usefulness and usability of standards and software developed in the LIRICS context

4 Recommendations by IAG members and further work (joint testing, etc.) (-> see D6.3A)

5 Any other business

6 Closure of the meeting at 17.00

2. Summaries of demonstrations and presentations

1 Opening of the meeting and welcome of participants

Laurent Romary and Gerhard Budin welcome the participants and explain that this workshop is at the same time a meeting of the Industrial Advisory Board of the LIRICS project and a demonstration workshop with the purpose of showing what has been implemented so far.

2 Presentation of the state-of-the-art in the LIRICS project – demonstration of results:

2.1 INTRO

2.1.1 Opening session: outlines of the project (Laurent Romary)

Laurent Romary gives a short overview of the project work, listing the achievements and results and focusing on the practical employability of the results.

2.1.2 LIRICS application contexts - overview and history of TC37 activities (Gerhard Budin)

Gerhard Budin explains the procedures of ISO standardization and gives an overview of ISO TC 37 standards in general, focusing on all standards and standards projects that are of particular relevance to language industry for implementation. With customers of language industry companies becoming more and more interested and knowledgeable in semantic web applications, knowledge organization systems, natural language processing, etc. the relevance of TC 37 standards in general and of the those standards projects that are run by the LIRICS consortium members in particular has significantly risen.

For instance, the current work item on word segmentation (all European expert members of the project team working on this item are part of the LIRICS consortium – Gerhard Budin, Gil Francopoulo, Laurent Romary – the other members are from China, Korea, Japan, USA) has received increased attention in translation industry

and localization, because word segmentation has become a major issue and bone of contention between language service providers and their customers when using translation memory systems. LISA (Localization Industry Standards Association) has been working on this topic and is expecting results from the work in LIRICS on this topic.

Even more relevant for language industry is the work on MLIF, the Multi-Lingual Information Framework that is run almost exclusively by LIRICS team members (including convenor and project editor from LORIA). This will become a process and workflow standard that integrates not only existing translation and localization standards in a workflow, but also multimedia and semantic web aspects (SMIL). Further details see below.

2.2 LANGUAGE CODES & DCR

2.2.1 The 639 family of standards (Lee Gillam)

Lee Gillam presents the recent developments in the ISO 639 framework of language coding. Rising from 200 to 30 000 codes, the ISO 639 multi-part standard has become one of the most important and most widely used standards of ISO. In fact, everybody using the Internet unavoidably uses ISO 639, as it has been deeply embedded in all Internet protocols for language identification. In addition, localization industry increasingly is interested in local language variations, also coded by ISO 639. Thus, the ecology of standards covered by the LIRICS project work depends on a dynamic development AND implementation of ISO 639 in a language industry implementation context. Web-based applications and cross-database integration efforts depend on consistent and interoperable implementations of the code tables of ISO 639. It is important to note that ISO 639 is referenced by all language resource standards (co-)prepared by the LIRICS group, so it is a kind of horizontal, fundamental pre-requisite standards for other, more specialized language industry standards to become fully operational in real-life applications.

The demo of this profile is part of 2.5.

2.2.2 Morpho-syntactic profile (Gil Francopoulo)

For the Data Category Registry (see 2.5) implementation several category selection profiles for ISO 12620 have been developed and are being implemented. The Morpho-syntactic profile presented by Gil is one of several profiles that have been specified. The profiling is important in order to provide language industry companies with manageable portions of data category modelling categories in order to enable them to compose their own user profiles on the basis of the thematically organized profile specifications.

The morpho-syntactic profile is a crucial one for many user communities in translation, terminology, lexicography, language engineering for natural language generation and analysis, etc. Basically it is an endless task, since all languages in the world in theory would have to be analyzed for potential specific needs not yet covered for other languages.

The demo of this profile is part of 2.5.

2.2.3 Syntactic profile (Thierry Declerck)

In a similar way Thierry presented the syntactic profile (also an extension model of LMF) that is also quite important for language engineering applications and an integral part of a coherent language industry product development strategy.

The demo of this profile is part of 2.5

2.2.4 Semantic profile and TimeML (Harry Bunt)

Harry presents the semantic profile, another crucial and indispensable part of any language processing application product. In this context the work on TimeML has gained significant momentum (meeting on TimeML a few days before the IAB meeting and demo workshop in Paris) and attention in the semantic web community in particular. For automated discourse analysis but also for natural language generation TimeML is crucial. The semantic dat cat profile is expected to be further extended in coming months, as it is a very complex sub-set with many ramifications into different specialized user scenarios. The typology of semantic relations is a good example of this basically open range of developments.

