LIRICS

Deliverable D7.1

Risk Management Plan

Project reference number	e-Content-22236-LIRICS
Project acronym	LIRICS
Project full title	Linguistic Infrastructure for Interoperable Resource and Systems
Project contact point	Laurent Romary, INRIA-Loria 615, rue du jardin botanique BP101. 54602 Villers lès Nancy (France) romary@loria.fr
Project web site	http://lirics.loria.fr
EC project officer	Erwin Valentini
Document title	Risk Management Plan
Deliverable ID	D7.1
Document type	Report
Dissemination level	Public
Contractual date of delivery	M2
Actual date of delivery	28 February 2005
Status & version	Draft
Work package, task & deliverable responsible	WP7, Task7.4, INRIA-Loria
Author(s) & affiliation(s)	Gil Francopoulo
Additional contributor(s)	Harry Bunt, Monica Monachini, Julien Nioche
Keywords	Risk, Consensus

Document evolution

version	date	version	date
1.0	28th February 2005		
1.1	29th August 2005		
1.2	30th August 2005		
1.3	15th August 2005		
1.4	19th September 2005		

1 Introduction

1.1 Goal

The goal of the current document is twofold:

1) to provide an assessment of the self-perceived risk factors;

2) to provide an overview of newly emerging technologies that are relevant to the Lirics project.

This self-assessment has been carried out in order to:

- Make the choices that have been taken in the project so far more explicit;

- Evaluate the choices: which choices are safe choices and which one are more risky;

- Evaluate whether the current technologies are most optimal for reaching the project goals, and determine possible directions for future work, using potentially more suitable technologies.

1.2 Contents

The report is built up as follows: chapter-2 gives an overview of the risk assessment. Chapter-3 gives the result of the technology watch.

2 Risk assessment

This chapter provides a self-assessment of risks that are common to all work-packages and risks per technical work-package.

2.1 Risks common to all work-packages

The risks identified at global project level are as follows:

- Delays in the acceptance of ISO documents regarding to CD ballot and delays in the work-plan;

- Industrial partners does not accept the standard proposed by Lirics;

- Take up and usage of standards are limited.

2.2 Risks for WP2 NLP lexica

The risk is to target one of the two extremes:

- One extreme is to define a too generic and abstract model without any connection with existing and foreseen lexicons;

- The other extreme is to focus on a too specific family of lexicons without taking into account the other lexicons.

The strategy chosen based on data categories permits to reach a vast scope of lexicons but these data categories must be recorded effectively and be of good quality.

The other potential risk is that the proposal developed in this work-package might not receive the acceptance at the ISO milestones with consequent delays. In order to avoid this situation, keeping close contacts with the ISO LMF working group is essential.

Other risk can be that the external research community does not take-up the proposal issued by LIRICS. In order to be sure that the proposal be adopted by consumers en mass, an industry driven strategy has been defined for the whole project and prospective end-users have been identified between key e-Content players. The LIRICS Industrial Advisory Board has been involved very early in the first phases of development of the lexical model, in order for us to receive early feedback from them and ensure that the specifications match their needs.

Other risk in an European initiative for standardization, is to have mainly an eurocentric worldview. In order to avoid as much as possible this situation, inter-relationships, continuous contacts and exchanges between LIRICS partners and ISO experts from America and Asia, in particular, are crucial. This is also useful in order to widen the frontier of our standards and give credibility to them in a truly international perspective.

2.3   Risks for WP3 Morpho-syntactic and syntactic annotations

The main risk lies in the complexity of the meta-model, since a lot of morpho-syntactic and syntactic features are to be considered.

For syntactic annotation, less industrial applications are including high-level syntactic annotations for the time being. What may prove difficult is to abstract over all language families of the world.

