



D4.1_LT-Observatory_specifications

Acronym: LT_OBSERVATORY

COORDINATION AND SUPPORT ACTION
INFORMATION AND COMMUNICATION TECHNOLOGIES

D4.1 LT-Observatory Specifications

GRANT AGREEMENT	644583
DELIVERABLE NUMBER	D4.1
DELIVERABLE TITLE	LT-Observatory Specifications
DUE DATE OF DELIVERABLE	31/03/2015
ACTUAL SUBMISSION DATE	08/05/2015
START DATE OF THE PROJECT	01/01/2015
DURATION	24 months
ORGANIZATION NAME RESPONSIBLE FOR THIS DELIVERABLE	LT-Innovate

DISSEMINATION LEVEL		
PU	Public	<input checked="" type="checkbox"/>
PP	Restricted to other programme participants (including the Commission Services)	<input type="checkbox"/>
RE	Restricted to a group specified by the consortium (including the Commission Services)	<input type="checkbox"/>
CO	Confidential, only for members of the consortium (including the Commission Services)	<input type="checkbox"/>

TYPE		
R	Document, report	<input checked="" type="checkbox"/>
DEM	Demonstrator, pilot, prototype	<input type="checkbox"/>
DEC	Websites, patent filings, prototype	<input type="checkbox"/>
OTHER		<input type="checkbox"/>

TABLE OF CONTENTS

ABBREVIATIONS & GLOSSARY	2
1. INTRODUCTION	3
2. OBJECTIVES	3
3. AUDIENCE	3
4. TECHNICAL SPECIFICATIONS	4
4.1 DOMAIN NAME.....	4
4.2 HOSTING	4
4.3 CONTENT MANAGEMENT SYSTEM	4
4.4 ARCHITECTURE	5
4.5 ROLES AND ACCESS PERMISSIONS.....	5
5. REQUIRED FEATURES	5
6. PROVISIONAL SITE STRUCTURE	5
6.1 HOMEPAGE.....	5
6.2 LIST OF MENUS	6
6.2.1 TOP LEFT MENU	6
6.2.2 TOP RIGHT MENU	6
6.2.3 LT-OBSERVE MENU	6
6.2.4 FOOTER MENU	6
6.3 MAIN SERVICES.....	6
6.3.1 DIRECTORY OF STAKEHOLDERS.....	7
6.3.2 PUBLIC POLICY OBSERVATORY.....	7
6.3.3 LANGUAGE RESOURCES OBSERVATORY	8
6.3.4 MARKET OBSERVATORY.....	9
6.3.5 LIBRARY.....	10
7. SOCIAL NETWORKS	10
7.1 BLOG	10
7.2 TWITTER.....	10
7.3 LINKEDIN.....	10
7.4 SCOOP.IT.....	10
8. SUB-CONTRACTING OF TECHNICAL FEATURES	11
9. PLATFORM DELIVERY AND ACCEPTANCE	11
NOTE	11

ABBREVIATIONS & GLOSSARY

DoA	Description of Action (Annex 1 / part A of the LT_Observatory project contract)
DSM	Digital Single Market
LRR(s)	Language resources repository/ies
LR(s)	Language resource(s)
LT	Language technology/ies
LTO	LT-Observe, the brand name selected for the service and platform established by the LT_Observatory project
MDSM	Multilingual Digital Single Market
MT	Machine translation



1. INTRODUCTION

The European Digital Single Market, one of the main goals of Europe 2020, is still fragmented due to language barriers. European society is multilingual, the diversity of its cultural heritage is an opportunity, but hampers trans-border eCommerce, social communication and exchange of (cultural) content. Languages without sufficient technological support will become marginalised. These barriers must be overcome, in particular by language technology (LT) solutions such as machine translation (MT). **One of the most fundamental building blocks to achieve a Multilingual Digital Single Market (MDSM) will be to join, benchmark the quality and facilitate the access to language resources (LRs).**

With this in mind, the LT Observe platform established by the LT_Observatory project aims at:

- creating a user-friendly one-stop-shop point of reference to language resources repositories (LRRs);
- checking these resources with regard to their quality/usability and assessing/benchmarking them;
- identifying and documenting European, national and regional funding sources (including Regional and Structural Funds) that may be used to fund additional work necessary in the field of LRs and MT ;
- providing an LT News Observatory, building on the news service developed by the LT COMPASS project (and operated by LT-Innovate);
- serving as a facilitator of cooperation between stakeholders, notably by the means of providing the online presence of targeted dialogue events and technical workshops;
- providing access to various tools (documents, links, etc.) geared at improving the quality of MT and LRs and supporting a high coverage of languages within the MDSM.

2. OBJECTIVES

The objective of work package 4 of the project (WP4) is to define and deploy the “LT-Observe” sub-platform (in the sense that it will be hosted on LT-Innovate’s existing platform), the online single access point for the information gathered throughout the LT_Observatory project including qualified information on language resources, best practice methodologies and European/national/regional funding sources.

3. AUDIENCE

Beneficiaries of this service / platform will include:

- LT / LR / MT providers
- LT / LR / MT users
- LRRs keen to promote the content of their repositories to potential users
- Researchers with an interest in the field of LT / LR / MT

D4.1_LT-OBSERVATORY_SPECIFICATIONS

LT_OBSERVATORY

- Policy makers with an interest in the field of LT / LR / MT
- Other stakeholders with an interest in the field of LT / LR / MT

4. TECHNICAL SPECIFICATIONS

4.1 DOMAIN NAME

The domain name of the LT-Observe sub-platform is: www.lt-observe.eu

The service name and domain name have been selected in view of sustainability: although the LT_Observatory project will end after 24 months, LT-Observe will remain on-line as a tool for the LT communities.

4.2 HOSTING

The LT-Observe sub-platform will be hosted on the dedicated server (Kimsufi) of LT-Innovate. It will be added to the already existing LT platform (www.lt-innovate.eu) as a sub-area (with a distinct, albeit compatible branding and look-and-feel). Furthermore, it will be linked to the CLARIN ERIC VLO – Virtual Language Observatory and other Language Resources Repositories already in existence.

4.3 CONTENT MANAGEMENT SYSTEM

The Drupal Open Source content management system (<http://drupal.org>) is the technology selected to develop the LT-Innovate website for the following reasons:

- Drupal is an open source platform which is widely used and supported with high quality, documented code.
- Drupal is coded in the open source scripting language PHP and supports as primary data sources the open source database formats MySQL and Postgresql.
- Drupal is standards based for both established and emerging standards. Specific target standards include XHTML and CSS.
- Drupal is modular and extensible. Based on a slim, powerful core, the application can be readily extended through a wide variety of custom plug-in modules.
- Low resource demands. To ensure excellent performance, Drupal puts a premium on low-profile coding (for example, minimizing database queries).
- Drupal has minimal, widely-available server-side software requirements.
- Specifically, Drupal should be fully operational on a server with Apache web server, PHP, and either MySQL or Postgresql.
- Drupal contains a full content management system, allowing content on the website to be easily updated.
- Ease of use. Drupal aims for a high standard of usability for developers, administrators, and users.

4.4 ARCHITECTURE

The LT-Observe sub-platform will follow the rules of the LT-Innovate website and will be designed for standard client-server architecture, running on a web server with the following characteristics:

- Linux operating system
- PHP for scripting
- MySQL and other versions as a database
- Cascading Style Sheets (CSS) for customization
- The drupal Views OAI-PMH module (https://www.drupal.org/project/views_oai_pmh)

4.5 ROLES AND ACCESS PERMISSIONS

The Roles and Access Permissions of the sub-platform will be based on the existing Roles and Access Permissions defined for the LT-Innovate platform. However, one additional role will be created: trusted user. This role will be made available to the persons untrusted with the initial benchmarking / assessment of the selected LRs. Hence, the following roles will be implemented for the sub-platform:

- Anonymous: any unlogged user of the platform
- Registrant: normal registrant, i.e. logged user of the platform
- Trusted user: user who has access to the LR benchmarking / assessment tools
- Editor: user who has the right to contribute content to selected parts of the platform
- Administrator: platform administrator

5. REQUIRED FEATURES

The following features are foreseen in the DoA of the LT_Observatory project:

- A Language Resources Catalogue: a one-stop-shop information service providing easy access to parallel corpora and LRs, searchable according to languages, countries, sectors/content, technology and users;
- An online Directory of Stakeholders (LR providers, solution providers, researchers, public authorities);
- A showcase of best practices in terms of MT/LRs (European and national initiatives);
- A catalogue of European and national funding sources for MT/LRs;
- A MT/LRs EcoGuide in modules customised for different stakeholders;
- A news curation service and social networks to support community building.

6. PROVISIONAL SITE STRUCTURE

6.1 HOMEPAGE

The Homepage of the LT-Observe sub-platform will include:

- A Header composed of the LT-Observe logo



D4.1_LT-OBSERVATORY_SPECIFICATIONS

LT_OBSERVATORY

- A rotating slider accommodating one or several banners (these can be used notably, for announcing project events and main achievements)
- 3 blocks highlighting the sub-platforms main “services”:
 - Public Policy Observatory
 - Language Resources Observatory
 - Market Observatory
- The prescribed logos of the European Commission and the following disclaimer “LT-Observe is supported by the LT_Observatory project funded by the Horizon 2020 Programme of the European Commission.”

6.2 LIST OF MENUS

6.2.1 TOP LEFT MENU

According to the LT-Innovate architecture, the top left menu will be displayed on the top left corner of all the pages of the LT-Observe sub-platform. This menu aims to quickly switch to the different sub-platforms of LT-Innovate. LT-Observe will be one of them.

6.2.2 TOP RIGHT MENU

According to the LT-Innovate architecture, the top right menu will be displayed on the top right corner of all the pages of the LT-Observe sub-platform. This menu offers shortcuts to all the social media accounts maintained by LT-Innovate (Blogger, Twitter, LinkedIn, YouTube, Scoop.it, rss) and the contact data (phone & e-mail).

6.2.3 LT-OBSERVE MENU

The LT-Observe menu includes the main pages of the LT-Observe sub-platform:

- Home
- Directory of Stakeholders
- Public policy
- Language resources
- Market
- Library
- Contact: the contact page will allow visitors filling in a webform (first name, last name, e-mail, and message) and sending it to contact@lt-innovate.eu.

6.2.4 FOOTER MENU

The Footer menu is a copy of the LT-Observe menu with all the items expanded for search engine optimisation (SEO) purposes.

6.3 MAIN SERVICES



D4.1_LT-OBSERVATORY_SPECIFICATIONS

LT_OBSERVATORY

6.3.1 DIRECTORY OF STAKEHOLDERS

The Directory of Stakeholders will be generated, using LT-Innovate’s main database as an engine. Relevant contacts will be added to this database with a “multilingual” tag allowing them to be displayed separately on the LT-Observe sub-platform. The Directory’s look and feel will be similar to LT-Innovate’s main [LT Directory](#).

