



LIRICS

Deliverable D7.4Bis

Annual Progress Report M12-M24

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 M12-M24
Deliverable ID	D7.4Bis
Document type	Report
Dissemination level	Public
Contractual date of delivery	M24
Actual date of delivery	31 st December 2006
Status & version	
Work package, task & deliverable responsible	WP7, Task7.3, INRIA-Loria
Author(s) & affiliation(s)	Gil Francopoulo
Additional contributor(s)	Adam Funk, Harry Bunt, Monica Monachini, Nuria Bel, Lee Gillam, Thierry Declerck
Keywords	Progress report

Document evolution

version	date	version	date
1.0	31st December 2006		
1.1	12th February 2007		

General progress and deviation

The main actions and events for this period were:

- LIRICS mid-term review
- LREC conference
- Production and CD ballot for LMF
- Annual plenary ISO meeting in Beijing

Some slight deviations in the software tasks have to be reported.

1 Progress in WP1, infrastructure for standards development and quality assurance

1.1 Progress and deviation

Task 1.1 training and infrastructure for developing standards

Together with partners from DFKI and CNR-ILC, Gil Francopoulo (INRIA) used and tested 'syntax' software that manages the data category registry. This work has been conducted in order to set up a basic set of values for the morpho-syntactic data categories for European languages.

A new version of the software dedicated to manage the data category registry is scheduled. MPI is in charge of the development of this piece of software and Sue Ellen Wright from ANSI is gathering the requirements for this new version. Gil Francopoulo provided in December 2006 a documents about the drawbacks of the current version, with a set of recommendations for the new version. Let's note that until the next version be fully functional, the current version is on line and functional.

Task 1.3 quality assessment

Mr Neil Newbold took up his post of LIRICS Research Assistant (RA) with effect from 3 January 2006. He has been continuing work undertaken in the first period of the project with specific reference to how *Quality Assurance* relates to document management and to terminologies and terminology products. The work covers the integration and use of supporting resources and components for the standards development process, including: a Plain English thesaurus; ISO TC 37 terminology via ISO 16642; syntactic annotation; and readability metrics. This is specifically relevant to T1.3 and WP5, with the intention of providing an assistive tool to authors of standards based around LIRICS efforts. A round of evaluation has been undertaken between Surrey and INRIA to assess potential for automation of the Plain English task in isolation and the integration of further components and tasks, and the inter-dependency between tasks and standards, will be assessed in the following periods.

Neil Newbold continued as LIRICS Research Assistant (RA) to the end of October, at which point he began working on Surrey's project on automatic annotation of video content - REVEAL - Recovering Evidence from Video by Fusing Video Evidence Thesaurus & Video Meta-Data, EPSRC GR/S98450/01. Use of LIRICS project results are under consideration in REVEAL as a result of this short contract, as envisaged in the original project outline. Neil will return to finalise efforts on LIRICS in February 2007. Work in this second period, with specific reference to *Quality Assurance* has included the development of components for GATE that annotate known terms imported via an ISO 16642 conformant format, that provide aspects of

automatic term discovery, and that control the application of annotations where embedded annotations are concerned such that (components of) known terms are not considered for simplification. Efforts in the final period will focus on updating the terminology resource and evaluation against the full range of emergent standards, with the final deliverable for this task intended to be suitable for use against further standards. Results from elsewhere in the project will be taken account of.

It should be noted also that the Surrey Morphology group from Surrey participated actively to the LMF technical discussions concerning definitions, item and arrangements strategies and bibliography.

1.2 Extended report on activities related to WP1, including LIRICS meetings

A two days meeting (20,21 Feb 2006) has been organized in Paris between DFKI, CNR-ILC and INRIA to work together on the data category registry.

Dr Gillam incorporated discussion of aspects of the LIRICS project in relation to his newly-scoped MSc course module (55 students) on "Challenges for Computing Professionals" with regard for issues of data protection, standards and standards bodies and professional ethics. The course included a guest speaker from the British Library.

GeoLang Ltd, a company to develop products and services oriented around ISO 639, particularly part 6, was formed with Dr Gillam as a director.

