MOLTO

Published on *Multilingual Online Translation* (http://www.molto-project.eu)

D10.4 MOLTO Dissemination and Exploitation Report

Contract No.:	ICT-FP7-ICT-247914 and 288317
Project full title:	MOLTO - Enlarged EU, Multilingual Online Translation
Deliverable:	10.4
Security (distribution level):	PU
Contractual date of delivery:	M39
Actual date of delivery:	30 May 2013
Type:	Report
Status & version:	Final
Author(s):	O. Caprotti, B. Popov, J. van Aart
Task responsible:	UGOT
Other contributors:	

Abstract

The final dissemination and explotation report discusses how the project MOLTO has informed the public of the results. The industrial partners of the Consortium, Ontotext and Be Informed are the main contributors to the exploitation plan for the technologies developed by MOLTO. Exploitation of MOLTO aims to pursue sustainability for the tools and technologies and to further their uptake.

1. Introduction

In the MOLTO initial plan for dissemination, we proposed to carry out the task in the following way:

> Dissemination on conferences, symposiums and workshops will be in the areas of language technology and translation, semantic technologies, and information retrieval and will include papers, posters, exhibition booths and sponsorships (by Ontotext at web and semantic technology conferences like ISWC, WWW, SemTech), and academic/professional events such as the Information Retrieval Facility Symposium. We will also organize a set of MOLTO workshops for the expert audience, featuring invited speakers and potential users from academy and industry.

Here we report on what has been done to popularize the work done in MOLTO and make the language

translation community aware of the project.

Additionally, this deliverable contains a plan of further exploitation of the project's results. In the longer run, as outlined by the Strategic Research Agenda for Multilingual Europe in 2020 by the META Technology Council, Language Technology is expected to *enable forms of knowledge evolution, knowledge transmission, and knowledge exploitation that speed up scientific, social, and cultural development*. Any exploitation of MOLTO results will have to take into account the themes of this research agenda. It is already clear that the trends have started. For instance, Theme 1, the translation cloud is the fitting trend for the MOLTO web services living in the cloud. Some of the MOLTO application grammars in the cloud do indeed provide "services for instantaneous reliable spoken and written translation among all European and major non-European languages".

2. Dissemination activities

During the lifetime of the project we have pursued many ways of informing the relevant stakeholders about the progress of the research and development of MOLTO tools. The user community for MOLTO technologies comprises academicians, working the areas of computational linguistics and semantic web, but also members of industry offering services such as translations of web pages and of online content, from e-Government and business logics to cultural heritage, patents in pharmacology and creators of resources for e-learning of mathematics.

Here below the ways in which this broad user community has been targeted.

2.1 International Conferences and Meetings

Because of the Open Access Clause, we had to make sure that the copyright policy for the proceedings of chosen conferences and meetings would allow distribution of the publication also on the partners' Open Access Servers. This is the list of Open Access servers that are also distributing the MOLTO publications:

- Göteborgs universitets publikationer http://gup.ub.gu.se/ and Chalmers Publication Library http://publications.lib.chalmers.se/
- UPCommons, open access to the knowledge of Universitat Politècnica de Catalunya: http://upcommons.upc.edu/
- University of Helsinki: Research database Tuhat https://tuhat.halvi.helsinki.fi/portal/en/ and Digital Archive HELDA https://helda.helsinki.fi/
- University of Zürich: http://www.oai.uzh.ch/

Here below is the list obtained from the web pages by fetching publications registered by the authors as **Conference Papers**.

[1253] Asiya: An Open Toolkit for Automatic Machine Translation (Meta-)Evaluation., Giménez, Jesús, and Màrquez Lluís , Fifth Machine Translation Marathon, Volume 94, Le Mans, (2010) Google Scholar BibTex Tagged RTF XML   Download: PBML-2010-Gimenez.pdf (170.97 KB)

Comparing human perceptions of post-editing effort with post-editing operations, Koponen, Maarit , Proceedings of the Seventh Workshop on Statistical Machine Translation, June, Montréal, Canada, p.181–190, (2012)

Computational evidence that Hindi and Urdu share a grammar but not the lexicon, Prasad, K. V. S., and Virk Shafqat , 3rd Workshop on South and Southeast Asian Natural Language Processing (SANLP), COLING 2012, (2012) Download: wssanlp-camera-ready.pdf (144.1 KB)

Controlled Language for Everyday Use: the MOLTO Phrasebook, Ranta, Aarne, Enache Ramona, and Détrez Grégoire , Controlled Natural Languages Workshop (CNL 2010), Marettimo, Italy, (2011) Download: everyday.pdf (647.5 KB)

[997] Creating Linguistic Resources with GF, Ranta, Aarne , LREC 2010, 05/2010, Valletta, Malta, (2010)

Deep evaluation of hybrid architectures: simple metrics correlated with human judgments, Labaka, Gorka, Díaz De Ilarraza Arantza, España-Bonet Cristina, Sarasola Kepa, and Màrquez Lluís , International Workshop on Using Linguistic Information for Hybrid Machine Translation, 11/2011, Barcelona, Spain, p.50-57, (2011) Download: LIHMT2011_cameraReady.pdf (183.94 KB)

Developing an interlingual translation lexicon using WordNets and Grammatical Framework, Virk, Shafqat, and Prasad K. V. S. , NoDaLiDa 2013, (2013) Download: uwngf.pdf (138.07 KB)

[1254] Document-level Automatic MT Evaluation based on Discourse Representations, Comelles, Elisabet, Giménez Jesús, Màrquez Lluís, Castellón Irene, and Arranz Victoria , 5th Workshop on Statistical Machine Translation (IWMT, 2010), (2010) Download: wmt10-cgmac.pdf (67.6 KB)

[1283] A Framework for Improved Access to Museum Databases in the Semantic Web, Dannélls, Dana, Damova Mariana, Enache Ramona, and Chechev Milen , RECENT ADVANCES IN NATURAL LANGUAGE PROCESSING, 09/2011, Hissar, Bulgaria, (2011) Download: ranlpWS-final.pdf (421.76 KB)

Full Machine Translation for Factoid Question Answering, España-Bonet, Cristina, and Comas Pere R. , EACL Workshop on Exploiting Synergies between Information Retrieval and Machine Translation, Avignon, Frnace, (In Press) Download: qasmt.pdf (201 KB)

[2012] General Architecture of a Controlled Natural Language Based Multilingual Semantic Wiki, Kaljurand, Kaarel , Third Workshop on Controlled Natural Language (CNL 2012), 09/2012, Volume 7427, Berlin / Heidelberg, Germany, p.110--120, (2012)

On generating coherent multilingual descriptions of museum objects from SemanticWeb ontologies, Dannélls, Dana , International Conference on Natural Language Generation (INLG), 2012, (2012) Download: inlg2012-final-proceedings.pdf (130.89 KB)

The GF Mathematical Grammar Library, Caprotti, Olga, and Saludes Jordi , Conference on Intelligent Computer Mathematics /OpenMath Workshop, 07/2012, (2012) Download: gf-mgl.pdf (219.79 KB); GF.matlib.xhtml (10.01 KB)