2.4   Risks for WP4 Semantic content

Standardization efforts concerning semantic annotation and representation have so far been limited to certain restricted domains of semantic information, such as time, space, and reference. Given that the work in this area is in a relatively early stage, there is a potential risk that the proposals developed in this work package might not be widely accepted. To deal with this risk, the LIRICS WP4 work is performed in close cooperation with the ISO TC 37/SC 4 Task Domain Group 3 (Semantic Content Representation), which was established in 2004, and by involving the ACL-SIGSEM Working Group on the Representation of Multimodal Semantic Information, which is led by researchers who participate actively in the LIRICS project As a first step, a joint LIRICS-sponsored two-day workshop of the ACL-SIGSEM Working Group and ISO TC 37/SC 4/TDG 3 was held in January 2005 in Tilburg, in conjunction with the 6th International Workshop on Computational Semantics.

2.5   Risks for WP5 Lirics reference implementation platform

This work-package depends on the successful definition of the relevant Lirics standards in WP2, 3 and 4. Therefore the main risks come from failure or delay in those WPs.

3   Technology watch regarding related projects and technologies

3.1   WP2 NLP lexica

Lirics partners must follow what's going on at W3C. In one direction, Lirics may take ideas from W3C and on the reverse direction, Lirics may influence W3C work. In this perspective, Nicoletta Calzolari and Gil Francopoulo have been nominated as ISO liaison officers between ISO/TC37/SC4/WG4 and the related projects within W3C.

Within ISO/TC37/SC2 a standard concerning presentation and representation of entries in dictionaries is being finalized. A Lirics member (Gil Francopoulo) follows this work and assists to all meetings of this group.

OLIF-2 must be taken into account. A study of the mapping between OLIF-2 and LMF has been made within Lirics. The project leaders are preparing OLIF-3. Moreover a set of extended examples of major existing lexicons have been pammed to the emerging LMF in order to study the way they fit into the LMF structure. The model must support data merging and interchange between two lexicons and as a consequence, permit to measure the difference between two differently conceived lexicons.

Liaison and continuous co-operation should be ensured with people belonging to Asian Federation of National Language Processing (AFNLP) in order to monitor, in this association, initiatives, events and technological advancements.

Similar activities towards the definition of standard practices for language resources undertaken at international level by some organizations such as ELRA should be undertaken and synergies established.

LIRICS members should interact with colleagues and experts in speech community who are undertaking parallel initiatives and try to define a common framework which accomodates both needs and requirements, in the areas where convergence is deemed possible.

Another area where LIRICS partners should concentrate efforts and standards appears to be "Multiword expressions". A study about the capability of the lexical framework is foreseen within the lifetime of the project and will proceed step-by-step with the ISO task force devoted to this issue.

Attention needs to be paid to latest developments in descriptive languages for multimodality, to be sure that the lexicon model will be able to accommodate different modalities, visual and spatio-temporal information (as it could be the case, for instance, for sign language lexicons).

3.2   WP3 morpho-syntactic and syntactic annotations

The ongoing project Technolangue/Easy in France in the domain of evaluation of syntactic annotations must be taken into account.

3.3   WP4 semantic content

Relevant for the work in LIRICS WP 4 is the development of annotation schemes indicating predicate-argument structures ('semantic roles') in the PropBank project, an effort to add a layer of semantic annotations to material in the Penn Treebank. In order to take this effort into account, a representative from this project participated upon invitation to the LIRICS-sponsored workshop on semantic annotation in Tilburg in January 2005. A workshop specifically intended to discuss the relevance of this effort and other recent North-American efforts is planned for late 2005.

3.4   WP5 Lirics reference implementation platform

The work done by WP2, 3 and 4 must be studied carefully during the life time of the project. It can occur that some statements considered as valid at the beginning of the project may change before the end of the project. A good communication with other WPs must be established and maintained.

The LIRICS reference implementation will be based on Web Services, thus using the Web Service Description Language (WSDL), see http://www.w3.org/TR/wsdl. In the context of the WP5, the LIRICS partners will follow the evolutions of the next generation of Web services specifications, including SOAP 1.2, WSDL 2.0 and any other emerging trends in Sercice

Oriented Architectures (SOA). In particular, a special attention will be paid to Grid computing, which can define new standards in close future.

4   Conclusion

This report provides a risk assessment for all work-packages of the Lirics project, as well as the results of a watch of the technologies applied in the project.

As far as possible, state-of-the-art technologies are applied in the project. So, given the goal of the project, identified risks can be considered acceptable.