A meeting was held with Gerard Meissjen of OmegaWiki to discuss a community-portal for language documentation of all kinds (8 December). OmegaWiki have expressed an intention to make full use, wherever possible, of the standards that emerge from the LIRICS project, including early uptake.

2 Progress in WP2, NLP lexica

2.1 Progress and deviation

Task 1.

In the reporting period, 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 being designed in *Task 2*.

In this task, the major outcome is the cooperation between WP2 and WP3 for establishing a harmonized definition of data categories for the lexicon and morpho-syntactic annotation. A meeting was held in Paris (20-21.2.2006) between Lirics/ISO members where a consistent set of morphosyntactic data categories has been decided and inserted in the ISO Morpho-syntactic Profile. As a consequence of this, a new version of the Deliverable 2.1 will be produced where the candidate set of lexical information proposed will be aligned along the different layers of the Morphosyntactic profile (PoS, MorphoFeatures, Case, FormRelated ...).

As far as revisions coming from ISO experts and suggestions from end-users are concerned, CNR-ILC, thanks to the involvement in international projects, is proposing the emerging set of LIRICS-ISO lexical data categories to NEDO and BootStrep partners in the aim of testing the adequacy and appropriateness and extending their coverage with respect to language-specific considerations (Asian languages) and application (creation of a specialized lexicon for data mining application in bio-medicine domain). The first stable set of Data Categories introduced in the Registry for the Morphosyntactic Profile has been presented at the Beijing meeting.

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 Contribution

CNR-ILC has continued the refinement and revision cycles of the lexical Markup Framework, the abstract meta-model for lexical representation. During the reporting period, three versions has been produced: revision-8, 9 and 13. On March 2006, the document has been submitted to ISO CD ballot. The ballot lasted three months and on June, 29 the SC-4 P-members voted. Comments coming from voting delegations will be taken into account and discussed during the next ISO Plenary meeting. During the reporting period, results of the ballot (March 2006) have become available: the national delegations expressed favourable opinion on the document and a lot of suggestions for improvement arose. Comments coming from voting delegations have been discussed during the ISO Plenary meeting held in Beijing (20-26 August 2006): there, the project coordinator unanimously with the editors and national delegates have decided to produce another document, integrating different feedback, and call for a further ballot in order to achieve a better acceptance rate. Different versions of LMF have been released: revision-10, based on comments coming from voting delegations, some intermediate versions (internally circulated after the meeting) and revision-13, the current version which has been submitted for the second ballot on November 2006. Results will be available on next March 2007.

As concerns advertising of Lexical Markup Framework to prospective users, a major scientific international conference was held in Genova (Italy), LREC06, where the lexical metamodel has been presented to a larger audience in many different occasions: satellite workshops, main conference and a special dedicated tutorial.

As concerns advertising of LMF to prospective users, a paper co-authored by INRIA, CNR-ILC and UPF, has been presented at the joint international conference held in Sydney (Australia) – July 2006, COLING-ACL06, where the lexical metamodel has been presented to a larger audience during the Workshop on Multilingual Lexical Resources and Interoperability. Furthermore, a paper by CNR-ILC has been presented at TSTT International Conference on Terminology Standardization and Technology Transfer, on August 26th 2006.

CNR-ILC is promoter of LMF in a number of International Projects where it acts as supervisor for the development of LMF-compliant lexical models. Specifically, in NEDO project (Japan grant) harmonized lexicons are being developed for Asian Languages and cross-language information retrieval purposes; in EU BootStrep project, a model for a lexicon in a Biomedical subdomain (gene regulation) is being developed for data mining applications. For this purpose, CNR-ILC has implemented both an RDF and XML version of the model and circulated some sample lexical entries, as a proof-of-concept of the model itself and in order to make partners able to instantiate LMF-conformant entries. Furthermore, a collaboration has been established by CNR-ILC and a representative of W3C from CNR-ISTC in Rome for the migration of LMF in OWL. In a meeting held in Rome, 29th November 2006 about the Semantic Dimension of Information Technology, LMF has been proposed for adoption in a new-born initiative *Senso Comune*, a linguistic knowledge base for Italian, built by people with new modalities of cooperation and collaboration through the web and openly available to humans for consultation and systems for elaboration.

