



ECP-2007-LANG-617001

FLaReNet

Action Plan

Deliverable number/name *D8.1 – Action Plan*

Dissemination level *Public*

Delivery date *29 May 2009*

Status *Final*

Version *0.3*

Authors

Nicoletta Calzolari, Claudia Soria, Paola Baroni, Tommaso Caselli, Sara Goggi, Monica Monachini, Valeria Quochi, Antonio Toral

Nuria Bel, Gerhard Budin, Khalid Choukri, Joseph Mariani, Jan Odijk, Stelios Piperidis



eContentplus

This project is funded under the eContentplus programme¹,
a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.

¹ OJ L 79, 24.3.2005, p. 1.



Table of Contents

| | |
|--|-----------|
| TABLE OF CONTENTS | 2 |
| 1 INTRODUCTION | 3 |
| 2 FLARENET GENERAL OBJECTIVES AND ACTIONS..... | 3 |
| 2.1 CREATE AND MOBILISE A UNIFIED AND COMMITTED COMMUNITY OF LANGUAGE RESOURCES AND TECHNOLOGIES PLAYERS..... | 4 |
| 2.2 ELICITATE “HOT TOPICS” | 5 |
| 2.3 DEFINE PRIORITIES AND FORMULATE STRATEGIES..... | 5 |
| 2.4 ISSUE RECOMMENDATIONS | 6 |
| 3 FLARENET SPECIFIC OBJECTIVES AND ACTIONS..... | 6 |
| 3.1 A UNIFYING THEME: AN OPEN RESOURCE INFRASTRUCTURE | 7 |
| 3.2 SUPPORT KNOWLEDGE AND AVAILABILITY OF LANGUAGE RESOURCES AND TOOLS..... | 8 |
| 3.3 CREATE CONSENSUS AROUND INTEROPERABILITY ISSUES AND PROMOTION OF STANDARDISATION..... | 10 |
| 3.4 PROMOTE LANGUAGE RESOURCE CREATION AND ENHANCEMENT | 12 |
| 3.5 LONG TERM ACTIVITIES INNOVATIVE METHODS AND MODELS FOR LR CREATION, MANAGEMENT, VALIDATION, AND ACCESS | 14 |
| 3.6 EVALUATION AND VALIDATION | 14 |
| 3.7 ENSURE INTERNATIONAL COOPERATION AND SYNERGIES | 14 |
| 4 APPENDIX- A LIST OF ACTIONS AND RECOMMENDATIONS FROM FLARENET..... | 17 |



1 Introduction

In this document we describe the Action Plan for the FLaReNet Thematic Network.

In FLaReNet, actions to be undertaken are related both to general objectives (overall, overarching goals of the project) and to specific objectives (i.e., those the Network identifies periodically on the basis of its work, the interaction with the Human Language Technology (HLT) community, and with other projects).

The document is structured as follows. The first chapter illustrates the general objectives of the Network, and the corresponding broad typologies of actions that are directed to achieve them. The second chapter narrows the focus on a smaller set of specific objectives and related actions that are to be seen as instantiations of the general types of actions described before. This also shows the strategy adopted by FLaReNet for the identification of actions to be undertaken by the project: specific objectives are selected on the basis of the particular themes, issues, and trends arising from the FLaReNet community at large. Specific objectives emerge from the interaction of people participating to FLaReNet meetings and events and are further pushed by the FLaReNet SC when perceived as worthy of consideration and further inspection.

In addition to keeping the main issues in focus, FLaReNet has the peculiar objective of being able to adapt to changing needs and requirements as they emerge from the community. It must be kept in mind that FLaReNet does not act independently from its surrounding environment. It pursues its objectives in close cooperation with major international bodies and groups, to try to formulate recommendations that come from worldwide consensus. At the same time, FLaReNet must be able to “sense” trends, needs and orientations well in advance, and act as a sentinel of changing conditions. If the Network wants to be successful in achieving this, then it will have to readjust and re-orient its actions from time to time. The Action Plan is therefore conceived as a high-level living document that will be revisited as the project progresses.