6.3.2 PUBLIC POLICY OBSERVATORY

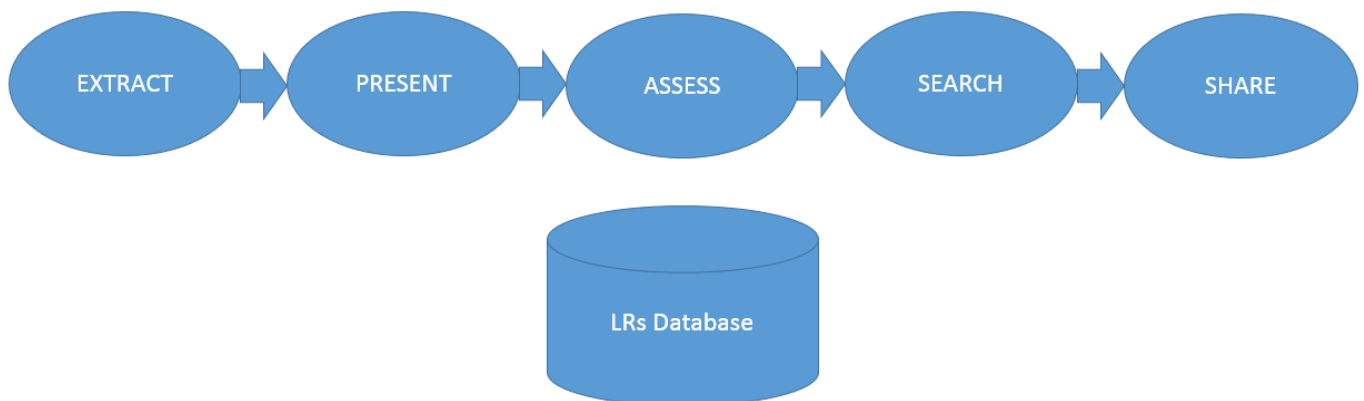
Public Policy is a basic page providing access to four areas or “services”:



- EU Language Policies will coalesce information (mostly in the form of links to relevant third party resources) about the EU’s policies and initiatives in the field of language technologies (including MT), language resources and multilingualism.
- National Language Policies will coalesce information (mostly in the form of links to relevant third party resources) about national policies and initiatives in the field of language technologies (including MT), language resources and multilingualism.
- EU Funding Opportunities will coalesce information (mostly in the form of links to relevant third party resources) about EU funding sources that may be used to fund additional work necessary in the field of LRs and MT.
- National / Regional Funding Opportunities will coalesce information (mostly in the form of links to relevant third party resources) about national funding sources (including Regional and Structural Funds) that may be used to fund additional work necessary in the field of LRs and MT.

6.3.3 LANGUAGE RESOURCES OBSERVATORY

The Language Resources Observatory will be implemented according to the following basic flow chart:



EXTRACT

The most useful language resources (selected on the basis of the methodology defined in WP1) will be extracted manually from all relevant LRRs. They will be appropriately tagged with keywords (to be defined in WP1) and referenced in the LRs database of the LT-Observe sub-platform.

PRESENT

The selected LRs will be presented in the Language Resources Observatory section of the sub-platform in the form of several lists (by language, by topic etc.). These lists will be generated automatically by the system on the basis of the keywords/tags allocated to each set of LRs during the “extract” phase. This will include a machine-readable access point of the Language Resources Catalogue, supporting the OAI-PMH protocol.

ASSESS

The online assessment/benchmarking system will comprise two features:

- A simple star rating system
- A comments system

Trusted users of the sub-platform will be designated who will have the right to allocate stars and make comments. Each allocation of stars (from 1 to 5 stars) will require an explanation through the comments function. The comments can be either selected from a standard list of comments (to be defined by WP1) or by contributing free text.

SEARCH

Initially, a simple search system will be implemented which will allow search by keywords / tags, language and number of stars. The search system will be discussed at the occasion of the Dialogue Days and Technical Workshops and may be expanded on the basis of the input received.



D4.1_LT-OBSERVATORY_SPECIFICATIONS

LT_OBSERVATORY

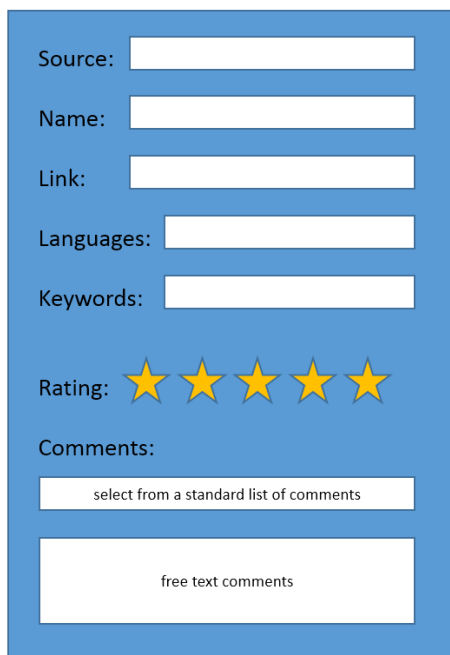
N.B. The subcontracting budget foreseen for technical support in view of platform development will be mostly reserved to refine the “assess” and “search” functions of the sub-platform.

