LIRICS

Deliverable D7.3bis

Annual Progress Report M1-M12

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	Annual Progress Report M1-M12
Deliverable ID	D7.3bis
Document type	Report
Dissemination level	Public
Contractual date of delivery	M12
Actual date of delivery	31st December 2005
Status & version	Draft
Work package, task & deliverable responsible	WP7, Task7.3, INRIA-Loria
Author(s) & affiliation(s)	Gil Francopoulo
Additional contributor(s)	All LIRICS members
Keywords	Progress report

Document evolution

version	date	version	date
1.0	31st December 2005		
1.1	16th February 2006		

1 General progress and deviation

In the beginning of 2005, one of the first things we did was to set up the LIRICS web site (see http://lirics.loria.fr) in order to be able to share effectively all our working papers, ISO documents etc. Meanwhile, a mailing list has been set up and managed by INRIA.

A lot of technical work was necessary this year in the various ISO tasks, in particular in the writting of the ISO documents.

In June, we organized the first Industry Advisory Group seminar in Barcelona in order:

- to present the outlines of the project;
- to get the foreseen users feedback in terms of needs, demands and questions.

In organizing very early such a seminar, we wanted to be sure that the specifications match their needs and avoid the situation where the results are presented at the end of the project without any possibility of change.

In August, in Warsaw all LIRICS partners actively participated to the ISO plenary meeting. The decisions have been taken to issue[1]:

- A draft for an International Standard for ISO-12620 (data category management) revision
- A committe draft for Morpho-syntactic annotation framework (MAF)
- A working draft for Syntactic annotation framework (SynAF)

2 Progress in WP1

2.1 Progress and deviation

Task 1.1 Training: Most of the LIRICS partners were aware of the ISO process at the beginning of the project but some are new in the ISO process. So, a certain amount of time has been spent by Laurent Romary and Gil Francopoulo in such training explanations. Let it be noted that deliverable D1.1 has been written for this purpose.

Philippe Sébire from INRIA during 16-17 March 2005 meeting explained how to manage a set of data categories by the means of an online connection with the Syntax tool hosted by INRIA in Nancy. This meeting has been completed by a series of mail exchanges between partners on this subject.

Task 1.2 Infrastructure: Some data category registry (DCR) users gave feedback by the means of the mail tool attached inside the DCR. Most of this feedback has been taken into account and the DCR is continuously maintained and improved by the INRIA team with respect to ISO-12620.

Task 1.3 Quality assessement: The task will lead to the development of criteria for assessing the quality of the standards documents being developed in LIRICS.

From an administrative point of view, the role of Surrey PI on LIRICS was transferred to Lee Gillam from Khurshid Ahmad.

---

[1] For the full text, see "http://lirics.loria.fr + event"

Lee Gillam supervised a visiting student, Kim van Zanten, from Maastricht School of Translation and Interpreting between March and June 2005. Kim worked on exploring, understanding and presenting mappings between extant ISO standards of TC 37 based on their content and relationships with other normatively referenced standards. Kim evaluated the application of Plain English and Simplified English to the simplification of language in standards documents, constructed a terminology from these standards that can be used to automatically mark terms in emerging standards documents, used System Quirk to automatically identify terminology within the documents and tested the so-called "principle of substitution" in ISO definitions. Results of this work provided additional UK commentary on the revision of ISO 860 for concept system harmonization.

2.2 Extended report on activities related to WP1, including LIRICS meetings and dissemination

Concerning the **ISO process:**

- A great number of days has been devoted to release the three versions of the Lexical Markup Framework documents: ISO 24613-revision-5, 6 and 7. This document is written by Gil Francopoulo (INRIA) and Monte George (ANSI) with the help of Nicoletta Calzolari (CNR).

- Laurent Romary (INRIA) worked on the revision of ISO-12620 (data category management) in order to prepare the Committee Draft version.

Concerning the **quality assessement**, the following papers were published and presented:

- Gillam, Tariq and Ahmad (2005) "Terminology and the Construction of Ontology". Terminology 11(1), pp55-81. John Benjamins Publishing Company.

- 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.

- Lee Gillam presented the paper "Pattern Mining Across Domain-Specific Text Collections" (Gillam and Ahmad) at the 4th International Conference on Machine Learning and Data Mining in Pattern Recognition (MLDM), Leipzig, Germany, July 9-11, 2005. The paper is published in Lecture Notes in Artificial Intelligence (LNAI), volume 3587. ISBN: 3-540-26923-1.

- The paper "Overcoming the Knowledge Acquisition Bottleneck?" (Gillam and Ahmad) was presented by Khurshid Ahmad at the 7th International conference on Terminology and Knowledge Engineering (TKE 2005, Copenhagen, 17-18 August) and is published in the proceedings, ISBN: 87-91242-46-0.