The GF Mathematics Library, Saludes, Jordi, and Xambó Sebastian , Proceedings First Workshop on CTP Components for Educational Software (THedu'11), 02/2012, Volume Electronic Proceedings in Theoretical Computer Science , Number 79, Wrocław, Poland, p.102–110, (2011)

How Much do Grammars Leak?, Angelov, Krasimir , COLING 2012, (Submitted) Download: gf-penn.pdf (229.75 KB)

[1285] Hybrid Machine Translation Guided by a Rule–Based System, España-Bonet, Cristina, Labaka Gorka, Díaz De Ilarraza Arantza, Màrquez Lluís, and Sarasola Kepa , Machine Translation Summit, 09/2011, Xiamen, China, p.554-561, (2011) Download: SMatxinT.pdf (398.46 KB)

A Hybrid System for Patent Translation, Enache, Ramona, España-Bonet Cristina, Ranta Aarne, and Màrquez Lluís , EAMT 2012, 05/2012, (2012) Download: hybridMOLTO4.pdf (156.49 KB); hybridMOLTO5.pdf (162.82 KB)

An IDE for the Grammatical Framework, Camilleri, John , Third International Workshop on

Free/Open-Source Rule-Based Machine Translation (FreeRBMT 2012), 06/2012, (2012)
 Download: An IDE for the Grammatical Framework - FreeRBMT 2012.pdf (507.44 KB)

MOLTO Enlarged EU - Multilingual Online Translation, Caprotti, Olga, and Ranta Aarne , 16th
Annual Conference of the European Association for Machine Translation, 05/2012, Trento, Italy,
(2012) Download: OnePageFinal4MOLTO.pdf (90.91 KB); EAMT2012.pdf (3.14 MB)

[1058] The MOLTO Phrasebook, Angelov, Krasimir, Caprotti Olga, Enache Ramona, Hallgren
Thomas, and Ranta Aarne , Third Swedish Language Technology Conference (SLTC-2010), 10//2010,
Linköping, (2010) Download: sltc2010.pdf (145.36 KB); MOLTO posterSLTC.pdf (387.85 KB);
MOLTO posterSLTC.pages (1.02 MB)

MT Techniques in a Retrieval System of Semantically Enriched Patents, Gonzàlez, Meritxell, Mateva
Maria, Enache Ramona, España-Bonet Cristina, and Màrquez Lluís , MTSummit 2013, Issue System
Demonstration, (2013) Download: molto_patents_demo.pdf (658.13 KB)

Multilingual Online Generation from Semantic Web Ontologies, Dannélls, Dana, Enache Ramona,
Damova Mariana, and Chechev Milen , www2012, 04/2012, Lyon, France, (2012)
 Download: wwweu2012_submission_16-final.pdf (1.22 MB)

[1108] Multilingual Packages of Controlled Languages: An Introduction to GF, Ranta, Aarne , CNL-
2010 (2nd Workshop on Controlled Natural Language), 14 September, Marettimo, Sicily, (2010)
 Download: cnl-2010.pdf (1.11 MB)

[2013] A Multilingual Semantic Wiki Based on Attempto Controlled English and Grammatical
Framework, Kaljurand, Kaarel, and Kuhn Tobias, Proceedings of the 10th Extended Semantic Web
Conference (ESWC 2013), (2013)

Multilingual Verbalisation of Modular Ontologies using GF and lemon, Davis, Brian, Enache Ramona,
van Grondelle Jeroen, and Pretorius Laurette , CNL 2012, Volume LNCS, (2012)
 Download: Multilingual Verbalization of Modular Ontologies using GF and lemon (Camera
Ready).pdf (212.21 KB)

[1299] Patent translation within the MOLTO project, España-Bonet, Cristina, Enache Ramona, Slaski
Adam, Ranta Aarne, Màrquez Lluís, and Gonzàlez Meritxell , Workshop on Patent Translation, MT
Summit XIII, 09/2011, p.70-78, (2011) Download: patentsMOLTO4.pdf (282.42 KB)

The Patents Retrieval Prototype in the MOLTO project, Chechev, Milen, Gonzàlez Meritxell, Màrquez
Lluís, and España-Bonet Cristina , WWW2012 Conference, Lyon, France, (2012)
 Download: patentsretrieval.pdf (397.01 KB)

[9999] Probabilistic Robust Parsing with Parallel Multiple Context-Free Grammars, Angelov, Krasimir ,
COLING 2012, (Submitted) Download: PMCFG.pdf (258.99 KB)

[1277] Reason-able View of Linked Data for Cultural Heritage, Damova, Mariana, and Dannélls Dana ,
The Third International Conference on SOFTWARE, SERVICES & SEMANTIC TECHNOLOGIES
(S3T), 09/2011, Bourgas, Bulgaria, (2011) Download: s3t2011_submission.pdf (1.18 MB)

[965] Robust Estimation of Feature Weights in Statistical Machine Translation, España-Bonet, Cristina,
and Màrquez Lluís , 14th Annual Conference of the European Association for Machine
Translation,EAMT-2010, (2010) Download: EAMT2010-EM.pdf (188.23 KB)

Smart Paradigms and the Predictability and Complexity of Inflectional Morphology, Détrez, Grégoire,
and Ranta Aarne , EACL (European Association for Computational Linguistics), 04/2012, Avignon,
(2012)

[986] Tools for Multilingual Grammar-Based Translation on the Web, Ranta, Aarne, Angelov Krasimir, and Hallgren Thomas , ACL 2010, July 2010, Uppsala, Sweden, (2010) Download: molto-demo.pdf (322.55 KB); MOLTO posterACL.pdf (982.89 KB)

Typeful Ontologies with Direct Multilingual Verbalization, Angelov, Krasimir, and Enache Ramona , Controlled Natural Languages Workshop (CNL 2010), 09/2010, Marettimo, Italy, (2011) Download: FinalSUMOCNL.pdf (252.38 KB)

Using GF in multimodal assistants for mathematics, Archambault, Dominique, Caprotti Olga, Ranta Aarne, and Jordi Saludes , 02/2012, Digitization and E-Inclusion in Mathematics and Science 2012, (2012) Download: ArchambaultCaprottiRantaSaludes.pdf (179.15 KB)

2.2 Journal Publications

Journal publication is a longer process than publication in conference proceedings so one might expect that it occurs after the end of a project's lifetime as archival medium for those results which are considered long lasting and of permanent value. In MOLTO we have already succeeded to list the following journal publication:

- Multilingual Sage, Saludes, Jordi, and Xambó Sebastian, Tbilisi Mathematical Journal, Volume 5, Issue 2, (In Press)

Books and proceedings has also been published and is currently being translated to Chinese,

- [1230] Grammatical Framework: Programming with Multilingual Grammars, Ranta, Aarne, CSLI Studies in Computational Linguistics, Stanford, p.350, (2011)
  - being translated to Chinese.

- Free/Open-Source Rule-Based Machine Translation, Online Proceedings of the FreeRBMT12, the Third International Workshop on Free/Open-source Rule-based Machine Translation, June 2012 Gothenburg, Sweden. Department of Computer Science and Engineering Chalmers University of Technology and University of Gothenburg Technical Report Number 2013:03, (2013). 1652-926X (ISSN).