ISO 12620 is profiting considerably from these profiling specifications for real-life implementations (see 2.5)

2.3 LEXICONS & ANNOTATION

2.3.1 LMF (Monica Monachini, Gil Francopoulo)

The work on the Lexical Markup Framework standard is presented by Gil and Monica. It is the most important but also most time consuming and certainly very complex project work in LIRICS. Version 14 has been presented in summer 2007. Work on LMF represents a major share of time spent in the LIRICS project because of its integrating nature (extension models, dat cat profile implementations in these extension models). Much is at stake with LMF, which is why non-European delegations also work very hard on LMF. Even more important is thus the significant participation of the LIRICS expert group as part of several national delegations in order to make sure that European language industry needs are taken into account.

2.3.2. MAF & SynAF (Nuria Bel, Thierry Declerck)

These two annotation frameworks have also been worked on in the last few months. They will also become important building blocks in annotation software

developments, especially in the context of web-based and semantic-web-oriented annotation workflows for different user communities and their specified purposes (annotation for discourse analysis, annotation for translation corpora, etc.)

2.4 MULTIMEDIA ENVIRONMENT

MLIF and video annotation are relatively new to the LIRICS project, but this thematic extension has been a significant step into integrated implementation strategies of the standards framework prepared in this context:

2.4.1 MLIF (Samuel Cruz-Lara)

As mentioned above, MLIF has become a success story for LIRICS. It integrates from a work-flow perspective existing industry standards such as XLIFF, TMX, TBX, etc. into a dynamic user context. Talks with tool developers in translation and localization industries are focusing on first implementations of MLIF in product suites planned for 2008. The work on MLIF in 2007 is focusing on arranging and specifying its building blocks. A new version is expected after the very successful August meeting (Provo) of TC 37 to become a Committee Draft. LISA is very much interested in promoting MLIF in their membership not only with technology providers but also service providers and vendors.

SMIL (Synchronized Multimedia Integration Language) has been developed for some time but has gained significant momentum and importance recently in the W3C context and now with its integration into the MLIF context. Media industry is interested in implementation. In May 2007 we had an industry meeting at the University of Vienna for media translation implementations. Several companies (e.g. Sysmedia from the UK, Titelbild from Germany, etc.) were present and got interested in MLIF+SMIL specifications as input to their future product generation and service provider spectrum (e.g. for live subtitling, automatic multilingual video annotation, etc.).

2.4.2 Controlled Authoring and Video Annotation (Lee Gillam)

A closely related topic is then presented by Lee Gillam in a fascinating outlook on controlled authoring and video annotation. As with MLIF+SMIL, the proposals presented by Lee are highly relevant to language industry in general and for authoring/editing/writing communities and multimedia industry groups in particular.

2.5 SOFTWARE Demonstrations

In this part of the workshop the reference software implementation is presented by Univ of Sheffield and by the Max Planck Institute for Psycholinguistics, where the software has been prepared. It integrates most of the topics described above.

2.5.1 Reference implementation for DCR management, lexicon Management, annotation (Adam Funk, Marc Kemps)

Three software components are presented:

- Data Category Registry (DCR) web service interface
- LMF web service interface
- GATE LMF plugin

They include the 12620/DCR profiles mentioned under 2.2 and the annotation schemes mentioned under 3.2.

The accompanying Power Point slide collection describes these software components.

The demo runs successfully. All functionalities are explained in detail to the audience.

On the basis of the workshop the sw development team has further developed the software, also in cooperation with Sue Ellen Wright. In August 2007 the latest version was successfully presented at the ISO TC 37 meetings in Provo, Utah, USA, also with the positive feedback from industry representatives and from people from the ISO Central Secretariat in Geneva, who now think about using this implementation in their standards as database initiative for metadata management.

3. Conclusions

The discussion after the presentations and demonstrations showed that the IAB members welcome the developments in the LIRICS project work and encourage the group to continue on this path.

The demonstrations were successful. As mentioned, subsequent presentations and demonstrations (e.g. in August this year) were also successful, thus encouraging the development teams to speed up their finalization work for the products to be delivered by the LIRICS project consortium.

In concluding we can state that the demonstration workshop was very successful.

The recommendations formulated by the Industrial Advisory Board are contained in document D6.3A.