- Lee Gillam attended the ISO TC37 activities in Poland, 22-26 August 2005 as a Principal UK Expert in the UK delegation. This was additional discussions regarding the UK position on the various emerging standards at two meetings of the UK shadow committee to TC37 (TS/1 and its subcommittees) on 3 October and 12 December, and in attendance at the meeting of the UK ICT co-ordination and strategy committee (ICT/-) on 8 December.

- Lee Gillam attended the UK's *e-Science* All Hands Meeting, Nottingham, 19-22 September 2005. The paper "Society Grids" (Ahmad, Gillam and Cheng) was presented by Khurshid Ahmad and is published in the proceedings, ISBN 1-904425-53-4.

- Lee Gillam gave an invited presentation on "Metadata, Terminology and Ontology" at the Information Management and *e-Social Science* Workshop, 5 October 2005, Lancaster University. A digital recording of this presentation is available at: http://redress.lancs.ac.uk/resources/Lee%20Gillam/Metadata%20Terminology%20and%20Ontology/Metadata-Terminology-and-Ontology.html

- The paper "Automatic Ontology Extraction from Unstructured Texts" (Ahmad and Gillam) was presented by Khurshid Ahmad at the 4th International Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE 2005), 31 Oct to 4 Nov 2005 in Cyprus, and is published in Lecture Notes in Computer Science (LNCS) 3761, pp1330-1346. ISBN 3-540-29738-3

- Lee Gillam attended the World Summit on the Information Society (WSIS), in Tunis, 14-18 November 2005 by invitation of the Nagaoka University of Technology, Japan. Possible opportunities with the Wikimedia Foundation are under discussion.

During the reporting period, LIRICS participants (Budin, Gillam, Romary) undertook further activities in relation to the extension of coverage of language codes of ISO 639. The metamodel approach adopted in LIRICS, as resulting from the SALT project and demonstrated in the "Terminological markup framework" of ISO 16642 has been applied to ISO 639 and should emerge in the principles and methods document of ISO 639-4 with a variety of data categories (metadata identifiers) under construction. The UK delegation submitted an ISO 639-6 document for committee draft (CD) ballot.

3 Progress in WP2

Progress and Deviations

For WP2 no deviation has to be reported with respect to the original planning in the TA.

Task 1. In the reporting period, CNR-ILC has started the analysis and evaluation of major past and ongoing efforts dedicated to standardization in the field of NLP lexicons, with the aim of collecting a set of linguistic information for populating the lexical layers of the data model. A draft unified inventory of lexical information, unified descriptors, short descriptions (a kind of *pre* Data Categories) has been identified as a coherent input to the Data Category Registry, to be formulated within Task 2. This activity has been carried out by (i) relying on the outcomes of the past projects EAGLES, PAROLE, MULTEXT, MULTEXT-East, SIMPLE and ISLE, (ii) searching for transversal coherence between lexicon and text annotation (WP3) and (iii) considering parallel standardization initiatives in the speech community. The results will be reported in more detail in the corresponding deliverable due by the end of September 2005.

Closely related to this work are the activities undertaken within the ISO-TC37/SC4/ WG4, where a task force for drafting a Data Category Registry is being composed. CNR-ILC plays a key role in this task force.

CNR-ILC, with the contribution from the partners involved in this Task, has continued the activities concerned with the formulation of a set of Data Categories needed for populating the different layers of the lexical data model to be designed in *Task 2*. This has resulted in the Deliverable 2.1 (D2.1, delivered in October 2005) that proposes a maximum set of *candidate* lexical data categories subdivided along the different layers of linguistic representation. As an added-value of this purely linguistic activity, we tried to offer an abstract model with an XML formalization and instantiation of the data. This effort is crucial in order to make this bulk of information more computationally manageable, thus helping and fostering next phases of Data Category creation. These Pre–Data Categories will receive the necessary adjustments

and modifications, after discussion among the LIRICS partners and feedback from connected LIRICS activities, revisions coming from ISO experts and, hopefully, suggestions issued from end-users which will test adequacy, appropriateness and coverage with respect to application and language-specific considerations.

In agreement with D3.1, this deliverable is to be seen as a document in-progress. The work of defining Data Categories for lexical description is, by its very nature, to be seen as an incremental one, since it will proceed step by step with the specification of the meta-model. Only the combination of the meta-model (its layers and components) with the Data Categories will concur to define the Lexical Model and the Lexical Mark-up Language necessary to instantiate it.

The identification and definition of Data Categories has many contact points with objectives defined in ISO/TC37, where a total revision of the ISO 12620 Data Category Registry is underway. The LIRICS Data Categories will serve as a coherent input to this Registry and will directly enrich it. They will be formalized according to the ISO standards, will be represented and ruled in conformance to ISO procedures, and, at the end of the project, will be maintained as an official ISO global resource.