- Controlled Natural Language, Third International Workshop, CNL 2012, Zurich, Switzerland, August 29-31, 2012. Proceedings Editors: Tobias Kuhn, Norbert E. Fuchs ISBN: 978-3-642-32611-0 (Print) 978-3-642-32612-7 (Online)

The following work has appeared as chapters of books:

- Creation and Integration of Reference Ontologies for Efficient LOD Management, Damova, Mariana, Kiryakov Atanas, Grinberg Maurice, Bergman Mark K., Giasson Frederik, and Simov Kiril, Semi-Automatic Ontology Development: Processes and Resources, 02.2012, Hershey PA, USA, p.162-201, (2012)

- Jeroen van Grondelle, Christina Unger: A 3-Dimensional Paradigm for Conceptually-scoped Language Technology [to appear], Towards the Multilingual Semantic Web, Springer, Autumn 2013.

Several junior members of the MOLTO team have completed their studies and written master or dissertation work related to tasks carried out as part of some work package. Some are continuing work that began in MOLTO as part of their thesis research. They include:

- Enache, Ramona (2011, Licentiate). *Automating the development of multilingual grammars*.

Göteborg : University of Gothenburg.

- Angelov, Krasimir (2011, PhD). *The Mechanics of the Grammatical Framework*. Göteborg: Chalmers University of Technology. Diss. ISBN/ISSN: 978-91-7385-605-8

- Virk, Shafqat (2012, Licentiate). *Computational Grammar Resources for Indo-Iranian Languages*. Göteborg : University of Gothenburg.

- Dannélls Dana (2012, PhD). *Multilingual text generation from structured formal representations*. Data Linguistica. University of Gothenburg. [pdf]

- Listenmaa I. 2012. Ontology-based lexicon management in a multilingual translation system – a survey of use cases. Department of Modern Languages, University of Helsinki. Download: listenmaa_masters_thesis_2012.pdf (1.05 MB)

- Ramona Enache (PhD, Forthcoming). Frontiers of Multilingual Grammar Development (prelim.). Göteborg : University of Gothenburg.

- John J. Camilleri (Master Sc., Forthcoming). A computational grammar and lexicon for Maltese. Göteborg : University of Gothenburg.
- Maarit Koponen (PhD, Forthcoming) Semantic Aspects of Machine Translation Quality Evaluation. Department of Modern Languages, University of Helsinki, Finland.

2.3 Project's Events

Project's Meetings

The project meetings were held every six months and always included a day which was open to participants from outside the Consortium: the MOLTO Open Day. The talks delivered by the MOLTO project members were targeted to a generic audience with no specific background assumed except for interest in the goals of MOLTO. The presentations are all available from the project's web site.

- MOLTO Project Kickoff Meeting. 9-11 March 2010, Barcelona, Spain.
- 1st MOLTO Project Meeting. 8-10 September 2010, Varna, Bulgaria.
- 2nd MOLTO Project Meeting. 9-11 March 2011, Gothenburg, Sweden.
- 3rd MOLTO Project Meeting. 31 August - 2 September 2011, Helsinki, Finland.
- MOLTO Enlarged EU (MOLTO-EEU) kickoff meeting. 12-13 January 2012, Gothenburg, Sweden.
- 4th Project Meeting. 7-9 March 2012, Zürich, Switzerland.
- 5th Project Meeting with MONNET. 17-21 September, Utrecht, the Netherlands.
- Final Project Meeting. 22-24 May 2013, Barcelona, Spain.

Workshops and Seminars

Additionally the project organized focused meetings:

- A. Ranta and R. Enache from UGOT gave a GF tutorial during the exchange visit to UHEL on May 4-5 2010.
- GF Meets SMT 1-5 November 2010, Gothenburg, Sweden.
- WP3 Seminar 13-14 June 2011, Helsinki, Finland.
- FreeRBMT 2012, the Third International Workshop on Free/Open-source Rule-Based Machine Translation was organized and hosted by UGOT between 13-15 June 2012 in Gothenburg,

Sweden.

- Public Sector Seminar 19 September 2012, Utrecht, the Netherlands.
- In house GF training for Be Informed. A [2-day GF course] (http://www.molto-project.eu/biblio/slide-presentation/gf-tutorial) was given at Be Informed (Apeldoorn) 11-12 Dec 2012 by Kaarel Kaljurand from UZH. The course covered all the material of the existing GF tutorials and additionally presented some GF-based applications (multilingual CNL-based semantic wiki system developed in MOLTO, and well as GF-based speech recognition grammars and smart phone applications).
- LEMON_GF workshop on 13-14 December 2013 in Bielefeld, Germany.

Summer Schools and Tutorials

The GF Summer School is a biannual event, these were partly sponsored by the project:

- Second GF Summer School, Frontiers of Multilingual Technologies. 15-26 August 2011, Barcelona, Spain.
- FORTHCOMING: Third GF Summer School, Scaling up Grammatical Resources. 18–30 August 2013. Frauenchiemsee island, Bavaria.

A number of GF tutorials were held during the past years during which the MOLTO tools are shown and actively used:

- Creating Linguistic Resources with GF LREC 2010 Tutorial, 17 May 2010. Malta.
- Grammatical Framework: A Hands-On Introduction. Tutorial at CADE-23, 1 August 2011. Wroclaw, Poland.
- Grammatical Framework Tutorial - ICFP 2012, Copenhagen 15 September 2012
- GF Tutorial at Shanghai Jiao Tong University, October 2012 by A. Ranta, 23-24 October, 2012.

2.4 Press Releases

The press releases were done at the beginning of the project and have been already reported in Deliverable 10.1. Each public event has been publicized by the local organizers via their official channels, so that announcements have appeared in calendars, bulletins and mailing lists.

2.5 Liaison Activities

Liaison with other EU-funded projects in the area of Computational Linguistics took place at international meetings where MOLTO was presented. The most relevant result to report however is joint work carried out with the MONNET project after organizing a joint meeting and a joint workshop reported before.

On December 13 and 14, 2012, PortDial members from Bielefeld met with Aarne Ranta (Grammatical Framework), Jeroen van Grondelle, Frank Smit and Jouri Fledderman from Be Informed, and John McCrae (lemon) in order to discuss the mapping from ontology-lexica to grammars, as well as the modular combination of induced domain grammars with dialog task grammars. The meeting gave rise to new ideas for the top-down grammar induction process being implemented. Moreover, the MOLTO-MONNET cooperation crystallized in the joint project proposal 611008- ADOPT coordinated by MONNET's coordinator Paul Buitelaar for combining the approaches, submitted as FP7-ICT-2013-SME-DCA but not granted.

MOLTO is a member of META-NET (http://www.meta-net.eu/) and more specifically of META-

Share. The MOLTO language technologies, resources, and tools are being distributed to members of the computational linguistics community under [LGPL](), consistent with the collaboration agreement signed with META-Share. As part of the liaison activities within META-NET, MOLTO also gave feedback for the final version of the strategy document for the META-NET agenda for Multilingual Europe 2020 (http://www.meta-net.eu/sra-en).

