



LIRICS

Deliverable D 6.1

Dissemination Plan

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	Dissemination Plan
Deliverable ID	D 6.1
Document type	Report
Dissemination level	Public
Contractual date of delivery	M2
Actual date of delivery	M8 updated version
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	dissemination, promotion, liaison, networking

Document evolution

Version	date	Version	Date
0.1	1st March 2005		
0.2	17th August 2005		
1.0	18th August 2005		

Introduction and Overview

The LIRICS Dissemination Plan is the first deliverable of WP 6 that focuses on dissemination and exploitation. The Dissemination Plan consists of the following strategies that are described in the subsequent chapters:

1. Direct promotion and dissemination with colleagues and communities in various target groups
2. Direct promotion and dissemination through multipliers in different communities of practice
3. Public dissemination and promotion by the project website
4. A discussion forum
5. Regular events of different types
6. Special strategy concerning internationalization/localization industry communities
7. Standardization bodies
8. Publications
9. Exploitation Perspectives

1. Direct promotion and dissemination with colleagues and communities in the following target groups:

1.1. Language industry companies

1.1.1. Language Service providers (companies focusing on this portfolio)

1.1.2. Tool/technology providers in the language industries (translation tools, machine translation systems, terminology management systems, language management tools, language processing tools, etc.)

- 1.1.3. Multimedia language Content providers (E-Learning (CALL) product publishers, speech corpus content publishers, public and private mass media/broadcast companies offering multimedia language content, etc.)
- 1.2. Individual language industry service providers such as translators, technical writers, etc.
- 1.3. Customers of language industry services and users of language technologies (language industry also includes translation services and technologies, localization/internationalization, technical authoring, information design, lexicographical products and services, etc.)
 - 1.3.1. Public institutions (governments/departments with specific requirements in terms of functionalities of tools/content from the perspective of procurement as well as language policy implementation support and guidance
 - 1.3.2. Companies interested in buying linguistic content and/or language services
 - 1.3.3. Communities of practice, professional associations (both service provider organizations as well as customer organizations)
 - 1.3.4. Individuals at large (“non-linguists”) simply interested in this area
- 1.4. Language technology scholars and experts (but also students in academic environments), in particular in the fields of:
 - 1.4.1. Computational linguistics
 - 1.4.2. Language engineering studies
 - 1.4.3. Translation technologies studies
 - 1.4.4. Localization studies
 - 1.4.5. Terminology engineering studies
 - 1.4.6. Knowledge engineering
 - 1.4.7. Technical authoring
 - 1.4.8. Information design
 - 1.4.9. Standardization

This strategy includes the preparation and publication of research articles in scientific journals as well as giving lectures and presentations at scientific events (congresses, workshops, etc.).

Acceptance of LIRICS in the scientific communities in the above mentioned disciplines is a crucial factor for the success of the project, it also makes sure that new experts enter the process of testing specifications, contributing to the creation of new specifications that in turn are transformed into standards.

2. Direct promotion and dissemination through multipliers in different communities of practice

2.1. Each consortium member institution of the LIRICS project (the consortium members were selected in the first place not only due to their scientific expertise and excellence but also due to the fact that they are all opinion leaders in various different constituencies, scientific disciplines and communities of practice, with national and/or regional and local authorities. This “oral propaganda” for LIRICS is also an important success factor.

2.2. Each company member of the Industry Advisory Group (e.g. with their customers). This dissemination channel is very important, as it also includes the actual implementation of the results of the standards-production work packages in real-world product and service environments. The **implementation strategy** developed in this work package is based on a contextual and embedded approach that is closely related to a specific kind of strategic consulting service provided to the advisory group member companies in the context of continuously working together with the consortium members. This means that on the basis of the specific needs and interests of the companies in the industrial advisory group pilot, implementations of the standards and specifications developed in LIRICS are pilot applications that are an additional input to the testing strategy.

3. Public dissemination and promotion by the project website

3.1. The project website (<http://lirics.loria.fr>) is an important tool for the public dissemination policy. On the public part of the website a whole range of documents are already publicly accessible. This instrument is highly relevant to the other dissemination strategies discussed in this document as it functions in the context of cooperation with other groups that are listed in the chapters 1,2, 4, 5,6, 7 for the purpose of referring to working documents to be distributed to members of these different groups and communities.