As a matter of fact the Pisa team took actively part in the ISO TC37/SC4 plenary meeting in Warsaw (21.08.05 - 26.09.05). In order to ensure constant links between LIRICS and ISO a formal resolution has been approved for the constitution of a WG4 sub-group for the purpose of designating generic data category sets for alignment with the levels of the metamodel (Members: Monica Monachini (co-chair), Gil Francopoulo (co-chair), Thierry Declerk)

Also related to the activities performed for this task, we can mention the investigation of the possibility of devising harmonized data categories for semantic roles, a topic covered by WP2 and WP4.

Task 2. CNR-ILC has started the formulation of a Lexical Mark-up Framework, i.e. (i) an abstract meta-model for lexicons as a set of structural nodes relevant for lexical description and (ii) a flexible environment, enabling specific implementations of user-defined mark-up languages (called LML) on the basis of common Data Categories.

A set of entries from existing lexicons have been mapped to the model to prove that it is able to represent many best practices and achieve unification.

Activities started very soon: a draft document was presented in Berlin in April, in conjunction with the ISO TC37/SC4 meeting, and to the LIRICS Industrial Advisory Board in Barcelona (21-22 June). Industrials have emphasized the importance of standards for lexicons, since they give credibility to products and are of relevance for high-quality lexical resources. Some industrials are already proceeding step-by-step with the standard framework emerging within WP2.

The LIRICS partners are carrying out the activities foreseen within this WP step by step with the ISO-TC37/SC4/WG4 (CNR-ILC is the Convenor of this WG), where the requirements of a standard for both NLP lexicons and Machine Readable Dictionary are exchanged between the two communities and a common core abstract model is being defined. In particular, within the ISO group, synergies are being created at international level between American and Asian experts of the field.

The following versions of the ISO working draft have been produced in conjunction with the two ISO editors: Gil Francopoulo (INRIA-Loria) and Monte George (ANSI):

19 March 2005: Lexical Markup Framework version-5

15 June 2005: Lexical Markup Framework version-6

10 August 2005: Lexical Markup Framework version-7

The LIRICS partners are carrying out the activities foreseen within this WP step by step with the ISO-TC37/SC4/WG4 (CNR-ILC is the Convenor of this WG and Gil Francopoulo is a co-project leader).
In particular, special attention has been devoted to the representation of fixed, semi-fixed or flexible multiword expressions patterns. An ad-hoc extension to the NLP lexicon meta-model has been foreseen and described in an informative Annex of the document.
As a by-product of its involvement in WP4, related to the definition of overall model for representation of semantic information, CNR-ILC has investigated the compatibility of the semantic information model with the lexico-semantic representation model, in order to ensure a shared approach between the two.

For these two tasks, no deviation has to be reported with respect to the original planning in the technical annex.

2. Extended Report on activities related to WP2, including LIRICS meeting and dissemination:

Lirics Meetings:

- Participation of Monica Monachini (CNR-ILC) at the kick-off meeting of LIRICS, 24.1.2005, in Luxembourg.

- Organization of a bilateral meeting between CNR-ILC – MPI, 15.2.2005, where Monica Monachini and Nilda Ruimy (CNR-ILC) presented the architecture of the PAROLE-SIMPLE lexicons and Mark Kemps-Snijders showed the functionalities of the LEXUS tool, designed to describe LMF lexicons conformant to the model emerging in WP2.

- Organization of a WP2 internal meeting, 16.2.2005 in Pisa, with the participation of Nicoletta Calzolari, Monica Monachini, Claudia Soria (CNR-ILC), Gil Francopoulo (LIRICS Technical Manager), Nuria Bel (IULA-UPF) and Mark Kemps-Snijders (MPI). Establishment of the basic structure of the meta-model for lexicons (core model + extensions) and definition of presentational strategies in the related document have been defined.

- Participation of Nicoletta Calzolari and Monica Monachini (CNR-ILC) at the LIRICS Meeting in Paris, March 16-17. Presentation of progress of work within WP2, discussion about synergies and convergences with the work being undertaken within WP3. Presentation of the standard core model for lexicons and the extensions for NLP lexicons.

- Bi-lateral meeting CNR-ILC – DFKI, 5.5.2005, in Pisa. Exchange and organization of morpho-syntactic and syntactic data extracted from former normalization activities and available corpora on both sides. ILC and DFKI discussed 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. The minutes are available on the web site under http://lirics.loria.fr + Events.

- CNR-ILC participated to the LIRICS Industrial Advisory Board Meeting, in Barcelona (21-22 June 2005).