In January 2012, A. Ranta presented MOLTO at Xerox Research Centre Europe, Grenoble, in a seminar that has been video recorded and is published online at http://videos.xrce.xerox.com/index.php/videos/index/618.

MOLTO has been hosting the [FreeRBMT]() conference in June 2012, with a special workshop day devoted to explore the possible cooperation between Apertium (http://www.apertium.org/) and MOLTO: results are already tangible, especially with respect to the adoption of the lexicons from Apertium.

2.6 Web Channels

MOLTO used the World Wide Web as its main channel for continuous dissemination and archiving. The project's web site, registered at http://www.molto-project.eu, has been designed mainly to support the internal management of the project, several sections are open only to registered members of the Consortium. Gradually, as work progressed and results became available, we added some public sections however we have leveraged the possibility to be present on popular social sites to push news to the readers outside the Consortium, most prominently Twitter and LinkedIn. Recently we added a Consortium-only Google community page, which could in the future help maintain informal ties among the Consortium members and those interested in the future of the MOLTO technologies. It is not yet clear for how long the URL of the project's website will be maintained but we plan to freeze the contents shortly after the end of the project and to produce an archival version. The most important documents will be stored also as multimedia showcase, as required in Appendix X to Annex I.

In addition to the project's site, MOLTO has published multimedia content on:

- YouTube (http://www.youtube.com/user/moltoproject),
- Livestream (http://www.livestream.com/moltoproject) and
- Ustream (http://www.ustream.tv/channel/molto-project-channel) for the live broadcast of the FreeRBMT talks.

Screencasts for some of the MOLTO tools have appeared on Screenr (http://www.screenr.com/user/MOLTOproject).

Events of interest have been advertised via newsletters and mailing lists (MT, EAMT) and social sites, in particular:

- project website http://molto-project.eu
- project Twitter account: http://twitter.com/moltoproject
- LinkedIn group http://www.linkedin.com/groups?gid=3703935&trk=hb_side_g
- Google+ Community page https://plus.google.com/u/0/communities/102188132491619323731 where we ran a few Google hangout online seminars
- elearningeuropa.info presence at http://www.elearningeuropa.info/directory/index.php?page=doc&doc_id=1621...

Partners have featuring the MOLTO work on their websites (searching `link:molto-project.eu` yields about 45 hits).

The project coordinator and the workpackage leaders have been reachable for questions by a contact

form accessible online. Recurring questions have been answered in the FAQ: http://www.molto-project.eu/view/faq, commonly edited by all registered users.

The publication list appearing on the website is an extensive reference list of the results of the project. It includes also software and other media. The RSS feed for the publications appearing in MOLTO is http://www.molto-project.eu/biblio and currently lists 224 items, many of which are the slide presentations delivered during the project's related events.

3. Exploitation Plan

Given the general direction of the field of language technologies, as outlined in the strategic research agenda for LT2020, the exploitation of MOLTO results will focus on the high-quality translation services in the cloud. These cloud services may serve the public sector as well as a more technical audience. The case studies have shown the versatility of the MOLTO technologies in terms of domain of application, scalability, and target audience.

Exploitation of the project's results and acquired experience also goes towards furthering the field of language technologies. This has already been demonstrated by e.g. the work on the lexicon resources also with respect to usability of publicly available semantic web resources.

3.1 Project's Outcomes

Several of the project's deliverable are of interest for further upkeep and will be maintained and developed in the future. MOLTO tools and technologies that have been released include several multilingual translation web services, grammar writing IDEs, guidelines and tutorials, a translation platform integrating the editing tools, and sample multilingual software such as a dialog system, query interfaces, a multilingual semantic wiki. In addition, the events that have been organized during the project's lifetime aimed at capacity building, both in academia and in industry. Young academicians worked with the commercial partners on very concrete problems and had to learn how to communicate with non-experts. Similarly, the industrial partners had to identify the tasks and issues that could best be solved by asking to the academic partners.

In what follows, the commercial partners outline the areas in which these newly created cooperation ties may in future be consolidated.

3.2 Exploitation Strategy

We identified three different strategies concerning exploitation:

- *Open Source Strategy*: the project has adopted this strategy for the release of the final products. All software and tools are available under LGPL. Some of the technologies are under continuous development, as it is usual in the open source community, and can be adopted and commercially further developed by branching the repositories.

- *Spin off Strategy*: this strategy is currently under discussion, interested parties are evaluating whether to provide a spinoff consultancy firm to provide GF and grammar/multilinguistic knowledge that companies might not have readily available.

- *Commercialization Strategy*: this is the strategy by the commercial partners, outlines in the following sections.

3.3 Structure of explanation plan

The project members discussed and agreed upon the use of the method from Stähler(1) to develop commercial exploitation plans. This results in the following contributions from each industrial partner:

- A company profile, which is a short description of the nature of the company: what it does; how large it is; how it makes money; and how it typically transfers research into products.
- A list of identified exploitation opportunities

For each promising opportunity the method from Stähler was followed to develop a plan to exploit outcomes. This results in the following sections corresponding to the method phases depicted in the figure below.
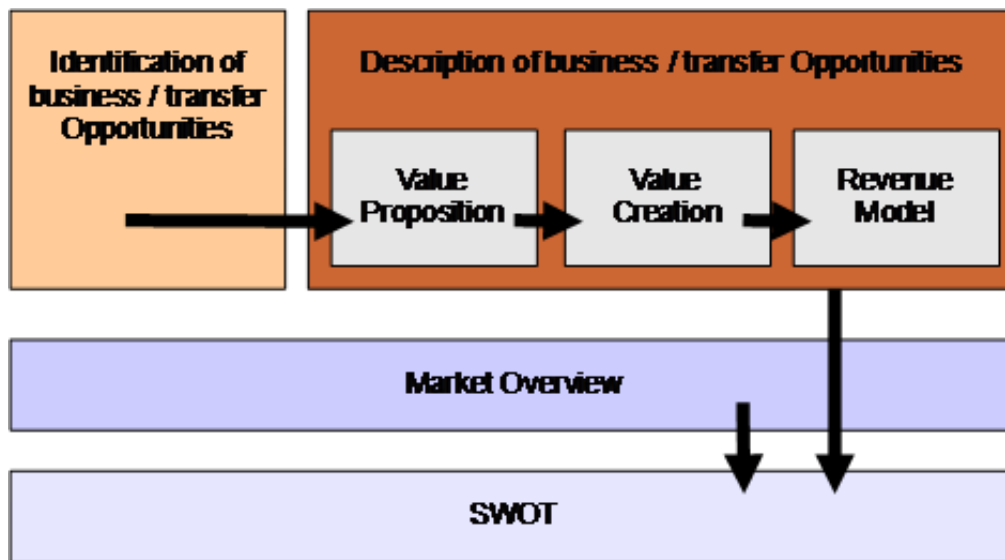
- a description of each opportunity identified
- a description of the markets associated with each opportunity
- a strengths-weaknesses-opportunities-threats (SWOT) analysis for each identified opportunity and market

(1)Stähler, Patrick; Geschäftsmodelle in der digitalen Ökonomie: Merkmale, Strategien und Auswirkungen. Josef Eul Verlag, Lohmar, 2001

3.4 Be Informed Exploitation Plans

Company Profile

Be Informed is an internationally operating, independent software vendor. The Be Informed business process platform transforms administrative processes. Thanks to Be Informed's unique semantic technology and solutions, business applications become completely model-driven, allowing organizations to instantly execute on new strategies and regulations. Organizations using Be Informed often report cost savings of tens of percents. Further benefits include a much higher straight-through processing rate leading to vastly improved productivity, and a reduction in time-to-change from months to days.