In the reporting period, CNR-ILC started taking into consideration different criteria for the selection of lexical samples for *test suites* which will accompany the standard. The lexical entries will be selected with different types or variable levels of granularity in mind, in order to represent different lexicon structures and architectures. As soon as LMF will be voted and the model considered stable, CNR-ILC will provide LMF lexical test suites for Italian for two very different lexical approaches, represented by WORDNET-IT and SIMPLE-IT. It should also be mentioned that activities undergoing in other international projects for lexicon development would offer different sample lexical entries for different languages and application domains.

It should be noted that CNR-ILC is designated as the representative of UNI (the Italian National body) for all standardization activities of ISO/TC37/SC4.

IULA-UPF Contribution

IULA-UPF has been working on the development of a tool for creating LMF compliant databases. Two more researchers in IULA has been trained in LMF, and they are working now in the development and uploading with inner materials of the database.

INRIA Contribution

Gil Francopoulo, as co-editor of LMF, together with Monte George (from ANSI) produced the ISO documents: LMF-rev-8, rev-9 and rev-13. But in fact, these documents include numerous technical contributions from LIRICS partners, like CNR and UPF.

2.2 Extended report on activities related to WP2, including LIRICS meetings

2.2.1 Lirics Meetings:

- Participation of Monica Monachini to the tri-lateral meeting (CNR-ILC, Loria, DFKI) held in Paris (20.02.05 - 21.02.06)
- Participation of Monica Monachini and Claudia Soria to the LIRICS meeting for the preparation of the mid-term review of LIRICS that was held in Genova (in conjunction with LREC06), 23.05.2006. Monica Monachini presented the results of WP2 in the first year of the project.
- Participation of Nicoletta Calzolari, Monica Monachini and Claudia Soria to the General ISO Meeting held in Beijing as LIRICS partners and Italian National Delegates (20.08.05 - 25.08.06).
- Participation of Nicoletta Calzolari, Monica Monachini and Claudia Soria to the LIRICS internal meeting held in Beijing on 24-08-2006, where comments and suggestions from the mid-term reviewers were discussed and future activities planned.

2.2.2 Events related to LIRICS

- 30.03.2006: Monica Monachini (CNR-ILC) participated to the DIAM Commission in Rome at the premises of UNI, where LIRICS, other standardization activities and involvements in ISO TC37/SC4 has been presented to UNI members.
- Active participation of Nicoletta Calzolari, Monica Monachini and Claudia Soria (CNR-ILC) at the LREC06 Conference (Genova 22-28.05.2006) where a lot of papers about LMF has been presented. Monica Monachini was among the proposers and presenters of the LMF Tutorial which received great success and had a larger audience, with more than 30 participants.
- 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.
- 19/27.07.2006: Participation of Nicoletta Calzolari, Monica Monachini and Claudia Soria (CNR-ILC) at the COLING-ACL'06 Conference (Sydney), where papers have been presented about LMF or about LMF-conformant lexical models and lexicons. Some other papers and/or posters have been presented about lexical models and

lexicons got inspired by LMF, as a sign of penetration of lexical community by LMF. The LMF paper has been accepted for publication in a Special Issue on "Multilingual Language Resources and Interoperability".

- 29.11.2006: Nicoletta Calzolari and Monica Monachini, together with other CNR-ILC representatives, participated in a meeting held in Rome, "the Semantic Dimension of Information Technology", with strong relations with LMF.

3 Progress in WP3, morpho-syntactic and syntactic annotations

3.1 Progress and deviation

DFKI Contribution

In this perioding report, DFKI has built on results of Deliverable 3.1 and contributed to both the on-going standardisation activity on Morpho-Syntax (helping in preparing the next CD ballot of MAF) and on elaborating the Working Draft (WD) document for SynAF (SYNtactic Annotation Framework), which has been submitted and accepted as an ISO New Work Item in the last period of LIRICS. A first version of this document has been delivered at the beginning of 2006 and has been discussed at the ISO meeting in Jeju (Korea) between the 19.1 and the 21.1.2006. As a consequence of this meeting, it has been agreed that a new version of the document will be produced, taking into account the comments of the ISO experts present at the ISO meeting in Jeju, and that this version should be distributed to all ISO experts, who have produced a vote on the NWIP of SynAF. This was done in April and comments of the experts have been included in the most recent version of SynAF, which should be then discussed a last time in Paris in July before being presented at the next plenary ISO TC37/SC4 meeting, to take place in August in Beijing. The schedules of SynAF are in line with the plan of the Technical Annex.