- Participation of CNR-ILC at the LIRICS Consortium Meeting, 22.6.2005, in Barcelona. Monica Monachini presented the work progress within WP2. In relation to morpho-syntax and the bulk of information coming from past normalization initiatives in the form of flat lists, CNR-ILC has designed a tool which allows the creation of a database of PoS with relevant morphological features for a given language and/or a specified project. The tool, which is an addition to the LIRICS work, allows the definition of constraints between features and values in the form of declarative rules that combine POS and morphological

features for a specified language and supports the definition of hierarchies between attributes and values while designing Data Categories. The resulting admitted combinations and constraints are saved in a database and can be exported in XML.

- Participation of Nicoletta Calzolari, Monica Monachini and Claudia Soria (CNR-ILC) at LIRICS meetings during the plenary meeting of ISO TC37/SC4 in Warsaw (21.08.05 - 26.08.05)

- Participation of Claudia Soria at the LIRICS WP4 Workshop in Nancy (8-9.12.2005).

- Organization of a WP2 internal meeting, 23-24.11.2005 in Pisa, with the participation of Claudia Soria, Valeria Quochi (CNR-ILC) and Gil Francopoulo (LIRICS Technical Manager). Establishment of the basic structure of the meta-model for representation of Multiword expressions.

Events related to LIRICS

- Nicoletta Calzolari (CNR-ILC) participated at the ISO TC37/SC4 WG 4 Meeting as Convenor of the group, 8-9.4.2005 in Berlin.

- 27.10.2005: Monica Monachini (CNR-ILC) participated to the DIAM Commission in Rome at the premises of UNI, where the candidature of Italy as P-member in ISO TC37/SC4 has been proposed for approval to the following UNI plenary council and CNR-ILC has been selected as reference expert (also due to actual efforts and active participation through LIRICS in ISO TC37/SC4 WGs).

- Participation of Nicoletta Calzolari, Monica Monachini and Claudia Soria (CNR-ILC) at the plenary meeting of ISO TC37/SC4 in Warsaw (21.08.05 - 26.08.05), where CNR-ILC is the Convenor of WG4.

- 23-24.11.2005 in Pisa: organization of a meeting of the sub-group of the ISO/TC 37/SC 4/WG 4 related to MultiWord Expressions and approved by formal resolution adopted at the SC 4 plenary meeting of ISO/TC 37/SC4 in Warsaw, 2005-08-25. The meeting has seen with the participation of Claudia Soria, Valeria Quochi (CNR-ILC) and Gil Francopoulo (LIRICS Technical Manager). Definition of the basic structure of the Extension for NLP related to Multiword expressions.

Dissemination

- Nicoletta Calzolari, invited speaker at the International Workshop on Generative Lexicon, 19-21.5.2005, in Geneva, presented the standardization activities being carried out in LIRICS WP2.

- Nicoletta Calzolari, invited speaker for the IULA 10th Anniversary, 16-17.6.2005, in Barcelona, presented the standardization activities being carried out in LIRICS WP2

- A paper by Monica Monachini and Nicoletta Calzolari, “Initiatives towards the Integration of Lexicons: MILE is Taking Steps Forward”, was accepted for presentation at the Workshop of GLVD Interest Group “Machine Translation”, 17.6.2005, Kothen, Germany. The paper has been also accepted for publication.

- In the framework of the language resource production activity of the ELRA PCom, it was decided that production of new language resources, especially those resulting from merging existing resources, will be performed according to the recommendations issued by LIRICS.

- 27.10.2005, DIAM Commission in Rome, at the premises of UNI: Monica Monachini (CNR-ILC) presents ILC-CNR activities connected to standardization, describing organization and work items within ISO TC/37 SC4, ILC actual involvement within various ISO sub-groups/thematic domain group and presenting LIRICS efforts and objectives.

- 27-28.7.2005: Nicoletta Calzolari gives an invited talk at ECOR (European Centre for Ontological Research) Inaugural Meeting, Saarland University, Saarbrucken, on the topic "Language Technology and the Semantic Web". The goal of this presentation is to stress the need of close collaboration of the language technology and semantic web communitites on issues related to standards.

- 20.9.2005: Nicoletta Calzolari gives an invited talk on the topic "Language resources: towards a framework for content interoperability" during the 9th Congress of AI*IA (Italian Association for Artificial Intelligence), Milan, Italy, where she presents the goals of LIRICS and ISO.

- 28-29.11.2005: Nicoletta Calzolari gives an invited talk on the topic "Language Resources: Priorities and Challenges" during the Symposium on Natural Language Processing and Image Recognition, Kyoto, Japan. In this talk the availability of standards is presented as a crucial challenge for advancement of the field.

