



Multilingual Ontologies for Networked Knowledge

D7.3.3 Annual Public Management Report year 1 (Publishable Summary of Annual Report)

Project ref. no	FP7-ICT-4-248458
Project acronym	Monnet
Start date of project (dur.)	01 March 2010 (36 Months)
Document due Date	28 February 2011 (M12)
Responsible for deliverable	Project Administration
Reply to	[sean.oriain, maria.smyth, paul.buitelaar] @deri.org
Document status	Final

Project reference no.	FP7-ICT-4-248458
Project working name	Monnet
Project full name	Multilingual Ontologies for Networked Knowledge
Document name	D7.3.3 Annual Public Management Report year 1
Security (distribution level)	PU
Contractual delivery date	28 February 2011 (M12)
Access	Public
Deliverable number	D7.3.3
Deliverable name	Annual Public Management Report
Type	Report
Version	Final
WP / Task responsible	WP7 (NUIG) - Sean O'Riain, Maria Smyth, Paul Buitelaar
Contributors	Project Administration, partners
EC Project Officer	Michel Brochard
Distribution List	Consortium Partners
Review list	Consortium Partners
Approved by	Project Administration, Consortium Supervisory Board, Research Supervisory Board
Document Location	https://dev.deri.ie/confluence/display/monnet/Deliverables

The Monnet project is concerned mainly with ontology localisation, i.e., the translation of the lexico-terminological level of ontologies (often referred to as the ‘ontology labels’). The project outcomes, as currently understood by the project members, can be described as a set of software components as follows, all of which can be used in combination as well as stand-alone:

- Ontology Lexicalization
- Ontology Localization
- Cross-lingual Ontology-based Information Extraction
- Cross-lingual Knowledge Access & Presentation

The following diagram illustrates the envisioned connections between these components, with their inputs and outputs:

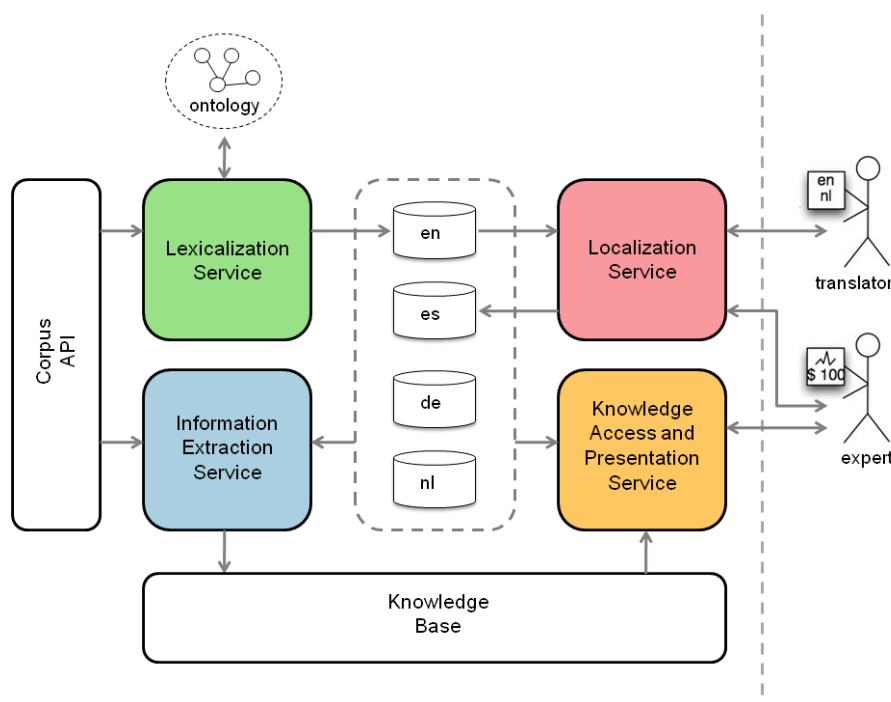


Figure 1: Monnet overview

The core objective of Monnet is the provision of advanced services for the translation of the lexico-terminological level of ontologies, which will be instantiated by the ‘localization service’. However, as ontologies often have only a very limited representation of lexico-terminological information, a first step will be to analyze a given ontology and enrich it with appropriate information on i) the terminological structure of ontology labels, ii) linguistic information on terminology items, and iii) analysis of implicit semantics where needed. Together we refer to these analysis and enrichment steps as “ontology lexicalisation”, which will be instantiated by the ‘lexicalisation service’ that takes as input an ontology and outputs an ‘ontology-lexicon’ for at least one default language (depending on the language that was used in defining ontology labels). A ‘corpus service’ will enable access to external domain corpus evidence for modelling and analyzing language use in the ontology labels. The ontology-lexicon will be represented on the basis of the so-called ‘lemon’ format¹, a lexicon model for ontologies that has been defined during this reporting period by the Monnet project for the appropriate integration of lexical/linguistic and terminological information in ontologies. The different lexicons will therefore be handled by use of the ‘lemon API’ as shown.