The role of Be Informed in MOLTO is to make sure that the solutions developed in the project can indeed be readily integrated into their solutions (the Be Informed Business Process Platform in

particular). Be Informed will build on its strong expertise in its domain to guide the project and make sure that the results are exploitable from a commercial point of view in the mid-term. Dissemination to, and feedback from, its client base, as part of the use case development in WP12, will increase the degree of suitability for exploitation.

Be Informed's exploitation strategy is tightly linked to its goal of quickly commercializing MOLTO results, and calls for a rapid and continuous flow of information to its sales force, existing client base and potential future customers. In addition, as an innovative company, Be Informed plans academic talks and publications.

Products Relevant to Opportunity

The outcome of MOLTO is relevant for Be Informed's Business Process Platform. For both client and server product components the translation services based on the GF based prototype can offer translation support at design time as well as runtime. This would enable several usage scenarios to deal with verbalization activities of customers business models and others artefacts.

For more detailed information about this product and its solutions see www.beinformed.com.

Research Transfer Process

The main approach of Be Informed Research and Innovation is based on co-innovation with customers, partners, and other third parties. These activities usually result in a working prototype. Prototypes which seem promising to get enough traction with customers are handed over to Be Informed Product and Solution development.

The MOLTO deliverables will be promoted to our clients and partner in the public sector. The prototype of the MOLTO multilingual verbalization component for integration with Be Informed Business Process Platform will be made available as an optional product component.

Relevant Trends in business domain

In this section we present a concise overview of relevant public sector trends and views within and across European Union Member States on future public services. This background information is not only necessary to understand the societal and political context in which multilingual public sector services take place, but also to detect synergies (and potential divergences) between visions about ontology driven services, language aspects and current developments within the public sector. The presented overview is inter alia based upon recent studies by the OECD (Towards Smarter and more transparent Government, e-government status spring 2010; OECD e-Government project; 25 March 2010; GOV/PGC/EGOV(2010)) and research results from the CROSSROAD Project (A Participative Roadmap for ICT Research in Electronic Governance and Policy Modelling; a support action under the European Commission 7th Framework Programme. http://crossroad.epu.ntua.gr/the-project/objectives/FP7-ICT-4-248458).

Within the context of this project we are dealing with public sector services that provide information and advice and perform transactions between citizens or companies and administrations. By using ontologies which contain concepts, their relations and respective rules, public sector services become decision centric and goal driven. This enables the public sector to become more agile, customer centric, efficient, effective and accountable as well.

In this section we will use the concepts of Governments and Public Sector interchangeably. Political institutions and administrative structures of counties are diverse, but regardless of their shape, they are all

part of the Public Sector ecosystem that provides public sector services to citizens and companies or institutions. Governments in Europe face an increasing number of challenges such as ageing populations, immigration, climate change and globalization, further reinforced by the financial crisis. The globalization trend has limited the freedom of governments to manage their national economies and new challenges such as immigration and an ageing population seem to fundamentally affect the scope of public sector activities. At the same time, society's expectations of public service delivery have by no means diminished as citizens from the 1980s onwards have become more concerned with choice and service quality. The paradox faced is one of open-ended demand versus a capped or falling resource share for actual delivery. Consequently, public administrations are under constant pressure to modernize their practices to meet new societal demands with reduced budgets.

In the Visionary Scenarios Design of the CROSSROAD Project, the researchers present a summary of the main trends with respect to ICT for governance and policy making in the wider context of an evolving public sector. They define a set of core policy trends across the governance and policy modelling domain, which also resonate with the use case settings of the MOLTO project.

1. Greater transparency and accountability of the public sector. A demand for a more transparent and accountable government can be discerned. Many EU Member States have put transparency and accountability policies in place.
2. Improved accessibility of public services. An increased awareness and perception of the needs and wishes of citizens, results in a drive towards more choice and accessibility of public services.
3. Quality, efficiency and effectiveness of the public sector. Many policies are aimed at delivering cheaper solutions while ensuring quality. An increased attention is given to efficiency, as in many sectors government institutions face considerable budget cuts. This trend is particularly driven by dwindling public finances.
4. New models of governance and the emergence and active participation of new stakeholders. A trend that can be discerned in most public sector domains is the emergence of new partnerships, the involvement of intermediaries and the acknowledgement of new stakeholder roles. Citizens, civil society, advocacy groups are increasingly empowered to organise themselves and play a role in public service delivery.
5. Stronger evidence based policy. A resurgence of governance models that value principles such as accountability, monitoring and evaluation reaffirms the principles of evidence-based policy as a necessity for making informed decisions.
6. Citizens' empowerment, expression of diversity, choice. The role of users is re-valued in a way that acknowledges their new found skills and growing empowerment. The principles of facilitating increased participation, user created content, user engagement, increased independence and ownership of public services applies to all public sector domains.
7. Improved digital competencies, bridging the digital divide. As in all domains technologies increasingly play an important role in the provision of public services, in all sectors questions arise as to the ICT skills of citizens required to have access to those services.
8. Promotion of independent living and self-organisation. Policy makers acknowledge that ICTs can play an important role for inclusion of all citizens and in order to achieve social equity and cohesion. In many countries ICT policies aim at enhancing the independence of citizens – for instance elderly or disadvantaged groups.

Within the context of this project we are dealing with public sector services which provide information and advice and perform transactions between citizens or companies and administrations. This type of services is decision centric by nature. They are dealing with rights, permissions and obligations, for instance in the domain of permits and grants. The activities that have to be supported by the services are knowledge intensive. Another characteristic is that they are event driven. This makes them perfect candidates for semantic enabled services. Ontologies are situated at the core of this kind of services.

We have to take into account that ontology support for public services is not only positioned at the end of the service chain, where government and citizen meet each other, but throughout the whole service

chain. Treating a request for a permit and deciding upon this request is based upon the same rules as getting advice whether one is entitled to acquire the permit. So, the concepts and rules that are used in ontologies apply as well to the citizens interactions as to the administrative officials interactions. The need for localization can however differ between these two target groups. In a traditional view public sector services are positioned at the execution and enforcement layer of the public sector infrastructure. This layer deals with policy implementation. For reasons of scoping we will focus in this stage of the project also on this policy implementation layer.

We foresee however a trend in which the use of ontologies will go more upstream towards the policy making process, since this will leverage the best outcome.

Be Informed: Multilingual verbalization of Business Models

Value Proposition

Main beneficiaries of the Molto outcomes are domain experts using Be Informed in an international context in the public sector. These public sector services provide information and advice and perform transactions between citizens or companies and administrations. This type of services relies by nature highly on interaction and communication on the one hand and the execution of regulations on the other hand. The quality of both aspects must be guaranteed. We will describe in brief scenarios of public sector actors like domain experts that are confronted with localization aspects for the services they are providing or intend to provide. These scenarios are:

1. National Government with International clients
2. National Government cooperating Internationally
3. National Government dealing with International Law/Policies
4. National Government in Multilingual Countries
5. International Government (European Union)

In all scenarios we can see that, although policy making and implementation seems to be mostly a local (national) issue, there are very often also international issues/aspects that have to be taken into account.