- 19-21-12-2005 Nicoletta Calzolari gives an invited talk on the topic "Linguistica Computazionale: sinergie con il progetto DLM" during the Workshop "Lessicologia e Metalinguaggio", Macerata, Italy. The talk addresses the importance of standards for synergies across Italian NLP lexicons.

4 Progress in WP3

4.1 Progress and deviation

- DFKI has started the overview of past initiatives that were dedicated to standardization in the field of morpho-syntax. This resulted in a first stable version of Deliverable 3.1 (finalized in October 2005). Due the amount of material found that is still to be described and integrated in the kind of document proposed by Deliverable 3.1, it was decided to propose in the first half of 2006 an new extended and revised version. The actual version of Deliverable 3.1 is at the same time the basis for the ISO standardisation work on syntactic annotation that is briefly discussed in the next dot. Related to this activity, DFKI began in the actual ISO work on standardization of morpho-syntactic annotation (the MAF document), participating as well at a MAF meeting in Berlin (8-9 April), and a member of DFKI is now as well in the editorial committee for the next step of the MAF document, the so-called CD ballot.

- Concerning the standardization of syntactic annotation, which is starting from scratch at ISO level, DFKI presented a first draft document for a so-called ISO New Work Item at the ISO meeting in Berlin (8-9 April 2005). The group of international experts present at this meeting approved basically the draft, which was accommodated to the comments made at this occasion, and this group supported a submission to ISO SC4/TC37 as a New Work Item. In May DFKI submitted the proposal for a New Work Item, together with a call for participation, to the secretary of ISO TC37/SC4, which distributed it to all the national standardization bodies that are related to the activities of ISO TC37/SC4. We can anticipate in this report that during the month of July a majority of national bodies has responded positively to the proposal and that the New Work Item is now officially an ISO document. So here we managed to tackle the first step toward an ISO standard for syntactic annotation and we are thus quite optimistic that this standardization activity will go quite far within the time scope of LIRICS. At the ISO TC37/SC4 plenary meeting in Warsaw (21.08.05 - 26.09.05) the new work item (NWI) SynAF (a central task of WP3) has been presented by Thierry Declerck (DFKI) to the experts of the various national

bodies present at the special session dedicated to this NWI, which has been approved by ISO in July 2005. In a second session dedicated to this issue, foreseen mainly for the LIRICS partners, detailed comments have been discussed, which form the base for the next step, the redaction of the Working Draft (WD) of SynAF, under the direction of Thierry Declerck. A first version of this document will be delivered at the end of 2005/beginning of 2006 and discussed at the ISO meeting in Jeju (Corea) between the 19.1 and the 21.1.2006.

No deviation from the original planning of WP3 has to be reported.

4.2 Extended Report on activities related to WP3, including LIRICS meeting and dissemination:

4.2.1 Lirics Meetings

- Participation of Thierry Declerck (DFKI) at the kick-off meeting of LIRICS, 24.1.2005, in Luxembourg

- 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 an 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 May. Exchange and organization of morpho-syntactic and syntactic data extracted from former normalization activities and available corpora at both sides. ILC and DFKI discussed 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 followed the IAG meeting. UPF presented and delivered a contribution to D3.1: Evaluation of initiatives for morpho-syntactic and syntactic annotation, with a survey of current practices for morpho-syntactic and syntactic annotation of texts.

- Participation of Thierry Declerck (DFKI) at LIRICS meetings during the plenary meeting of ISO TC37/SC4 in Warsaw (21.08.05 - 26.08.05)

- Bi-lateral meeting Tilburg-DFKI in Tilburg, discussings also joint issues for WP3 and WP4 (the interface between syntactic and semantic annotations). The meeting took place in Tilburg (16-17.11.2005).

- Participation of Thierry Declerck at the LIRICS WP4 Workshop in Nancy (8-9.12.2005).

- Bilateral meeting DFKI-Tilburg, this time mainly related to issues of WP4 and how DFKI can contribute to sub-goals of WP4. Place: Saarbruecken. Date (14-16.12.2005).

4.2.2 Events related to LIRICS

- Thierry Declerck (DFKI) worked as a member of the DIN mirroring group NAAT 6

- 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 presented 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.

- DFKI organizeed the meeting of "DIN NAT AA 6 Sprachressourcen", the German Mirror Committee of TC37/SC4 on 13.5.2005. LIRICS was discussed in detail and possible collaboration with the eContent project "EurotermBank" was addressed by a DIN Member being also partner of the eContent project.

- 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) was 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 will provide support for the preparation of the official documents. DFKI contributed also to discussion on the ISO MAF proposal

- Participation of Thierry Declerck (DFKI) at the plenary meeting of ISO TC37/SC4 in Warsaw (21.08.05 - 26.08.05). Presentation of the ISO SynAF initiative (a task of WP3). Within this plenary meeting, participation at the ISO TDG3 session (TDG3 is dedicated to the Representation of Multimodal Semantic Content, and Thierry Declerck is in charge of the item dedicated to the relation between argument structures and semantic roles).