4. A discussion forum

4.1. A discussion forum will be set up once there are concrete documents presented for public discussion. This is expected to be in October 2005. At the moment such discussion lists (like the lirics discussion list) are for communication within the consortium and between the consortium and the Industry Advisory Group. It is essential not to start the public discussion forum too early in order to avoid the danger that participants will lose interest soon when there is no constant traffic on the list for lack of relevant discussion issues. After October 2005, on the other hand, there will be a whole taxonomy of discussion issues that will be suggested to the future members of the discussion list.

5. Regular events of different types

5.1. Internal workshops

Internal workshops are used, in addition to discussing the actual technical work items and details of document writing, as fora for strategic discussions of how to deal with the different target groups of the project and of how to constantly adapt, adjust and fine-tune the various dissemination strategies mentioned above and below, to prepare upcoming events and to coordinate the dissemination activities of members of the consortium.

5.2. Workshops with the Industry Advisory Group

5.2.1. The first workshop of this kind was successfully held on 20/21 of June 2005 in Barcelona on the premises of University Pompeu Fabra. Members of the LIRICS consortium presented the details of the work plan and interim results of the various work packages to the members of the Industrial Advisory Group. These in turn presented their expectations and their advice to the project group. The presentations that were given are available on the Lirics-website.

5.3. Public workshops for informing about the progress of work and for attracting new members in user communities

5.3.1. Some workshops of this type have been held and will continue to be held by consortium members in their home countries and at local level to local industry as well as computational linguistics groups.

5.3.2. European-wide public workshops will be held in conjunction with international/European events such as the LREC Conference in May 2006 in Genoa.

A LIRICS workshop has been held in Copenhagen on the occasion of the TKE '05 (Terminology and Knowledge Engineering) on the 18th of August 2005, focusing on discussing the TMF document and its relation to SKOS of W3C. Alan Melby and Sue Ellen Wright presented together with Gerhard Budin the workshop content. Workshop participants were particularly interested in the mapping between the SKOS vocabulary (draft) and the ISO 12620 data categories as they are implemented in the TBX and TMF standards. The topics of ontology engineering formats (OWL) and conceptual schema formats such as SKOS will play an important role in the LIRICS project as many language engineering processes and language technology standards need to refer to ontologies and knowledge organization systems in their implementations. Interest in this particular link between language engineering and ontological knowledge engineering is rising in commercial and public application environments as well as in academic research communities and standardization bodies.

5.3.3 ESSLLI 2005

Thierry Declerck presents at the summer school ESSLLI 2005 in Edinburgh the LIRICS standardization work in connection with a lecture dedicated to MPEG-7 within a course entitled "NLP for Multimedia Applications". (August 8-19 2005) (The main topic: relation between TC37/SC4 and the MPEG-7 Standard for the description of multimedia content representation)

5.3.4 Other events

Nuria Bel has presented LIRICS at the Internet Global Conference that was held in Barcelona (6-10 June 2005).

Nicoletta Calzolari was an invited speaker at the International Workshop on the Generative Lexicon that was held in Geneva, 19-21 May 2005, as well as invited speaker at the IULA 10th anniversary conference in Barcelona, 16-17 June 2005.

Such presentations significantly contribute to the visibility of the LIRICS project in the research communities addressed in chapter 1 of this report. Such presentations (as well as the publications listed under chapter 8) will continue to be an important strategy for the success and the sustainability of the project's mission.

5.3.5 Future Activities, future events

Therry Declerck is co-chair of a workshop dedicated to semantic annotation: "SemAnnot", which is a satellite event to the 4th International Semantic Web Conference (ISWC 2005), to take place between the 6th and 10th of November 2005. The workshop will be endorsed by LIRICS.

Gerhard Budin succeeded in being included in the organization of a major event to be held in December in Berlin, focusing on the topic of language industry standardization for global business and eGovernment. He has been invited to present the LIRICS project there, not only as an individual contribution to language technology standardization, but also as a platform for international and inter-regional standards development as well as the co-ordination among national standardization activities. The draft programme will be finished soon.