The Appendix represents the most “living” part of the Action Plan: it contains, divided by topic, a list of Actions and Recommendations mostly issued by FLaReNet first meetings. Actions and Recommendations are intended not only as to be carried out by FLaReNet, but also as indications of work for existing and upcoming EU projects, in an attempt to coordinate and map current and future initiatives onto a broad common goal.

2 FLaReNet General Objectives and Actions

The general objectives of FLaReNet are those deriving from its statutory goals, i.e.:

- the consolidation of a community in the field of Language Resources and Technologies;
- the definition of priorities and strategies to be undertaken, to be presented under the form of a roadmap;
- the development of recommendations for the community in general, and the policy-makers in particular.



2.1 Create and mobilise a unified and committed community of Language Resources and Technologies players

FLaReNet has the challenging task of creating a network of people around the notion of Language Resources and Technologies. To this end, FLaReNet is bringing together leading experts of research institutions, academies, companies, funding agencies, as well as users (either as individuals and as members of professional organisations) with the specific purpose of creating consensus around short, medium and long-term strategic objectives. It is of foremost importance that the FLaReNet Network be composed of the as widest as possible representation of experiences, practices, research lines, industrial and political strategies; this in order to derive an overall picture of the field of Language Resources and Technologies that is not limited to the European scene, but can also be globally inspired.

Actions targeting the objective of creation and consolidation of a community are largely encompassing dissemination activities (Deliverable 1.4). All actions are conceived so as to have an effect, in the end, on the community that FLaReNet wishes to consolidate and strengthen. Here, we sum up the most relevant ones.

In order to constantly increase the community of people involved in FLaReNet, as well as to ensure their commitment to the objectives of the Network, an active permanent recruiting campaign is always open. People wishing to join the Network can do so by filling an appropriate web form available on the FLaReNet web site. The FLaReNet Network is open to participation by public and private, research and industrial organizations. Invitation to join, either personal or by means of mailing lists are used in order to enlarge the community as much as possible.

The Network is currently composed of around 200 individuals belonging to academia, research institutes, industries and government, and it is steadily enlarging through new subscriptions. Such a community needs to grow not only in number, but also with reference to the type of disciplines involved, from the core ones (Natural Language Processing, computational linguistics, Language Engineering) to “neighbouring” ones, such as cognitive science, semantic web, etc. Participants are expected and encouraged to express their views individually as experts but also their organizations views and concerns.

A check point is foreseen every six months to assess the number, type (academic vs. industrial, institutional, etc.) and geographical distribution of people composing the FLaReNet community. Meetings, either thematic or general ones², are the primary means for attracting new members and to reinforce participation of existing ones, but participation is expected and encouraged also by means of online discussions, forum threads, and collaborative documents. A team of online discussion moderators from the core team helps to instigate discussion and channel it into useful directions.

Other general ways for sensitizing and attracting people, as well as for making former members aware of the Network activities, is a massive use of advertising material, publishing of the Newsletter, and participation in conferences and major events related to Language Resources and Technologies.

² For a description of the different types of meetings foreseen in FLaReNet, see D1.4, *Dissemination Plan*.



Apart from actions for enlarging the FLaReNet community, those aimed at consolidating it are especially important. Participants to the community need to feel they belong to a group of people that is actually shaping the field of Language Resources and Technologies by delineating its direction for the next future. The User Forum, the creation of Thematic Group and “think-tanks” of experts and the launch of closed meetings are the privileged ways for creating close and connected groups of people.

2.2 Elicit “hot topics”

FLaReNet has the challenging goal to act as a “sensor” of current and future trends in Language Resources and Technologies. In order to do this, it must be able to make most pressing issues emerge from its community of players. A number of actions globally converge toward this goal:

- thematic meetings;
- encouragement to propose discussion themes (e.g. through our wiki site);
- requests for topic proposals for Thematic meetings / provoking issues;
- link with major (new) projects & initiatives.