- Thierry Declerck is co-chair of a workshop on "Semantic Annotation" held as a satellite event of the 4th International Semantic web conference in Galway (6-11.11.2005). The workshop, held on the 7th of November, is endorsed by LIRICS, since standards play a siginificant role in this workshop.

4.2.3 Dissemination

- Accepted submission of a Paper about the DIN activities in Germany, where LIRICS is being introduced as the European Project on standardization. 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 on Biomedical Ontologies and Text Processing** which is part of the 4th European Conference on Computational Biology (ECCB). This paper was accepted in July. The paper proposes an interface between syntactic structures and domain-specific ontologies. This is a topic that is central to the interface between WP3 and WP4 in LIRICS, but also for possible relations between TC37/SC4 and W3C standardization activities.

- Thierry Declerck (DFKI) presents the goals of LIRICS and ISO TC37/SC4 in combination with the standardisation work on Multimedia content analysis and indexing (the MPEG initatiative at ISO) at the ESSLLI summer school in Edinburg (06.08.05 bis 20.08.05) in a lecture called "NLP for Multimedia Applications").

- Thierry Declerck, in his quality as a co-chair, presents briefly the goal of LIRICS to the participants of a workshop on "Semantic Annotation" held as a satellite event of the 4th International Semantic web conference in Galway (6-11.11.2005). The goal of this presentation is to stress the need of close collaboration of the language technology and semantic web communitites on issues related to standards.

- Thierry Declerck presents a paper by Mihaela Vela and himself “Linguistic dependencies for the extraction of domain-specific semantic relations” at the Workshop on Biomedical Ontologies and Text Processing which is part of the 4th European Conference on Computational Biology (ECCB). This paper was accepted in July 2005. The paper proposes an interface between syntactic structures and domain-specific ontologies. This is a topic that is central to the interface between WP3 and WP4 in LIRICS, but also for possible relations between TC37/SC4 and W3C standardization activities.

- Thierry Declerck gives an invited talk at JRC in ISPRA (5.10.2005) on the topic "Event extraction from text", and he presents extensively the goals of LIRICS and ISO.

- Thierry Declerck is co-author of a German (with Thorsten Trippel and Ulrich Heid) paper describing the DIN. ISO and LIRICS goals, that has been published in the GLDV- Journal for Computational Linguistic and Language Technology (Band 20 – Heft 2 – Jahrgang 2005 – ISSN 0175-1336). Title of the paper: “Sprachressourcen in der Standardisierung”.

5 Progress in WP4

5.1 Progress and deviation

In the reporting period, UTiL has started: (1) the overview of recent initiatives in the area of standardisation of semantic annotation schemes; and (2) the formulation of methodological principles for standardization in semantic annotation. Closely related to this work are the activities undertaken during this period in the SO TC 37/SC 4 Task Domain Group on Semantic Content Representation (TDG 3), in which UTiL has played a key role and participated in meetings in Tilburg in January and in Berlin in April. The results of these activities will be reported in more detail in the corresponding deliverable due by the end of September 2005.

No deviation from the original planning of WP4 has to be reported.

5.2 Report on activities related to WP4, including LIRICS meetings and dissemination:

5.2.1 Lirics Meetings:

- Participation of Harry Bunt (UTiL) at the kick-off meeting of LIRICS, 24 January 2005, in Luxembourg.

- Participation of Harry Bunt (UTiL) at the LIRICS project meeting in Paris, 16-17 March 2005.

- LIRICS IAG Meeting: Participation of Harry Bunt (UTiL) in the meeting with the LIRICS Industrial Advisory Group in Barcelona, 21-22 June 2005.

- Bilateral meeting between MPI and USFD at M10.

5.2.2 LIRICS-related events

- Participation (as general chair) of Harry Bunt (UTiL) in the two-day meeting of the ACL-SIGSEM Working Group on the Representation of Multimodal Semantic Information and the ISO TC37/SC4 Thematic Domain Group on Semantic Content Representation (TDG3 of ISO TC37/SC4), Tilburg, 10-11 January 2005, endorsed by LIRICS. Harry Bunt presented work on the definition of dialogue acts (activity 2 of the ISO/TC37/SC4/TDG3). Thierry Declerck (DFKI) presented work on predicate-argument structures, which is closely related to the interface between WP3 and WP4 in LIRICS.