At a meeting in Paris (6-7 July 2006) at the premises of AFNOR, both MAF and SynAF have been extensively discussed, whereas Thierry Declerck also represented the German DIN NAAT6 committee, which made an extensive list of comments on MAF (and also on LMF, comments to be discussed at the plenary ISO TC37/SC4). As a result of this meeting updated versions of MAF and SynAF will be prepared for the ISO TC37 plenary meeting to be held in Beijing (21-26 August 2006). The reports of MAF and SynAF at this meeting were the base for very interesting discussions and as a result it was decided to go for a new version of the MAF proposal (to be submitted for a CD in October 2006). This was also due to the fact that MAF/SynAF were not discussed in isolation, but together with the other TC37/SC4 projects LAF, LMF and the NWIP SemAF (new work item proposal). A need for harmonization of the terminology in used in the various documents was detected, and it was decided that the LMF proposal, although accepted at the CD ballot, would be updated and re-submitted for a CD ballot in November 2006. For MAF and SynAF, this means that they would be submitted for a CD ballot later as foreseen, since they build partly on LMF.

MAF is delayed wrt to the original ISO calendar. Partly for the reasons mentioned above. Also we have to take into account that the chief editor of MAF is not a member of the LIRICS consortium, and that we cannot influence directly the progress on this project. But LIRICS partners are contributing to the MAF document, and so we know the state of this project, and we can thus continue working on SynAF, which is partly building on MAF, and SynAF will be submitted to a CD ballot in March 2007. We expect SynAF to be available as a DIS proposal within the lifetime of LIRICS.

- DFKI also cooperated with WP2 and WP4 on data categories for the lexicon, the morpho-syntax, the syntax and the semantic. For establishing a consistent definition between WP2 (lexicon) and WP3 (morpho-syntax and syntax), a meeting took place in Paris in February 2006, and it was decided to introduce a profile for syntactic annotation. The cooperation between WP3 and WP4 (semantics) was also discussed at a more general meeting of ISO TC37/SC4 (TDG3 on the representation of semantic content), which was

held with the most relevant American colleagues end of April near Los Angeles. In fact DFKI spent a higher effort as foreseen in WP4, this due to the very successful development of the proposal for semantic annotation within ISO TC 37/SC4, which is ahead of time, compared to the LIRICS plans. Since a group at DFKI is working on issues related to the representation of temporal information in ontologies, DFKI is preparing material for the LIRICS and the New Work Item on semantic annotation (part1 of which is dedicated to the annotation of temporal expressions). Thierry Declerck presents this proposal at the plenary session of TC37 in Beijing and also at a joint LIRICS-TC37/SC4 and TimeML workshop in Boston (26-28 October 2006)

A slight delay on MAF and SynAF (3 months for the CD ballot) has to be reported, but we expect to be on time again for the DIS ballot.

UPF Contribution

IULA-UPF has been working in the survey of syntactic tags to be taken into account for the data category selection. A report on existing tree-banks tags and its relation to datacategories has been produced.