2.3 Define priorities and formulate strategies

Language technologies and language resources are the necessary ingredients for the development of applications that will help bridging language barriers in the global and unified information space, in a variety of means (the Web as well as other communication devices) and for a variety of channels/media (spoken and written language alike but also other associated modalities e.g. gesture). It is of utmost importance, however, to identify a set of priority themes as well as short, medium, and long-term strategic objectives in order to avoid scattered or conflicting efforts. The major players in the field of Language Resources and Technologies need to consensually work together and indicate a clear direction and priorities for the next years, under the form of a roadmap for Language Resources and Technologies. This is the kind of results at which meetings are especially targeted. Actions foreseen to this end are centred on the activity of thematic, general and liaison meetings (see Deliverable 1.4 for further details).

2.4 Issue recommendations

The FLaReNet Network must propose priority themes under the form of consensual recommendations and a plan of action for EC Member States, other European-wide decision makers, companies, as well as non-EU and International organizations. This in order to sensitize those responsible for implementing the financial and political frameworks that are necessary to sustain the actions to be implemented on a long term. Actions toward this goal are:

- publication of FLaReNet recommendations and their distribution to international associations, consortia, bodies that have outreach power;
- white papers;
- relations with industrial and users communities;



- relations with funding bodies and government institutions;
- relations with EC and its funded projects.

3 FLaReNet Specific Objectives and Actions

These objectives are to be seen as the FLaReNet “foci”, i.e. those issues that are identified by the Network as the major dimensions around which the community should converge and for which different types of actions should be put in motion. These foci not only represent the issues being addressed and analysed by FLaReNet in order to distil priorities and recommendations, but they also act as the glue that will make all FLaReNet activities and messages cohesive and help structure and scale the community activities.

A first set of specific objectives are those that emerged from the FLaReNet so-called “Vienna Forum”. This was a major event that gathered around 100 active players in the field with the specific purpose of setting up a task force to brainstorm and make emerge the technological, market and policy challenges to be faced in a multilingual digital Europe.

It was organised as a series of working sessions where leading experts were invited to present their vision on hot topics in the field of Language Resources and Language Technologies. A new formula was experimented, whereby the FLaReNet Steering Committee prepared for each session a background document highlighting a set of relevant issues and questions to be addressed by the speakers. Such documents were provided to the speakers and participants well in advance to ensure high level and well thought of contribution.

Over a two-day programme, the participants to the Forum had the opportunity to start elaborating an accurate picture of the current landscape of the Language Resources and Technologies field and to propose emerging directions of intervention, not necessarily constrained by the event or FLaReNet framework.

Some messages recurred repeatedly across the various sessions, as a sign both of a great convergence around these ideas and also of their relevance and importance to the field. The Forum validated ideas that have been “in the air” for several years and, in some cases, fostered and/or developed by specific groups, as having entered the main stream of thought and practice within the language technology community.

3.1 A unifying theme: an Open Resource Infrastructure

Among the many themes that emerged from FLaReNet Vienna Forum, infrastructure building seems to be the main message and the most urgent issue ahead of us. With inductive methods dominating the current paradigms in Language Engineering, language resource building, annotation, cataloguing, accessibility, availability and clearance from IPR is what the research community is calling for. The message coming out from the FLaReNet Vienna Forum is only one among the multiple concurring signs



now indicating that time is ripe for establishing an Open³ Resource Infrastructure (ORI), which allows networking of language technology professionals and their clients, as well as easy sharing of data, corpora, language resources and tools that are made interoperable and work seamlessly together. An ORI is something that the Language Resources community has been pushing since some time, and that is now increasingly recognized as a necessary step for building on each other achievements and avoid scattered or conflicting efforts. As a response to the community needs, FP7 Call 4 foresees a Network of Excellence addressing such an infrastructure.

FLaReNet does not have the necessary resources to put such an infrastructure into place, but it can pave the way to later actors with a number of preparatory initiatives.

We foresee that FLaReNet can play a valuable role in:

- sensitizing the community, by spreading awareness of the need and justification for such an infrastructure;
- catalyzing a number of activities around this major theme, that can be seen as “the” endeavour of the LR community in the next future;
- mobilizing the community by setting up expert groups that will work in synergy on a number of different, yet concurring issues that are crucial for the development of a language resource infrastructure, such as a preliminary assessment of the temporal, technical, organisational, and legal constraints involved in the implementation of an open language infrastructure;
- boosting cooperation between the existing infrastructures and initiatives;
- in this way, prepare the field for a future implementation of an open resource infrastructure by early elicitation of user needs and requirements on a range of issues, from technical to organisational.