National Government with International clients

A very common pattern in the world is the provision of public sector services in the field of immigration. Immigration services have to be provided to immigrants who want to work and/or live in another country and to companies or organizations who want to hire labour resources from another country. A specific kind of stakeholder is the group that wants to bring family members to the country they live in. The main process is the issuing of permanent or temporary/provisional permits for admission and residence. A crucial characteristic of this process is that the rules for admission and residence are changing frequently and sometimes with short notice. Since immigrations offices are communicating with 'the whole world', one cannot expect them to translate their services into all languages. Normally they will use the language or languages of their own country and maybe one or a few other languages that can be understood by the majority of their customers. And, in specific cases, they will want to translate a part of their information to a specific target language. This can be the case for instance as due to a certain incident a new group of immigrants from an individual country 'threatens' to flood the country. So they need a process that supports the translation of services to the current languages on a regular and flexible basis and an approach to deal with incidents that require instant translations in the non-current languages. In all cases it is a challenge to translate the complicated immigration laws and procedures into comprehensive services for national and international users.

National Government cooperating International

An example of a government agency that has to cooperate internationally is the Dutch Emission Authority (NEA). Emissions trading is a flexible policy instrument which governments use to improve the living environment. In the Netherlands there are two emissions trading systems, one for emissions of carbon dioxide (CO_2) and one for emissions of nitrogen oxides (NOx). Emission trading requires an infrastructure for issuing permits, monitoring and allocating emission allowances. Emissions trading is inevitably an international business that requires cross boundary cooperation, information and communication. The public services of the Dutch Emission Authority must therefore be available and accessible in more than one language. In this case NEA wants to make its service also available in the English language.

Trading requires international agreement on standards and preferably also on service patterns. By using one information concept it becomes easier to exchange information and to innovate. In such a case the ontology supported infrastructure of a frontrunner in the specific domain, such as NEA, could be used as a basis for internationalisation and standardisation.

National Government dealing with International Law/Policies

The times of splendid isolation are over (if they ever existed); we are living in a dynamic international world and an increasingly more global market. One of the government parties that is affected daily by this trend is Customs. They have to deal not only with local laws, but also with common market regulations, international trade regulations etcetera. The regulations, they have to comply with, and have to enforce, change frequently, based upon incidents, new insights and political developments. And within a set of regulations, the priorities for enforcement can change too.

Customs have to deal with international treaties about traffic of goods between countries and the limitation thereof. For example for importing certain goods from China, one has to apply for an export license in China which is transformed to an import permit in the country of destination. This leads to multilingual public services that are delivered in different countries of the world. Depending on the types of goods there might be an additional import tax to protect a country's internal market from being 'flooded' with low price goods from low cost countries.

In order to be able to levy additional tax on certain goods one must be able to classify these goods. The EU defined the Combined Nomenclature, which is in fact a taxonomy of goods and their codes that can be used to classify goods that enter a country. This taxonomy is available in all official countries of the European Union. The taxonomy is based on the Harmonized Commodity Description and Coding System7 which is run by the World Customs Organization. The harmonized system is used by 137 countries and the European Union

National Government in Multilingual Countries

Many countries are bi-lingual or multilingual. This means that all official publications and services have to be provided in more than one language. Often the pilot language, the language in which a document is written first, depends on the preferred language of the author. By using an ontology, the meaning of the document in the pilot language can be expressed abstractly and unambiguously in concepts and rules. They can then be translated into a particular language to express the meaning using the vocabulary and syntax of that language.

Value Creation

Be Informed captures policy in ontologies. These ontologies are used throughout the policy lifecycle from choosing/deciding on policy, communicating the agreed upon policy to all stakeholders to running

the supporting applications. As a consequence, verbalizations of these ontologies could be used in a number of scenarios throughout that policy lifecycle.

Review, Validation and Feedback of Models

For the ontologies to be used as the basis of actual applications, it is crucial they contain a correct representation of the requirements and constraints. Review and validation before deploying and the ability to provide feedback on the model after deployment is very important. A natural language representation of the models can help stakeholders to exercise these tasks. Special verbalization choices might have to be made to create texts that are effective in this specific scenario.

Text based Editing of Models

The most effective way of business user involvement is of course allowing them to create models themselves or, often more realistic, to maintain and alter existing models. In [EKAW2010] we explored editors that do use a textual metaphor to present models to the users, but that do not use typing text as editing metaphor.

Self Documenting Models

Typically, systems need to be well documented for IT organizations to be able to support production use and perform maintenance. The online, navigational access to the models is then often not acceptable, and conventional documentation sets need to be generated.

Textual UI's for Model Driven Applications

Classically, business applications have used tables of data to present detailed information that is available in a business process. When involving customers in business processes, they find it hard to interpret the data. Verbalization into natural language can be a great way to present, for instance, process progress data to laymen, as the data can be presented in a self explanatory way.

Communicating Model Based Decisions

The ontologies capturing legislation and policy are used to drive decision services, applying the policy to actual cases. These decisions taken are communicated to the stakeholders and need to be documented and explained. Verbalization of the model could be extended to verbalization of the decisions based on the models.

Revenue Model

The proposed exploitation path would increase revenues of existing products like the Be Informed Business Process Platform. Be Informed will offer the Molto verbalization engine as an optional product component. It is difficult to predict the size of the increase at this stage of development.

Market Overview

Ontology translation systems are usually created using general-purpose programming languages, such as LISP or Java, and the mappings between expressions in the source and target languages are neither well-documented nor explained. Integrated tooling as part of Be Informed's Business Process Platform is at this stage unique.

3.5 Ontotext Exploitation Plans

Introduction

Ontotext's business model combines the development of products (including some open source versions) with the provision of research, consultancy and development services. Many commercial projects combine all four elements. For **Ontotext** MOLTO will bring the unique opportunity to strengthen its position in the semantic technologies and knowledge-driven text analytics market, with development and adoption of intelligence methods that support ontology-based multilinguality. This will be possible due to the fact that MOLTO adds to the semantic technologies the GF formalism, which operates as an interlingua on language level and thus, localizes the ontologies in appropriate ways. More precisely, the main directions of future development will be as follows:

- Interoperability and grounding in Linked Open Data resources and domain ontologies
- High throughput, multilingual text processing
- Robust cross-lingual translation of various domain data within search and retrieval services

The business strategies will be as follows:

- The company will put its technology (in this case, the KIM Semantic Annotation Platform, http://www.ontotext.com/kim; and Publishing tools) into a stronger multilingual context.
- OntoText will strengthen its synergies between the semantic and world-knowledge infrastructure modules, and the MT services.
- The task of Combining the GF model and the ontology standards would test and enrich the reasoning platform, developed and maintained at Ontotext.

Company Profile

Ontotext AD is the strongest semantic technologies company in Europe and a world-leading supplier of core semantic technology, text mining and web mining solutions.