The LIRICS group has also been invited to organize a workshop in conjunction with the WordNet Conference in January 2006 in South Korea.

Not surprisingly, LREC 2006 will be THE major public event for LIRICS in 2006. A one-day LIRICS workshop is in preparation for Genova in May 2006. This is also true for the Open Forum on Metadata Registries 2006 in Kobe, after the successful contribution of LIRICS members to the 2005 OFMR event in Berlin (see below under 6.1).

Harry Bunt is planning the next International Workshop on Computational Semantics for January 2007 in Tilburg. In conjunction with this event, LIRICS will be actively involved in presenting its results in the framework of the ACL-SIGSEM WG on the Representation of Multimodal Semantic Information. This event will have to be prepared in working group meetings that will be held in the WPs of LIRICS in 2006.

6. Special strategy concerning internationalization/localization industry communities

6.1. Unicode Consortium and Open Forum for Metadata Registries

The organizers of the Unicode Conferences had asked Gerhard Budin to host the Unicode Conference in April 2005. He proposed to combine this conference with the Open Forum for Metadata Registries that he also co-organized. So both events took place in Berlin in April 2005.

Gerhard Budin gave a paper at the International Unicode Conference and organized ad hoc meetings with crucial members of the Unicode consortium in order to discuss standardization aspects relevant to LIRICS. Several other members of the LIRICS consortium (including the project coordinator Laurent Romary and several others) presented papers at the Open Forum for Metadata Registries and discussed the role of language technology standards in enhancing semantic interoperability in federated information sources and databases. Cooperation between ISO/TC 37 and ISO/IEC JTC 1 is of crucial importance to the success of the LIRICS project. The same is true for the UNICODE Consortium! Similar events are planned for 2006 and 2007.

6.2. LISA

The Localization Industry Standards Association is another crucial partner group for the LIRICS project, in all respects – for finding out industry expectations and requirements in terms of language technology standards, for jointly developing or further refining standards (e.g. TMX, TBX) and for developing joint strategies for promoting the consistent use of language technology standards in all industry sectors where localization and internationalization activities are relevant. It is foreseen to organize a LIRICS workshop at a LISA event in 2006.

6.3. Co-operation with Other groups

Other institutions such as the Localization Institute, the Localisation Research Centre, and ASIS/XLIFF are other important institutions and communities that will be dealt with in the coming months of the project work.

In the framework of the language resource production activity of Pcom, it was decided that production of new lexical resources, especially those resulting from merging of existing resources, will be performed according to the recommendations developed by the LIRICS project.

7. Standardization bodies

7.1. ISO

Obviously the work with ISO is the major focus of LIRICS in order to reach its strategic goals. Since the LIRICS project has started, several meetings have taken place. The TC 37 meeting week in Warsaw in August is clearly a crucial week for discussing all work items currently in preparation in the various sub-committees of TC 37 (in particular SC 4). The members of the

LIRICS consortium are presenting their intensive work in work packages 2-5 to the relevant ISO working groups in order to speed up and improve this work and in order to steer these work items in a direction that actually mirrors the interests, needs and requirements of all user groups.

Also other ISO committees are relevant for the LIRICS project: TC 46, JTC 1, TC 184, TC 176 and others.

Additional activities in relation to LIRICS have been continuing with respect to developments of the international standards for language codes, ISO 639. Budin (Vienna), Gillam (Surrey) and Romary (INRIA/CNRS) have been involved with activities progressing through the ISO standardisation process in relation to 4 current ISO projects that are at various stages:

- ISO 639-3 - Codes for the representation of names of languages – Part 3: Alpha-3 code for comprehensive coverage of languages
- ISO 639-4 - Codes for the representation of names of languages – Part 4: Implementation guidelines and general principles for languages coding
- ISO 639-5 - Codes for the representation of names of languages – Part 5: Alpha-3 code for language families and groups
- ISO 639-6 - Codes for the representation of names of languages – Part 6: Codes for comprehensive coverage of language variation

7.2. CEN/ISSS

Several Workshops in the CEN/ISSS area have been established in the last two years that have turned out to be very useful for LIRICS work: eCAT (electronic product catalogs, product classification), ADNOM (Administrative Nomenclature), eGOV (electronic government), Knowledge Management, E-Learning Technologies, etc. Cooperation has started with these groups and will be intensified in autumn 2005. Due to the fact that language-related standardization has increasingly been accepted as a necessity in the CEN framework, LIRICS has been receiving increased visibility. Gerhard Budin as chair of CEN/ISSS/ADNOM can use this role to link LIRICS to CEN activities.