- Participation (as general chair) of Harry Bunt (UTiL) in the 6th International Workshop on Computational Semantics, Tilburg, 12-14 January 2005. Susanne Salmon-Alt and Laurent Romary (INRIA) presented recent work on co-reference annotation, which is one of the topics in LIRICS WP 4. Harry Bunt presented new work on the representation and annotation of quantification and modification structures in natural language using typed feature structures.

- Participation of Harry Bunt (UTiL) at the ISO TC37/SC4 WG 2, WG 4 AND TDG 2 Meeting in Berlin, 8-9 April 2005.

- Participation of Harry Bunt and Yann Girard (UtiL) in DIALOR'05, the 9th Workshop on the Semantics and Pragmatics of Dialogue, in Nancy, 9-11 June 2005. They presented new work on the design of generic multidimensional taxonomies for dialogue act annotation.

- Working visit of Amanda Schiffrin to INRIA-LORIA in Nancy, 20-24 June 2005, in order to become better acquainted with the LIRICS project's background in standardization activities in ISO TC 37, in particular in TC 37/SC 4, and with the software facilities developed at LORIA for supporting the construction and archiving of data categories.

5.2.3 Dissemination

A paper has been accepted about issues in the design of generic multidimensional taxonomies for dialogue act annotation. The paper, jointly authored by Harry Bunt and Yann Girard (UtiL), was presented at the DIALOR'05 conference in Nancy, 9-11 June 2005, and published in the DIALOR proceedings: Harry Bunt and Yann Girard, "Designing an Open, Multidimensional Dialogue Act Taxonomy"; in Gardent, C. and Gaiffe, B (eds.), *Proceedings of the 9th Workshop on the Semantics and Pragmatics of Dialogue*, pp. 37-44.

6 Progress in WP5

6.1 Progress and deviation

WP5 participants have worked on tasks 5.1, 5.2, 5.3, and 5.4.

Task 5.1 is currently working on APIs (Application Programming Interfaces) following the LIRICS standards for the ISO Data Category Registry (WP1), NLP Lexica (WP2), Morpho-syntactic annotations (WP3).

The work on DCR API and the accompanying DCR Usage platform are tackled by MPI, (D5.1.A v1.0 completed in M6) whereas USFD has focused on the Morpho-syntactic annotations API (D5.1.C v1.0 completed in M12). USFD and MPI are designing collaboratively the NLP Lexica API (D5.1.B v1.0 completed in M12). The APIs will be defined following the guidelines from deliverable D1.1.

Task 5.2 has started and MPI have worked there on D5.2.A DCR Reference Implementation.

Task 5.3. also started in M6 and USFD is currently designing the LIRICS service integration platform, first version of which will be available at M18.

Task 5.4 is focusing on the Data Category Usage platform, first version of which is expected at M18 (D5.4)

No deviation from the original planning of WP5 has to be reported.

6.2 Report on activities related to WP5, including LIRICS meetings and dissemination

6.2.1 LIRICS meetings

- Participation of Kalina Bontcheva (USFD) at the kick-off meeting of LIRICS, 24 January 2005, in Luxembourg.

- Participation of Kalina Bontcheva and Julien Nioche (USFD) at the LIRICS project meeting in Paris, 16-17 March 2005.

- LIRICS IAG Meeting: Participation of Julien Nioche (USFD) in the meeting with the LIRICS Industrial Advisory Group in Barcelona, 21-22 June 2005.

6.2.2 LIRICS related events

The LIRICS standards under development have been presented at a meeting of the SEKT project on semantically enabled knowledge technologies (www.sekt-project.com), with a view to making the relevant ISO-compliant linguistic resources, once the standards will have reached sufficient maturity.

7 Progress in WP6

7.1 Progress and deviation

A deviation from the original planning of WP6 is reported: presentation of version 1.0 of deliverable D6.1 was deferred from month 2 to month 8, because the evolution of the project activities at the strategic level is to be seen in close interaction with and inseparable from any sustainable dissemination plan. This is even more true for exploitation activities to be planned. So the first version of D6.1 was a first guess that co-evolved with the activities in all WPs.

All LIRICS events that have been held since the beginning of the project were obviously used for group discussions and updates on the planning and design of D6.1 (e.g. the Paris event in March 2005, the Berlin meetings in April 2005, the August meetings in Warsaw 2005, etc.). Thus, version 1.0 of D6.1 mirrors the results of this continuous work. So the justification for a new delivery date of D6.1 emerged from this approach.

7.2 Report on activities for WP6 dissemination and exploitation

7.2.1 Word on D6.1

The structure of Deliverable 6.1 is obviously a mirror of the different kinds of activities in this work package. Important is the overall strategy that includes each individual item listed here and the interaction between all of them, e.g. events and publications to promotion of LIRICS in research communities and in corporate environments, the work within standardization communities and the impact of this work on the product strategies of language technology industry players, etc.