The theme of laying down the basis for an open resource infrastructure can thus be seen as providing a unifying framework around which other sub-goals are organised. The sub-goals, in their turn, correspond to the main issues at stake on which FLaReNet can provide advice and recommendations, also in synergy with the work to be carried out under FLaReNet work-packages.

Main sub-themes are:

- fostering the knowledge dissemination and availability of language resources and tools that will eventually enter the open language infrastructure;
- the creation of consensus around interoperability issues and the promotion of standardisation activities;
- promoting the creation of new language resources and the enhancement of existing ones;
- giving a first assessment of the various constraints involved in the implementation of an open resource infrastructure, such as, for instance:
 - **Temporal:** what is the time frame under which a first implementation can be reasonably expected (considering various steps and building blocks)?

³ Openness has to be understood here as an intrinsic multidimensional characteristic of the ORI taking into consideration dimensions related to technical basis, legal aspects, licensing possibilities and schemas, sharing/distribution and pricing policies, etc.



- **Technical:** which should be the main principles underlying its architecture, assuming some evolution over time to account for new developments (web2.0 ...)?
- **Organisational:** how the involved centres and resources should be regulated? Which are the best organisational and governance models that can ensure efficiency and sustainability?
- **Legal:** what are the legal and IPR issues that could be involved by the realization of an open language infrastructure?

3.2 Support knowledge and availability of language resources and tools

During its first months of activity, FLaReNet has already gained the role of pole of attraction of the Language Resources and Technologies community. Thanks to that, FLaReNet has the potential to act as an early incubator for language resources and tools that will eventually become part of an open resource infrastructure. A capillary activity is needed in order to map existing resources, inform about them, enhance their accessibility. FLaReNet has the power to do so. A first set of actions, in coordination with other relevant organisations, associations and initiatives, is detailed below. These actions are broadly targeted at

- 1) locating and describing currently used Language Resources and Tools;
- 2) providing a comprehensive, community-built repository of information about Language Resources and Tools, in order to start creating consensus about topics and terminology, and
- 3) ultimately enhancing accessibility to Language Resources and Tools.

Actions

Short Term

- FLaReNet will launch the first steps in designing a Wikipedia-like repository of information dedicated to Language Resources and Technologies, and eventually some internal resources will be allocated to its implementation. This is intended as an informative resource, to which all FLaReNet members and supporters will be asked to actively contribute. This Wiki could either be linked to Wikipedia, or be an independent resource.
- FLaReNet will act in synergy with ELRA and LREC2010 to draw the “Map of Language Resources, Technologies and Evaluation”, in order to survey new, existing and most used language resources and tools. A dedicated interface will be set up in conjunction with LREC submission procedure for eliciting information about the language resources and tools that will be presented or represent the background of the scientific work submitted to the conference. The LREC2010 Map will be linked to the ELRA Universal Catalogue. Such work may be furthermore conducted on the publication of previous LREC conferences (and may be others, such as ACL, Interspeech, etc.) if some (internal) funds could be secured of this purpose.



Mid Term

- Organise meetings with representatives of the European Parliament, Euronews, DGT, and other owners of large volumes of data to investigate modalities for their sharing

3.3 Create consensus around interoperability issues and promotion of standardisation

In the long term, interoperability will be the cornerstone of a global network of language processing capabilities. Interoperability of language resources in the future will be ensured if we now establish the necessary framework and a corresponding infrastructure (i.e. standards and technologies) that will be later made operational. This can only be achieved through a coordinated, community-wide effort that will ensure both comprehensive coverage and widespread acceptance. FLaReNet can act as coordinating factor to ensure wide permeation of the notion of interoperability in the community, so as to pave the way to wide acceptance of its use.