7.3 National Standardization Activities

Germany:

DFKI organises the meeting of “DIN NAT AA 6 Sprachressourcen” the German Mirror Committee of TC37/SC4, 13.5.2005. Thierry Declerck gives a detailed description of LIRICS. Scope and goals of the project are highly appreciated by the DIN members, who acknowledge the importance of carrying out standardisation activities at European Level. Possible concertation (also at national level) with the eContent project

EurotermBank is suggested by Dr. Klaus Schmitz (member of DIN AA5 and partner of EurotermBank).

Austria:

Gerhard Budin as chair of the Austrian Standards Committee on Terminology (FNA 033) presents the LIRICS project to the members of this committee and discusses with them the ways of active contribution to the LIRICS project from the perspective of Austrian terminology standardization. Several new members were found for this committee due to the expanded perspectives of language-related standardization.

France:

Laurent Romary and Gil Francopoulo actively contribute to national standardization activities in the AFNOR framework, where LIRICS has received substantial visibility.

8. Publications

During the period January to August 2005, the following papers were either published or accepted for publication in relation to the LIRICS project, and specifically to activity on document quality management (WP1 T1.3), by the University of Surrey:

- Gillam, Tariq and Ahmad (2005) "Terminology and the Construction of Ontology". Terminology 11(1), pp55-81. John Benjamins Publishing Company.
- Gillam and Ahmad (2005). "Pattern mining across domain-specific text collections". In Perner and Imiya (Eds.) Proc. of 4th International Conference on Machine Learning and Data Mining in Pattern Recognition (MLDM) 2005, Leipzig, Germany, July 9-11, 2005. Lecture Notes in Artificial Intelligence, Vol. 3587 pp 570-579
- Gillam (2005). "Metadata descriptors: ISO standards for terminology and other language resources". Proc. of 1st International e-Social Science Conference. Manchester, June 2005.
- Ahmad, Gillam and Cheng (2005). "Textual and Quantitative Analysis: Towards a new, e-mediated Social Science". Proc. of 1st International e-Social Science Conference. Manchester, June 2005.
- Gillam, Ahmad and Dear (2005). "Grid-enabling Social Scientists: some infrastructure issues". Proc. of 1st International e-Social Science Conference. Manchester, June 2005.
- Gillam and Ahmad (2005). "Overcoming the Knowledge Acquisition Bottleneck?". Proc. of 7th International conference on Terminology and Knowledge Engineering (TKE) 2005.
- Ahmad, Gillam and Cheng (2005). "Society Grids". Proc. of e-Science All Hands conference (*In press*).

- Ahmad and Gillam (2005). "Automatic Ontology Extraction from Unstructured Texts". Proc. of International Conference on Ontologies, Databases and Applications of Semantics (ODBASE) 2005. Lecture Notes in Computer Science (LNCS) (*In press*).

The 5 published conference papers were each accompanied by presentations; the 2 conference papers currently *in press* will be presented at the respective conferences also. Continued peer-review of the ideas to be encompassed in LIRICS will be undertaken through further publication.

DFKI:

Lirics is briefly described in a paper summarizing activities on standards for language resources at the German conference GLDV 2005. Title of the paper:

"Standardisierung von Sprachressourcen: Der aktuelle Stand". Authors: Thorsten Trippel, Thierry Declerck, Ulrich Heid (all member of DIN NAT AA 6).

Participation of Thierry Declerck (DFKI) at the Lirics Meeting in Paris, March 16-17. Presentation of actual work on extracting morphological relevant descriptors from past and on-going normalization initiatives on morpho-syntactic annotation. The results of the work are already available in a XML Schema. Especially Eagles and Multext East have been considered, but also work done by ISO representatives. Also discussion on the submission of a ISO New Work Item Proposal on syntactic annotation, to be adapted first at the ISO Meeting in Berlin (8-9 April 2005). For this support is provided by INRIA-LORIA.