¹ <http://lexinfo.net/>

After the generation of an appropriate ontology-lexicon by the ‘lexicalisation service’, the ‘localization service’ can now operate on this for the translation process, i.e., note that Monnet thus implements ontology localisation by translation of the ontology-lexicon. A hybrid machine translation (MT) approach will be explored in this that will build on state of the art statistical MT as provided by the MaTrEx system² of Dublin City University, with which Monnet initiated an informal collaboration during this reporting period, and knowledge-based approaches using domain knowledge provided by ontologies and associated domain terminologies. The ontology localisation service will be developed and tested in the context of two real-life use cases on the translation of i) financial services taxonomies in XBRL (eXtensible Business Reporting Language) developed by Monnet partner XBRL Europe, and ii) public services ontologies developed by Monnet partner Be Informed for the Dutch government.

To provide a further proof of concept in a real-life use case, the Monnet project will implement also a cross-lingual ontology-based information extraction component that will exploit the XBRL-based ontology-lexicons for various languages for fact extraction from multilingual financial reports. The ‘information extraction service’ will instantiate this functionality by use of the lemon API (for accessing lemon-based ontology lexicons) and the corpus service (for accessing linguistically/semantically annotated domain corpora) to generate a knowledge base (‘KB’) of financial facts.

Finally, the ‘knowledge access & presentation service’ will instantiate a component for the multilingual querying and visualisation of this KB, which will provide a further proof of concept in the real-life use case of cross-lingual business intelligence. A particular focus here has been on supporting the internationalization of web applications by using linguistic knowledge available in the form of lemon models. A rule language called LAIF has been developed in this reporting period that eases the task of localizing a web application to different languages. This component will be able to provide cross-lingual, ontology-based access to financial information by the use of XBRL semantics in combination with multilingual ontology-lexicons.

During the current reporting period the following achievements were met in implementing the objectives as outlined above (corresponding to the Monnet milestone I ‘Foundations’ and the first part of Monnet milestone II: ‘Focus and Development’):

- WP1: initial Use Case Definition; initial version of Use Case Data
- WP2: Ontology-Lexicon Model completed; initial version of Web Access Component; initial version of Ontology Access and Linguistic Pre-processing Component
- WP3: Definition of Interface between Information Extraction and Ontology-Lexicon Model
- WP4: Methodology for Multilingual User Interface Localization
- WP5: 1st Iteration of Prototype Component Integration; 1st Performance Evaluation and Usability Study
- WP6: Dissemination & Exploitation Plan; Monnet Project Website
- WP7: Project Work Plan & Quality Plan; Project Self-Assessment Plan

To inform interested parties on the objectives of the Monnet project and to keep them informed of new developments, such as the lemon model, a web site has been set up during this reporting period at the URL <http://www.monnet-project.eu/> (corresponding to deliverable **D6.1.1 Monnet Project Web Site**). The web site has been designed to present information on the Monnet project objectives and aims, the research partners, work packages, and use cases. Additionally, web site visitors will be kept up to date on scientific publications on Monnet research, industrial and scientific events organized by Monnet partners, invited talks by Monnet researchers on relevant topics, Monnet dissemination in the media, and on available Monnet downloads and demos. The following screen shots illustrate the look and feel of the website:

² <http://www.openmatrex.org/>



Figure 2: Monnet web site – homepage

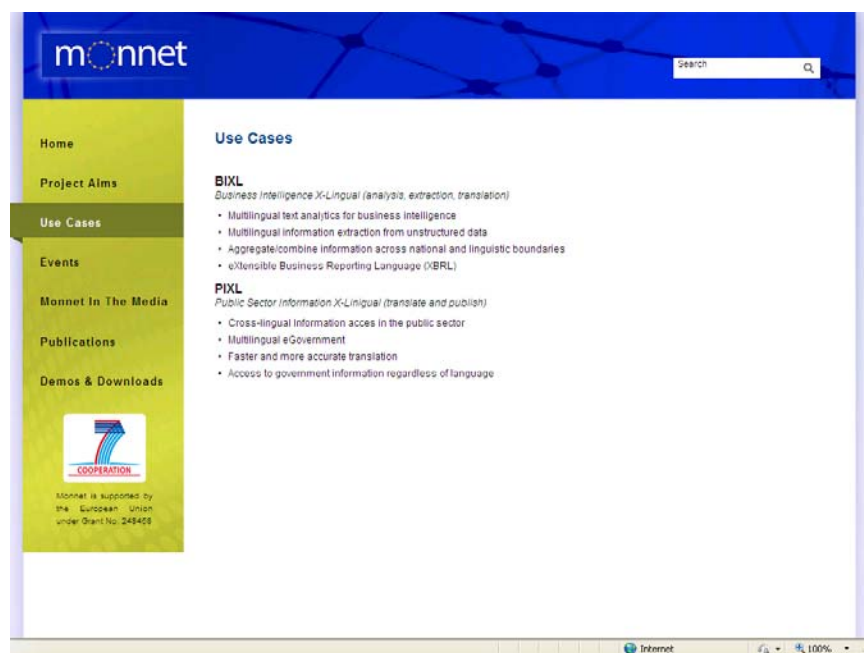


Figure 3: Monnet web site – use case descriptions