Standardisation is at the core of interoperability. Standardisation issues currently show substantial convergence of opinion and practice, which needs now to be supported to become operational. For instance, standards now need tools that support them – this will promote and ensure their adoption. Standards for interchange formats (for I/O among linguistic processing tools and systems) were agreed to be the appropriate means to pursue interoperability. Repositories for tools and language data, but also for information on data (documentation, manuals, metadata, etc.) should be established that are universally and easily accessible by everyone. Not only are data formats to be standardised, but also metadata. The area of metadata has already achieved a broad consensus and substantial work has been performed, so that it appears a reasonable, well defined area to start from. Moreover, establishment of well-designed and widely agreed metadata is a necessary requirement for any implementation of an open language infrastructure. Another intervention can be foreseen in the area of criteria for annotating and producing language resources. The availability of common annotation guidelines and specifications is perceived as a viable solution to current problems in the production of language resources, such as efficiency, quality and interoperability. To this end, the creation of a repository with data formats, annotations, etc. – where to find the most frequently used and preferred schemes – is proposed as a major help to achieve and promote standardisation.

FLaReNet, in line with its mission, will ensure the broadest possible community engagement in the development and agreement on strategies, priorities and best approaches by means of sessions, open meetings, and special workshops at major conferences in the field, together with active maintenance of and involvement in open web forums and Wikis. Several task-forces will be established to address the above issues. The privileged means of interaction of the task forces will be online collaboration through the FLaReNet web site, both in the section dedicated to the Steering Committee and in those of the various working groups.



Actions

Short Term

- FLaReNet will create a working group in the area of high-level specification of metadata, to work in synergy with other major initiatives, such as, for instance, the working groups dedicated to this topic in CLARIN and ELRA
- FLaReNet will launch a shared repository with data formats, annotations, guidelines, etc. and eliciting user requirements towards this goal
- FLaReNet will activate collaboration with OLAC and LDC
- FLaReNet will support/finance workshops and tutorials on how to use standards, especially in those areas where there is more consensus

Short/Mid Term

- FLaReNet will identify the major efforts on standards development and interoperable system design together with existing and under-development technologies, and surveying success cases in interoperability. “Best practices” in standards will be identified from the point of view of their usability, usefulness, viability, outreach etc. Conversely, FLaReNet will analyse what is missing to make standards operational, in terms of community awareness, usability, usefulness, supporting infrastructure needed, facilitating tools, etc. A survey will be launched to elicit the community’s view on this issue. Those areas for which interoperability would have the broadest impact in advancing research and development and significant applications will be identified.
- FLaReNet endorse/support initiatives for the development of tools fostering the use of standards, for instance web services providing format conversion facilities from proprietary representation formats to standardised ones
- FLaReNet, will encourage/promote the adoption and use of standards in new projects.

Long Term

- FLaReNet will agree upon a common roadmap for the adoption of encoding standards for language resources and tool (in conjunction with CLARIN, DARIAH and ELRA).

3.4 Promote language resource creation and enhancement

There is an urgent need to reduce the divide between languages that are wealthy in language resources and those that are not, a prerequisite that is necessary to guarantee an adequate development of NLP technologies for under-represented languages. Support should be given so that basic language resources are developed for all languages. This is also in line with the current trend to “de-globalize” human language resources and focus on local languages/cultures. A re-alignment of language resources as for languages covered is needed on the long term. New EU accessing countries should be especially supported in order to attain a complete coverage of the range of language resources needed to support the development of most important language and speech applications. Existing resources, on the other hand, need to be maintained and improved so as to achieve better quality. High quality of language technology is essential and this



in turn requires high quality and large quantity of language resources. To this end, more investments and effort on automatic means should be devoted to research and deployment of solution on the automation of the production of the large quantity of resources demanded, and of enough quality to get acceptable results in industrial environments.

At the same time, interest and research should also be encouraged to devise better methodologies, frameworks and initiatives for the validation and evaluation of the resources produced, especially automatically.

FLaReNet either directly takes or encourages actions toward the goal of promoting language resource creation, validation and evaluation, either for new languages and with innovative methods.