3.2 Extended Report on activities related to WP3, including LIRICS meeting

3.2.1 Lirics Meetings.

- Participation of Thierry Declerck (DFKI) at ISO TC37/SC4 meeting in Jeju, Korea (17-24 January 2006). Discussion of SynAF with Asian and North American representative of standardisation bodies. A lirics meeting takes place on the issue of semantic annotation.
- Participation of Thierry Declerck (DFKI) at the tri-lateral meeting of LIRICS (ILC, Loria and DFKI) 19-21 February 2006 in Paris
- Participation of Thierry Declerck (DFKI) at the joint ISO TC37/SC4, TDG3 and LIRICS meeting in Marina Del Rey: discussing with American partners the issue of semantic annotation, including the relation between syntactic and semantic annotation (19-24 April 2006)
- Participation of Thierry Declerck (DFKI) at the LIRICS meeting preparing the mid-term review of LIRICS, that took place on the 23rd of May in Genova (during the LREC conference). Thierry Declerck presents the results of WP at this review.
- Participation of Thierry Declerck (DFKI) at the joint LIRICS - ISO TC37/SC4 meeting in Paris (06-07 July 2006). Discussion of MAF, SynAF and preparing the SemAF proposal.
- Participation of Thierry Declerck (DFKI) at the ISO TC37 plenary meeting in Beijing (21-27 August), and at the satellite LIRICS events. Thierry Declerck is also representing the German DIN NAAT6 committee at this plenary event, and explains the comments made by DIN NAAT6 to the LMF and MAF proposals. Thierry Declerck presents SynAF and a proposal for a Time Ontology at the SemAF session.
- Participation of Thierry Declerck (DFKI) at the joint ISO TC37/SC4, TDG3, TimeML and LIRICS meeting in Boston (26-25 October 2006). Thierry Declerck presents an updated version of the DFKI proposal for a time ontology.
- Thierry Declerck presents actual state of the LIRICS project at the CLARING Workshop in Budapest (30-31-October 2006). The CLARIN workshop is about a European Infrastructure for Language Resources. Representatives of Institutions from the 25 European countries are present at this workshop.

3.2.2 Events related to LIRICS

- Thierry Declerck (DFKI) worked as a member of the DIN mirroring group NAAT 6
- Thierry Declerck and Mirjam Kessler present LIRICS work at the annual meeting of the DGFS society (21-24.2 2006 in Bielefeld)
- Participation of Thierry Declerck (DFKI) annual meeting of “DIN NAT AA 6 Sprachressourcen”, the German Mirror Committee of TC37/SC4 on 12.5.2006.in Berlin. LIRICS was discussed in detail and the DIN committee is very pleased to hear that the works in LIRICS on SynAF and also the semantic annotation is going well. The DIN committee also work on the CD ballots of LMF and MAF and produces a very exhaustive list of comments, which will be discussed in the ISO TC37/SC4 plenary meeting in Beijing (August 2006).
- Thierry Declerck (DFKI) participates very actively to the LREC conference held in Genova (22nd-28th May 2006).
- Thierry Declerck presents LIRICS and ISO work at a W3C workshop dedicated to multimedia semantic annotation (Amsterdam 10-11 July).
- Thierry Declerck presents updates of LIRICS and ISO work at a W3C workshop dedicated to multimedia semantic annotation (Athens 8-9 December just after the SAMT conference).

4 Progress in WP4, semantic content

4.1 Progress and deviation

Following on from the two bilateral meetings between UvT (Tilburg, Netherlands) and DFKI (Saarbrücken, Germany) on 16th November and 14th-16th December 2005 concerning how to isolate a comprehensive set of semantic roles, work in this area began to change from the more theoretical and methodological aspects of semantic annotation towards the development of a more concrete set of data categories for practical annotation. An early attempt at a preliminary list of data categories for temporal and dialogue act annotation was presented and discussed alongside the respective metamodel representations as early as the meeting held at LORIA (Nancy, France) on 10th-12th December 2005. By this point, as discussed in the document D4.1, it had been decided that there would be four areas of consideration for semantic annotation in the data category registry, namely: temporal information, reference, dialogue acts and semantic roles.

In the first half of 2006, data was gathered in earnest on the potential approaches towards a recommendation for a list of semantic roles. The hope was to be able to develop a set of data categories that were as well defined for all areas of consideration. With that view in mind, a literature review was made and some criteria for the consistent definition of semantic roles were formulated. Comparative studies of the various schemes already in existence were carried out. Large areas of overlap were found to exist between different systems of annotation, but little in the way of overall consensus. We wished to identify a rigorous, well-defined and, as far as possible, comprehensive set of data categories that have broad coverage of semantic roles in a variety of different and influential schemes. These data categories would cover what corresponds to the top-level, generic semantic roles, which might then be further refined and specified according to the type of event or action or state defined in the text. These preliminary data categories for semantic roles (along with an updated set of temporal and dialogue act data categories) were presented in document D4.2, and are more or less stable.