SHARE

The metadata collected in the LR Observatory will be made available again for harvesting by other LLRs, such as the Virtual Language Observatory (www.clarin.eu/vlo). This will enhance the total impact of the collected and curated data. In exchange, LRRs will be required to include a link to the LR Observatory, which should also guide additional users towards the LTO’s benchmarking interface.

DATABASE

The database will be composed of individual LR records, structured as follows:



Source:

Name:

Link:

Languages:

Keywords:

Rating: ★★★★★

Comments:

6.3.4 MARKET OBSERVATORY

This section will consist of two main parts:



D4.1_LT-OBSERVATORY_SPECIFICATIONS

LT_OBSERVATORY

NEWS CHANNELS will be composed of the existing LangTechNews channel of LT-Innovate and a new LangPolNews channel which will reflect news related to the EU and national language policies.

MARKET STUDIES will provide access to LT-Innovate's existing market monitoring / research services. These resources will be supplemented by additional information sources (documents, links, etc.) specifically relevant to LRs / MT.

6.3.5 LIBRARY

The Library section will initially provide access to:

- the News Channels: LangTechNews and LangPolNews
- a restricted document repository available for trusted users only

7. SOCIAL NETWORKS

Social networks channels of LT-Innovate will be used for all communication purposes because these benefit already well-targeted and engaged LT communities.

7.1 BLOG

The layout of the blogger has to be updated to welcome articles from the partners. Articles will be displayed as a magazine using the dynamic views of Blogger.

A new metatag "LT-Observe" will be created to categorize efficiently all the articles of the partners. A new page will be created to list all the articles of LT-Observe.

7.2 TWITTER

LT-Innovate's Twitter accounts (@LTInnovate and @LangTechNews) will be the relay of the LT-Observe communication on Twitter. All tweets for LT-Observe will include the following hashtag: #LTO

7.3 LINKEDIN

LT-Innovate's existing "Partnering in LT" group can be used to exchange and discuss topics relevant to the LT_Observatory project.

7.4 SCOOP.IT

LT-Innovate's existing channel LangTechNews will be used for market related news. Whenever relevant to LRs, news items will be tagged "language resources".



D4.1_LT-OBSERVATORY_SPECIFICATIONS

LT_OBSERVATORY

An additional channel LangPolNews will be created to provide coverage of EU and national language policies. Project partners will be encouraged to scoop interesting information related to the project's scope.

8. SUB-CONTRACTING OF TECHNICAL FEATURES

While the LT-Observe platform will be launched and maintained by LT-Innovate's staff, the DoA foresees the sub-contracting of certain technical features such as the realisation of a technical benchmarking / assessment system of LRs and a powerful and versatile search function.

9. PLATFORM DELIVERY AND ACCEPTANCE

The beta version of the LT-Observe sub-platform will take three months to be built. As the present specifications will be validated by the end of March, we plan to have the beta-version online by the end of June 2015. A fully operational sub-platform will be available by the end of September 2015. It is expected that fine-tuning will be necessary throughout the project. The sub-platform will be declared as delivered and accepted if and when the green light is received from the project partners, the European Commission and the project reviewers in the various capacities.

NOTE

This document has been drafted at a very early stage in the project, taking into consideration all features mentioned in the DoA, but at a time when essential contributions from other work packages, which will shape the outputs of the project are not yet available and, in many cases, have not yet been initiated. Furthermore, essential feedback to be obtained from the service / platform user communities through Dialogue Days and Technical Workshops is also missing. It is therefore to be expected that the platform specifications will have to be adapted, revised and fine-tuned as the project progresses in terms of fine-tuning and launching the platform (first beta version in October 2015 and final platform in June 2016). The WP leader will adapt the platform throughout this period within the limits of the implementation efforts and budget foreseen.