Bi-lateral meeting DFKI-ILC in Pisa at the 5th Mai. Exchange and organization of morpho-syntactic and syntactic data extracted from former normalization activities and available corpora at both sides. ILC and DFKI discuss the fundamental issues for the submission of the New Work Item on Syntax (SynAF) to the ISO. This NWI will propose a Standard for Syntactic annotation as defined in WP3 of LIRICS.

LIRICS IAG Meeting: Thierry Declerck (DFKI) presents the ISO MAF and SynAF initiatives at the LIRICS IAG meeting in Barcelona (21-22 June 2005).

A bilateral meeting between DFKI and UPF is following the IAG meeting: UPD delivers its works on extracting LIRICS relevant data for the morpho-syntactic and syntactic annotation of Spanish text.

Participation of Thierry Declerck (DFKI) at the 2-day meeting of the SIGSEM Working Group on the Representation of Multimodal Semantic Information and ISO TC37/SC4 Thematic Domain Group of Multimodal Content (TDG3 of ISO TC37/SC4), endorsed by LIRICS. Thierry Declerck presents actual work on the Syntax-Semantic Interface, (activity 6 of the ISO/TC37/SC4/TDG3): “semantic roles and argument structures”. This is highly related to the interface between WP3 and WP4 in LIRICS.

Participation of Thierry Declerck (DFKI) at the ISO TC37/SC4 WG 2, WG 4 AND TDG 2 Meeting in Berlin (8-9 April 2005). The New Work Item Proposal on Syntactic Annotation (WP3 of LIRICS) is discussed and positively supported by a large majority of ISO representatives present at this meeting. Thierry Declerck will be the project leader of this item, corresponding to the WP3 of Lirics. INRIA-LORIA is providing support for the preparation of the official documents. DFKI contributes also to discussion on the ISO MAF proposal

Accepted submission of a Paper about the DIN activities in Germany, where LIRICS is being introduced as the European Project on standardisation. GLDV Conference 2005. Authors: “Thorsten Trippel, Thierry Declerck, Ulrich Heid , Title “Standardisierung von Sprachressourcen: Der aktuelle Stand“. Paper presented by Thorsten Trippel , 1. April 2005

Thierry Declerck and Mihaela Vela submit in June a paper “Linguistic Dependencies for the extraction of domain-specific semantic relations” to the Workshop “**Workshop on Biomedical Ontologies and Text Processing** “ which is part of the 4th European Conference on Computational Biology (ECCB). This paper is being accepted in July.

The paper proposes an interface between syntactic structures and domain-specific ontologies. A topic that is central to the interface between WP3 and WP4 in LIRICS, but also for possible relations between TC37/SC4 and W3C standardisation activities

CNR:

Monica Monachini and Nicoletta Calzolari published a paper: “Initiatives towards the Integration of Lexicons: MILE is Taking Steps Forwards”, Proceedings of the Workshop GLDV Interest Group “Machine Translation” held in Köthen, Germany, 17th of June 2005

9. Exploitation Perspectives

This topic is to be seen under the 7 previous chapters: all these activities are the basis for any future exploitation activity of LIRICS results. Acceptance of current LIRICS technical standards research and development is the best bet on real-life exploitation – in this respect the expectations of the members of the IAG are crucial. Technical exploitation means the implementation of standards in commercial as well as public information infrastructures and language technology environments. Exploitation also means accepted consulting competence for all the members of the LIRICS consortium who distribute the results in their constituencies, in their communities.

Since the nature of the results of the work in the Work Packages of the project is to produce public and open source standards, the concept of exploitation is of course to be seen at the strategic level of global information infrastructures as well as European language industry stakeholders and key players. Companies implementing such standards are expected to be nearer to the needs of their clients since the standards themselves are modelled according to user needs.

In the field of translation technology, for instance, a technology push is urgently needed, since available products in the field (translation memories, MT systems) are based on concepts that are not very new, to say the least. The R&D activities of SYSTRAN are an excellent example. They are experimenting with new methods that complement existing and well proven methods. Semantic annotation is one hope in particular, the LMF project is another, holistic approach that will help MT to innovate its technology base.