D 6.1 has the following structure in terms of 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

7.2.2 Industry Advisory Group

In item 5 b (below) the strategy concerning the IAG will be explained in more detail.

7.2.3 Public dissemination and promotion by the project website

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 D6.

7.2.4 Discussion forum

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.

7.2.5 Regular events of different types

7.2.5.1 Internal workshops

Internal workshops are used, in addition to discussing the actual technical work items and details of document writing, 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.

7.2.5.2 Workshops with the Industry Advisory Group

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.

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

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

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.

Thierry Declerck presented 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 was the relation between TC37/SC4 and the MPEG-7 Standard for the description of multimedia content representation).

7.2.5.4 Other events

Nuria Bel has presented LIRICS at the Internet Global Conference that was held in Barcelona (6-10 June 2005). The title of the presentation was:" Basis and Motivation for an open linguistic infrastructure" (see http://www.igcweb.net/default.php).

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 section 1 of this report. Such presentations (as well as the publications listed under section 8) will continue to be an important strategy for the success and the sustainability of the project's mission.

7.2.5.5 Future Activities, future events

Thierry 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 Genoa 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 7.2.6.1).

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.

7.2.6 Special strategy concerning internationalization/localization industry communities

7.2.6.1 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.

7.2.6.2 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 that a LIRICS workshop at a LISA event in 2006 will be organized.

7.2.6.3 Co-operation with Other groups

Other institutions such as the Localization Institute, the Localization 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 the merging of existing resources, will be performed according to the recommendations developed by the LIRICS project.

7.2.7 Standardization bodies

7.2.7.1 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 cleary 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 standardization process in relation to four 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

7.2.7.2 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.

8 Progress in WP7

8.1 Progress and deviation

Financial coordination: Virginie Tessier (INRIA) established and maintained the financial records, particularly the payment of the EC advance amounts.

Administration: The INRIA team worked to prepare the LIRICS kick-off meeting in Luxembourg, 24 January 2005. Gil Francopoulo prepared two sets of slides to present the

project: a short version and a long version. The short version has been used by Laurent Romary to present the project[2].

Project Coordination: In order to provide a good management the followings tools have been set up:

- The LIRICS web site that is used internally and externally by the partners in order to share technical information.

- The LIRICS mailing list in order to ask technical questions, push news, etc.

- The management board that is maintained by INRIA and regularly sent to the partners. This is not an official LIRICS deliverable. This is MS-Word array that sums up dynamically where we are with respect to the LIRICS time frame as stated in the technical annex. The purpose of this document is to allow each partner to know where he is and where the other partners are. In order to ease reading, the covered period is not the whole project period: only the past and the next months are covered. The array is structured as follows: for each deliverable, there is a line. All lines are chronologically ordered with respect to the delivery date. Each line is composed of the following cells: date, deliverable identifier, deliverable full name, author, status with possibly comments.

The following LIRICS technical meetings have been organized by INRIA:

- 24 January 2005 in Luxembourg after the project presentation to the EC officers.

- 16-17 March 2005 in AFNOR offices in Paris. Gil Francopoulo presented the slides of the files LMF and morphoSyntacticProfile.
  (see http://lirics.loria.fr + Events + LIRICS meeting March 16-18 + LMF)
  (see http://lirics.loria.fr + Events + LIRICS meeting March 16-18 + Morpho-syntactic profile)

8.2 Extended report on activities related to WP7, including LIRICS meetings and dissemination

The INRIA team participated at:

- the lexicon meeting in Pisa hosted by CNR on 16 February 2005. Gil Francopoulo wrote the minutes.
  (see http://lirics.loria.fr + Events + LIRICS meeting Pisa February 16).

- the ISO-TC37/SC4 WG 2, WG 4 meetings in Berlin hosted by MPI, on 8-9 April 2005. Within the LMF project, Gil Francopoulo presented the slides "Opinion regarding to four different points about the work in progress".
  (see http://lirics.loria.fr+ Events + ISO/TC 37/SC4 Working group meetings April 8-9)

- the Industry Advisory Group meeting in Barcelona hosted by UPF on 21-22 June 2005. Gil Francopoulo presented the slides "Data Category Registry, LMF, TMF etc. What is the situation?"
  (see http://lirics.loria.fr + Events + LIRICS Industry Advisory Group meeting).
  Gil Francopoulo wrote the minutes.
  (see http://lirics.loria.fr + Events + LIRICS Industry Advisory Group meeting)

- the LIRICS meeting in Barcelona on 22 June 2005.

[2] The long version is for the LIRICS partners in case they need to pick slides. These two versions are available on the LIRICS web site.

- In the ISO plenary meeting in Warsaw, Laurent Romary chaired the ISO-TC37/SC4 sessions.