Actions

Short Term

- Draw up a list of application-driven needs for LR production: which applications require which types of LRs.
- Draw up a list of applications/technologies of “greatest exposure” that incorporate multilingual aspects.
- Liaise with representatives from Wikipedia, AnaWiki Project social networks, so as to investigate how social-based methods can be exploited for the production of Language Resources.
- Reflection should be devoted to elaborate an adequate business model supporting the vision of an open resource infrastructure, so that a preliminary assessment of the various requirements is given.
- The methodologies developed by the PAROLE/SIMPLE partners for the enrichment and tuning of legacy resources will be promoted as best practice models for use of LRs.
- Improve awareness about legal issues concerning intellectual property.
- Identify needs of less-resourced languages and assess current status of development of LR&Ts in the European, Euro-Mediterranean and regional framework.
- Provide a definition of BLARK. Provide an estimate of the overall costs of a BLARK (Basic LAnguage Resource Kit) for the 23 EU languages (FLaReNet WP2, Jan 2010) and Find reliable methods for assessing the depth and breadth of the gaps in language resource coverage.
- Investigate different resource production models, such as the formation of consortia with players with similar interests, wiki-like models, social networks, etc. Take into account their “dynamic” aspect.

Mid Term

- Evaluate/assess the difficulty to address the evaluation of all technologies for all languages – produce a sort of BLARK for evaluation (FLaReNet WP5, from Jan 2010)

Long Term

- *Cooperative Services.* Another strategy proposed and partially implemented is to set up *cooperative services by multiple local players* to counter one big global player. Since each local player can often offer better quality than the global player



for the particular local language or domain, the cooperating local players together can compete against the global player and offer customers the best available quality

- Investigate different solutions for addressing the problem of task- vs. application-oriented evaluation, such as:
 - A general evaluation framework, including both kinds of evaluation, such as the ISLE Framework for Evaluation in Machine Translation (FEMTI) approach.
 - An integrated evaluation platform.
- Discuss and investigate frameworks for remote evaluation distributed over the Internet, which allows interchanging components, comparing various approaches, and examining the influence of single components on the whole system.

3.5 Ensure international cooperation and synergies

Given the issues at stake, cooperation cannot be limited to a European landscape. The field of Language Resources and Technologies now needs a global approach, similar problems are encountered and solved everywhere, and similar solutions must be found if we want the field to have a coherent direction instead that many dispersed ones. To this end, FLaReNet will do the following:

- A brainstorming exercise has been set up in order to elicit main needs, justification, characteristics of an ORI, starting from concrete scenarios. This will result in a document/report/memorandum.
- FLaReNet will exploit the coordination with its sister project SILT (US funded) in order to agree on common strategies for eliciting feedback from the community.
- Permanent connections will be sought and established with major national and international players (such as AFNLP, ELRA, LDC, COCOSDA, TEI, Oriental-COCOSDA, ALTA, ISO, ETSI, etc.) in order to both elicit feedback and disseminate results and recommendations.
- FLaReNet has established an International Advisory Board grouping major players outside the EU. The International Advisory Board will prepare and commit to a Memorandum of Understanding that will refer to the main foci described above.

Actions

Short Term

- FLaReNet will take the lead in assembling relevant people, institutions, and organisations around the world into a collaborative network to which the institutions and individuals involved are committed (and really, have funding for) whose goal is to collaboratively work toward interoperability.
- Facilitate more interaction and co-operation between different communities active in the broad field of Language Resources and Technologies: Machine Translation technology providers, Computer Assisted Translation technology providers, translation and localisation service providers, large institutional language services, Language Resource producers in different domains and



different languages, research organisations interested in these topics and processes, and publishing houses.

- Establish an International Advisory Board that will act as an International Forum.
- Prepare a Memorandum of Understanding with the main issues discussed and ask members of FLaReNet to sign it when joining the Network.
- Promote the need of a communication vector for open source resources and tools.
- FLaReNet can formally promote a new worldwide language infrastructure for easy access to Language Resource and Technologies, in a web-based resource and technology grid.

Mid/Long Term

- Produce a White paper summarising ideas for directors of programs of funding agencies, and organise a Forum of directors of funding agencies.