- **Established in year 2000**, today Ontotext has over 65 employees located in Bulgaria (Sofia and Varna), USA (Fairfield, CT) and UK (London);
- After acquiring VC funding in 2008, at the end of 2010 Ontotext reached break-even and since then doubles its commercial revenues annually.
- We are global leader in **semantic database engines**, successfully competing with ORACLE, IBM, and Microsoft in this field;
- Our unique competences are backed by heavy investment in R&D – over the last 12 years we have invested more than 300 person-years in semantic technology. We know what works and what does not!

We have unmatched portfolio of world-class technology and expertise in:

- **Semantic Databases**: high-performance RDF DBMS, scalable reasoning;
- **Semantic Search**: text-mining (IE), Information Retrieval (IR) ;
- **Web Mining**: focused crawling, screen scraping, data fusion;
- **Linked Data Management and Data Integration**.

The main differentiator between Ontotext and other semantic technology vendors is that we deliver robust technology, proven in multiple high-profile projects that justify its maturity and usability. The best example in this direction is the usage of OWLIM (our RDF database engine) in the BBC FIFA World Cup 2010 website where most of the pages were generated dynamically through queries to OWLIM –

millions of requests per day, hundreds of updates per hour, handled by a cluster of few servers. Following the success of this project, BBC extended the use of Ontotext technology for the BBC Sport website and for the London Olympics 2012 website.

Ontotext's clients span across several sectors:

- **Pharma**: AstraZeneca, UK, and UCB, Belgium;
- **Media and publishing**: BBC and Press Association, UK; and Publicis, Germany;
- **Telecommunications**: Telecom Italia and Korea Telecom;
- **Archives and cultural heritage**: The National Archive, UK, the British Museum, the Dutch Public Library;
- **Government**: Department of Deffence, USA; and House of Commons and TNA, UK; Natural Resource Canada.

Considering the substantial number of clients of Ontotext in UK, we are running in London regular open **training courses** "Semantic Technologies with OWLIM", usually scheduled at roughly once per quarter.

Products Relevant to Opportunity

- OWLIM is an industrial-scale semantic database, using Semantic Web standards for inference and integration/consolidation of heterogeneous data.
- KIM Platform is a semantic search engine, using text analysis to provide hybrid queries involving structured data and inference.
- FactForge is a public service that represents a reason-able view to the web of data.
- Linked Life Data is a platform for semantic data integration trough RDF warehousing and efficient reasoning that helps to resolve conflicts in the data.
- Publishing platform – semantic publishing platform, ingesting and enriching thousands of news with linked data daily; enables the publishers and third-parties to explore innovative business models and alternative revenue streams.

Research Transfer Process

On the one hand, the research goes into products through the traditional ways:

- creating prototypes in use case domains, and then
- scaling these prototypes into systems for real usage.

On the other hand, the developed technology within the project is applicable to other related areas of NLP services application. It can be either used as stand-alone applications, or be integrated into larger and more complex architectures. Both business opportunities have significant added value.

The first direction is exemplified by the envisaged use case domains: Patents in medical domain and artefacts in Cultural heritage domain. The second direction goes to areas that apply strongly Question Answering, Information Retrieval and MT. Such areas are: Publishing, Social Media and Pharma. The related products are highly commercial and thus, precision and relevance of the retrieved information are crucial features for the clients. GF formalism would be useful for the smoothing of the multilingual retrieval and translation results. It must be noted that the component shared by all targeted products of Ontotext is the ontology-based knowledge that relates to LOD and multilingual settings.

All the EU research projects that Ontotext has been involved in, have lead to the improvement of the current technology as well as to the creation of new products, that have been explored in commercial projects. In this way, we might view the Research as an Investigation, Preparation and Compilation

phase, while the applications in Industry – as Adaptation, Harmonization and Real Setting evaluation phase. Below some synergies of the aforementioned kind are given:

1. RENDER is an ongoing project that aims at providing a comprehensive conceptual framework and technological infrastructure for enabling, supporting, managing and exploiting information diversity in Web-based environments. It also would leverage very large amounts of content and metadata: news, blog and microblog streams, content and logs from Wikipedia, news archives, multimedia content and reader comments, discussion forums, etc. This data is managed by a highly scalable data management infrastructure, and enriched with machine-understandable descriptions and links referring to the Linked Open Data Cloud. **This development would lead OWLIM and KIM technologies to handle diverse data, which would widen their data coverage and management.**

2. CUBIST is an ongoing project that aims at *Combining and Uniting Business Intelligence and Semantic Technologies with a special focus on unstructured data mining. Being central to the project goals, the semantic technology supports a persistent layer – a semantic Data Warehouse.* **The project adds to the better Visual Analytics, whose improved characteristics would be important for providing more competitive user interfaces in industry.**

3. MediaCompaign is innovative in Ontology creation for cross-media modelling of media presence and campaigns; Semantic cross-market product data interlinking; Identification and tracking of new media campaigns in different media and countries. MediaCompaign focused mainly on advertisement campaigns and their impact on attitudes and opinions. **Thus, the publishing services, provided by Ontotext, will be enriched with sentiment analysis additionally to the knowledge-based analysis. Thus, Ontotext will have a social-aware service.**

4. NoTube project concentrated on personalized semantic news; personalized TV guide with adaptive advertising as well as Internet TV in the Social Web. It relied on the key role of the semantic technologies, taking into account the community aspects and is built on multilinguality. **The results strengthened the personalized component in the retrieved information in commercial publishing platforms.**

5. PHEME project (will start in October 2013) has as its main goal the development of scalable methods for Social Semantic Intelligence, across media and languages. I also aims at modeling not only facts and opinions, but also the parameters of reliability of the information sources. Additionally, PHEME focuses on more concrete and socially crucial cases in recent years, such as crowdsourcing, citizen journalism and bioinformatics. PHEME goes beyond official media campaigns - to social network dimensions and beyond the opinions – to rumour and misinformation detection. **This project will lead to a large-scale social media bound OWLIM and KIM platforms. Also, it will add to its services the identification of misinformation, which would be very valuable facility in the personalized component for the end users.**

In all productizing areas, listed below, the following underlying NLP technology is assumed:

- Multilingual semantics based question answering
- Cross-language retrieval
- Public Translation Service (combining GF + ML)

Relevant Trends in business domain

With the globalization processes and harmonization of large groups of documents in EU, the requirements for particular data management systems is rapidly growing. Additionally, virtual space has become more populated, shared, explored and multilingual. For example, virtual tours in famous museums; virtual storage and access to EU legislation; interactive online digests; electronic government;

digital preservation storages; social networks etc.

For these reasons, it is not surprising that some of the most active business domains at the moment are the Cultural heritage stakeholders (DARIAH, CLARIN, EUROPEANA); Pharma (Astra Zeneca); Media Publishing (BBC, NDP, Press Association) and Social Media (Pheme project).

Ontotext is involved in all of the aforementioned domains through research projects and commercial projects.