The data categories for reference annotation, although also presented in D4.2, are however still under development, and will be subject to considerable further revision pending contributions from other LIRICS participants.

For temporal information a preliminary set of data categories has also been devised. These will serve as input to the newly planned ISO work on the development of an international standard for the annotation of temporal information (see below).

For dialogue acts a set of data categories has been developed that has been discussed at several LIRICS and ISO meetings. In order to design a comprehensible and coherent set of data categories in this area, it was felt important to further develop a multidimensional approach to this aspect of semantic annotation. This approach has been worked out in some detail, has been investigated for its practical use by multiple annotators, and presented at several international conferences. The resulting set of data categories seems stable and extensible.

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

Other activities during this time frame include:

- ISO/LIRICS meeting in Jeju, South Korea on 19th-21st January 2006 (ISO TC 37/SC 4 meeting where LIRICS work was discussed). Discussion of metamodels for semantic annotation in the areas of LIRICS WP 4; methodological aspects of the design of semantic data categories; and planning of work.
- LIRICS/ISO meeting of experts in semantic annotation in Marina del Rey, Los Angeles, USA on 20th-22nd April 2006 that could be summarized as follows:
 - i) An intense discussion concerning all areas of semantic annotation (except reference annotation) – mainly planned as a cross-Atlantic workshop to secure greater US support for the development of interoperable semantic data categories.
 - ii) It was decided to submit a New Work Item Proposal forward to the ISO organization for the development of an international standard for the annotation of temporal information, taking TimeML as starting point, with a project team comprising James Pustejovsky, Bran Boguraev, Harry Bunt, Kiyong Lee and Nancy Ide.
 - iii) It was agreed to have another joint LIRICS/ISO meeting during the week of the IWCS-7 conference in Tilburg, Netherlands January 2007. Most of those experts present agreed to attend.
 - iv) There was extensive discussion of semantic roles with Martha Palmer (representing PropBank) and Michael Ellsworth (from FrameNet). There was broad agreement with the idea of the development of top-level data categories (roughly corresponding to the top-level of FrameNet); this has subsequently been implemented.
- LREC conference presentation of Task 4.1. An important opportunity to inform the research community of the effort to produce standards in (particularly) semantic annotation. A lot of interest was shown, especially in the methodological aspects.

5 Progress in WP5, LIRICS reference implementation platform

5.1 Progress and deviation

In the reporting period, WP5 participants have worked on tasks 5.1, 5.2, 5.3, and 5.4.

Task 5.1 worked on further development of APIs (Application Programming Interfaces) following the LIRICS standards for the ISO Data Category Registry (WP1), NLP Lexica (WP2), Morpho-syntactic annotations (WP3). USFD and MPI delivered the LIRICS reference architecture (D5.1.E). USFD delivered version 2 of the APIs for lexica (D5.1.B), morpho-syntactic annotation (D5.1.C) and syntactic annotation (D5.1.D).

Task 5.2 has continued and MPI delivered version 2 of the DCR Reference Implementation (D5.2.A). The LMF reference implementation (D5.2.B) is still being postponed because a workable XML interchange format has to be developed; MPI and other partners are working on this (which will also involve modifying the existing LMF standard) and it should be ready by M27. We are preparing to deliver the reference implementation for morpho-syntactic analysis (D5.2.C) in M25. UtiI and USFD are discussing the provision of a sample annotated corpus to demonstrate the semantic annotation standard instead of providing software for D5.2.D.

Task 5.3. continued and USFD delivered version 1 of the LIRICS service integration platform (D5.3) at M18. Version 2 of the LIRICS service integration platform (D5.3) will be delivered at M30.

Task 5.4 continued and MPI delivered the first version of the Data Category Usage platform (D5.4) at M18. The DCR usage platform has been integrated into LEXUS and has become a stable component of the standard LEXUS framework, and testing has been completed.

Deviations from the original planning of WP5: delay in the LMF implementation; slight delays in D5.2.C; proposal pending to amend D5.2.D.

5.2 Report on activities related to WP5, including LIRICS meetings

DFKI, MPI and USFD met briefly for technical discussions in M17 (at LREC in Genova in conjunction with the project review).