Ontotext : Productize MOLTO Technology in novel areas

Publishing platforms and Digital Journalism

The cross-media analytics is a typical case of business intelligence, developed at Ontotext. Ontotext's technology covers preferably (but not only) publishing agencies (such as, Press Association, NDP, Oxford, etc.) and government data management (US government). From a language point of view, the company has been working systematically on commercial project for Dutch and English. However, lately, it started to expand the multilingual set to Bulgarian, German, Chinese, etc. Having in mind these facts, GF formalism as well as RDF-GF interoperability from MOLTO would be the natural extension of the information extractors, thus facilitating the interaction between the users' queries and their machine processing. More precisely, the following extensions are envisaged: embedded translator service, tuned to the domain (sports, finance, politics, etc.); embedded converter from RDF representation to GF and then to language, and vice versa.

Also, internally, the semantic annotation tool will be augmented with language localization modules that would support the annotators.

Related markets: publishing; electronic government

SWOT Analysis:

- **Strengths:** Improvement of the existing multilingual modules; creation of new functionalities to the customers, such as viewing the same result in various languages; improving the annotation process and text analytics; better communication between the ontology and user queries.

- **Weaknesses:** Domain adaptation of the MOLTO modules might be needed, when addressing a new domain or even a subdomain of a specific domain.

- **Opportunities:** There might be the possibility to create a publishing platform of new generation, which provides a typological core for many languages and thus – is easily adaptable to new languages. Additionally, to see Dutch news in English within the publishing system itself, for example, would extremely facilitate the customers.

- **Threats:** The online real time applications might be unstable initially due to the complex architecture.

Social Media

In Ontotext projects the existing LOD resources (such as, Linked Life Data) are applied for different socially aware domains and across languages. For example, the entity extraction tool LUPedia as well as

the linked data concept store [FactForge](#) will be used in enhancing the socially marked knowledge. These modules will be extended by the language generation tool from MOLTO in order to improve the accuracy of the extracted information. This step is manageable, since the MOLTO rule-based translation technology is extended with the help of statistical approaches.

Related markets: education, tourism

SWOT Analysis:

- **Strengths:** MOLTO gives the possibility of applying a structured approach to unstructured data for the purposes of good understanding of big amounts of data.

- **Weaknesses:** MOLTO might support better some forms of Social Media (publicly available), while some others (restricted) – not so well.

- **Opportunities:** The social media might be viewed as a network of subdomains and addressed by MOLTO technology in a step-by-step way.

- **Threats:** No visible possibility is foreseen at the moment for using MOLTO modules directly in sentiment and opinion analysis.

Pharma

Ontotext regularly participates in projects that consider health care and life sciences data management (there is Life Science project running now). Here the available domain ontologies are explored together with the NLP processing. GF will be extremely useful since both the prescriptive and diagnosis languages are controlled. There is an additional level of translation here, namely: from the specialized prescription and anamnesis language of doctors into the common natural language of the users.

Related markets: Medical producs sales, health care

SWOT Analysis:

- **Strengths:** MOLTO is best performing in controlled and structured domains. Pharma is a good example of such a combination. In addition, there is an already working prototype on Patents in this domain.

- **Weaknesses:** Pharma would be better manageable from doctors' production point of view, rather than from patient perspective, since professional language is better controlled.

- **Opportunities:** Improvement of multilingual search and relevance of the search results.

- **Threats:** Pharma is one of the well elaborated domains from a processing point of view. Thus, the real added value of MOLTO is to be tested in the future.

Ontotext : Productize in the use case domains: Patents and Cultural Heritage

Patents

Part of the commercial projects, carried out within Ontotext, are connected to pharmacies. In this

respect, the developed prototype in bio-medical and pharmaceutical domains will be employed directly in the workflow processes.

Related markets: Administration, government, science, businesses

SWOT Analysis:

- **Strengths:** The usage of patents is a common and necessary activity in industry. Thus, the created structure model for handling patents in one specific domain would be applicable to patents in other domans, too. Retrieval services in a strongly cross-lingual context would be also very attractive features for exploitation.

- **Weaknesses:** If the MOLTO modules are used in another domain of patents (for example science) some adaptation will be needed, although the patent structure itself would be stable beyond specific domains.

- **Opportunities:** The usage of the patent service would facilitate and speed up the process of managing Pharma policies with respect with new development in healthcare.

- **Threats:** The patent service might not cover all the query requirements of the users due to the limitations of the controlled language or the incompleteness of the corpus.

Cultural Heritage

There are many stakeholders in this area, since lately the related initiatives have grown considerably. Here we have in mind the specific ones: Europeana, British Museum, ConservationSpace and CLARIN.

Europeana already provides search facilities. However, they cover only metadata and are not connected to ontologies. Also, the translation from one language to another is done via machine translation(MT) only, without any grammatical formalism behind it. British museum is a partner which would try the MOLTO services, profiled specially for museum objects. They already use Ontotext's semantic repository OWLIM. This service supports semantic search, semantic RDF data sources, Web Publication. MOLTO will add to the better search functionality as well as to the multilingual information extraction. Similar projects are: Gothenburg City Museum (Sweden); Polish Digital National Museum; Yale Center for British Art (USA): Linked Open Data publishing of museum collection. ConservationSpace project is managed by the National Gallery of Art (USA) and 7 other institutional partners from the USA, UK and Denmark. It handles the data management. MOLTO services might also contribute to the better preservation of the documents through adopting the GF formalism as a mediator between the users' queries and SPARQL queries. Similar projects are: FP7 CHARISMA: Synergy for a Multidisciplinary Approach to Conservation/Restoration; FP7 3D-COFORM: 3D documentation and collection formation of tangible cultural heritage; CLARIN is a pan-European initiative, which aims at elaborating also a globally shared service, among other services, for exploration of cultural artefacts. Ontotext is a participant in this initiative. It might provide the same facility to the consortium as in the above opportunity. Similar projects: FP7 V-MUST: Virtual Museum Transnational Network, a Network of Excellence.

Related markets: tourism, education

SWOT Analysis:

- **Strengths:** coverage of many languages with language specific mediated filtering (through GF formalism); high precision of the retrieved content due to the controlled language; easy adaptability to other areas of cultural artefacts and languages.

- **Weaknesses:** Since one of the use cases in MOLTO considers museums, the application to other subdomains of Cultural heritage might need adaptation of the grammars and ontologies.

- **Opportunities:** The service can be adopted by various virtual cultural databases and adapted to them.

- **Threats:** There might not be available resources for certain languages or language variants; language generation might not be efficient enough for all languages.

4. Conclusions

The project dissemination activity has focused on three major stakeholders' groups: researchers in NLP, public sector, semantic web technologists. These have been reached by organizing events at international meetings, on online social platforms, and face to face. A major outcome of the project is the ongoing discussion between the academic developers of the MOLTO technologies and the commercial partners based on the work carried out. This discussion concerns the future mechanisms that should be created so that the MOLTO results can be successfully adopted for exploitation. It has become clear during the case studies and the evaluation of the project that the fast developing technologies used in MOLTO need to become mature before they can be used commercially. Moreover, it would be desirable to be able to offer professional support, consultancy services and training in order to promote the uptake of the project's translation services.

Source URL: http://www.molto-project.eu/wiki/living-deliverables/d104-molto-dissemination-and-exploitation-report