Technical DCR meeting at MPI, 31 August to 1 September.

Technical DCR meeting at UtiI, 11 to 12 December

6 Progress in WP6, dissemination and exploitation

6.1 Progress and deviation

According to the fact that a certain number of normative documents are stable now, the partners wrote and presented a great number of scientific communications, in the main conferences of the domain and all over the world.

6.2 Report on activities for WP6 dissemination and exploitation

- 15.02.2066 NEDO meeting in Pisa. Monica Monachini and Claudia Soria presented the LMF meta-model and RDF instantiations of the model has been provided.
- 30.03.2006, DIAM Commission in Rome, at the premises of UNI: Monica Monachini (CNR-ILC) briefly presents ILC-CNR activities connected to standardization, describing the LMF meta-model. The UNI Committee expressed the positive vote on

LMF and produced a set of comments that will be object of discussion to the next general ISO TC37 SC 4 meeting in Beijing.

- 13.04.2006 BootStrep kick-off meeting. Monica Monachini presented LMF as the state-of-the-art model for lexicon creation, stressing the importance for a project as BootStrep to work in compliance to this emerging ISO standard. RDF and XML instantiations of the model has been provided to partners.
- 22.05.2005: The LIRICS partners gave a tutorial on LMF during LREC conference. This tutorial was organized by G. Francopoulo, M. Monachini, L. Romary, S. Samon-Alt.
- 22-28.06.2006: LREC, Main conference: Lexical Markup Framework (LMF), G Francopoulo, M. George, N. Calzolari, M. Monachini, N. Bel, M. Pet, C. Soria.
- 22-28.06.2006: LREC, International workshop towards a research infrastructure for language resources: The relevance of standards for research infrastructures, G. Francopoulo, T. Declerck, M. Monachini, L. Romary.
- 22-28.06.2006: LREC, International workshop: acquiring and representing multilingual, specialized lexicons: LMF for multilingual, specialized lexicons, G. Francopoulo, M. George, N. Calzolari, M. Monachini, N. Bel, C. Soria.
- Thierry Declerck presented the paper by Thirsten Trippel, Ulrich Heid and Thierry Declerck "Standardisierungsarbeit für Sprachressourcen bei DIN (NAT AA 6))" at DGFS 2006.
- Mirjam Kessler presented a paper by Mirjam Kessler and Thierry Declerck "Data Categories as the link between Language archives and NLP tools" at DGFS 2006
- Thierry Delcerck presented the paper "SynAF: Towards a Standard for Syntactic Annotation" at LREC 2006.
- Thierry Declerck presented a paper by Declerck Thierry, Pérez Asunción Gómez, Vela Ovidiu, Gantner Zeno and Manzano-Macho David, "Multilingual Lexical Semantic Resources for Ontology Translation" at LREC 2006.
- Thierry Declerck presented a paper by Thierry Declerck, Mihaela Vela "Generic NLP Tools for Supporting Shallow Ontology Building" at LREC 2006.
- Thierry Declerck co-authored a paper, in which he stressed the role that can be played by LMF in linguistic descriptions in ontologies: Paul Buitelaar, Thierry Declerck, Anette Frank, Stefania Racioppa, Malte Kiesel, Michael Sintek, Ralf Engel, Massimo Romanelli, Daniel Sonntag, Berenike Loos, Vanessa Micelli, Robert Porzel, Philipp Cimiano [*LingInfo: Design and Applications of a Model for the Integration of Linguistic Information in Ontologies*](#). In: Proc. of OntoLex06, a Workshop at LREC, Genoa, Italy, May 2006.
- Khurshid Ahmad, Lee Gillam and David Cheng. (2006). Sentiments on a Grid: Analysis of Streaming News and Views. Proc. of 5th Intl. Conf. on Language Resources and Evaluation (LREC), 22-28 May 2006.
- Lee Gillam and Khurshid Ahmad. (2006). Financial data tombs and nurseries: A grid-based text and ontological analysis. Proc. of 1st Intl. Workshop on Grid Technology for Financial Modeling and Simulation (Grid in Finance 2006), 3-4 February 2006.
- Invited talk: Dr Gillam gave the talk "No place for Sentiments?". Presented over *Access Grid* as part of the National Centre for e-Social Science's Access Grid Seminar

Series. Audience: University of Surrey; Institute for Advanced Studies - Lancaster University; e-Science North West (ESNW) - Manchester University; Freeman Institute - Universities of Brighton and Sussex (8 June 2006).

- Invited talk: Dr Gillam gave the talk "Sentiment Analysis and Financial Grids" at JISC/ESRC funded workshop on Bridging quantitative and qualitative methods for social sciences using text mining techniques. Hosted by *National Centre for Text Mining / National Centre for e-Social Science*, Manchester Conference Centre. (28 April 2006)
- Nuria Bel: LIRICS, Linguistic Infrastructure for Interoperable Resources and Systems". Poster in the VII Congreso de Lingüística General, Barcelona, April 18-21, 2006
- 19/27.07.2006: Participation of Nicoletta Calzolari, Monica Monachini and Claudia Soria (CNR-ILC) at the COLING-ACL'06 Conference (Sydney). LMF has been discussed as a model for fostering the development of Multilingual Lexical Resources and Interoperability. The following paper has been presented: Francopoulo G., Bel N., George M., Calzolari N., Monachini M., Pet M. and Soria C. 2006. "Lexical markup framework (LMF) for NLP multilingual resources". In Proceedings of the COLING-ACL Workshop on Multilingual Lexical Resources and Interoperability, Sydney, Australia, It should be noted that some other papers and/or posters that have been presented about lexical models and lexicons are got inspired by LMF, as a sign of penetration of lexical community by LMF. The LMF paper has been accepted for publication in a Special Issue on "Multilingual Language Resources and Interoperability.
- 26.08.2006: TSTT-2006 International Conference on Terminology, Standardization and Technology Transfer. Nicoletta Calzolari, in collaboration with Monica Monachini and Claudia Soria (CNR-ILC) presented of CNR-ILC's international standardization activities across the years and LIRICS related results were extensively discussed.
- 22.09.2006: Valeria Quochi, CNR-ILC, presented LMF to BootStrep partners. The meta-model was proposed as the basis for starting defining the Bio-Lexicon, a specialized terminological lexicon in the domain of molecular biology. A specialized DTD drawn from LMF and some LMF-conformant XML entries has been produced and studied between partners.
- 28.11.2006: Monica Monachini took part in a Conference Call together with the NEDO partners, during which part of the LMF syntactic-semantic extensions were discussed.
- 29.11.2006: Nicoletta Calzolari and Monica Monachini, together with other CNR-ILC representatives, participated in a meeting held in Rome, "the Semantic Dimension of Information Technology". LMF has been proposed for adoption in a new-born initiative *Senso Comune*, as a model for a new conception of a computational lexicon for Italian, built by people openly available to humans and software systems.
- 25-26.08.2006 Lee Gillam, Debbie Garside, Chris Cox. "Information volumes and linguistic diversity: meeting the challenges for content management". 3rd International Conference on Terminology, Standardization and Technology Transfer, Beijing PRC.

7 Progress in WP7, management

7.1 Progress and deviation

Financial coordination: In the reporting period, Virginie Tessier (INRIA) established and maintained the financial records and payments.

Administration and project coordination:

- INRIA with the help of the other LIRICS partners organized the LIRICS mid-term review in Genoa (23rd May 2006).
- INRIA maintained the LIRICS web site and the associated forum where a great number of technical messages were exchanged.

7.2 Extended report on activities related to WP7, including LIRICS meetings

- Gil Francopoulo and Monte George had a two days technical meeting in Washington (10, 11 January 2006) in order to stabilize LMF core model.
- Gil Francopoulo organized the Thematic domain group-2 on data categories for morpho-syntax in Jeju (Korea) and attended the meetings on syntactic annotation and word segmentation (19, 20, 21 January 2006).
- Gil Francopoulo attended WordNet conference in Jeju following ISO meetings.
- In June 2006, Gil Francopoulo has been nominated as French head of delegation for ISO-TC37/SC3 (Terminology and other resources) and ISO-TC37/SC4 (Language resource management).