



RENDER
FP7-ICT-2009-5
Contract no.: 257790
www.render-project.eu

RENDER

Deliverable D6.3.2

Report on clustering and liaison activities Year 2

Editor:	Delia Rusu, JSI
Author(s):	Delia Rusu, JSI; Fabian Flöck, KIT
Deliverable Nature:	Report (R)
Dissemination Level: (Confidentiality)	Public (PU)
Contractual Delivery Date:	30 September 2012
Actual Delivery Date:	30 September 2012
Suggested Readers:	Everybody interested in the clustering and liaison activities carried out during the second project year
Version:	2.0
Keywords:	Collaboration activities, research projects

Disclaimer

This document contains material, which is the copyright of certain RENDER consortium parties, and may not be reproduced or copied without permission.

In case of Public (PU):

All RENDER consortium parties have agreed to full publication of this document.

In case of Restricted to Programme (PP):

All RENDER consortium parties have agreed to make this document available on request to other framework programme participants.

In case of Restricted to Group (RE):

The information contained in this document is the proprietary confidential information of the RENDER consortium and may not be disclosed except in accordance with the consortium agreement. However, all RENDER consortium parties have agreed to make this document available to <group> / <purpose>.

In case of Consortium confidential (CO):

The information contained in this document is the proprietary confidential information of the RENDER consortium and may not be disclosed except in accordance with the consortium agreement.

The commercial use of any information contained in this document may require a license from the proprietor of that information.

Neither the RENDER consortium as a whole, nor a certain party of the RENDER consortium warrant that the information contained in this document is capable of use, or that use of the information is free from risk, and accept no liability for loss or damage suffered by any person using this information.

Full Project Title:	RENDER – Reflecting Knowledge Diversity
Short Project Title:	RENDER
Number and Title of Work package:	WP6 - Dissemination, exploitation, and community building
Document Title:	D 6.3.2 - Report on Clustering and Liaison Activities Year 2
Editor (Name, Affiliation)	Delia Rusu, JSI
Work package Leader (Name, affiliation)	Ioan Toma, STI
Estimation of PM spent on the deliverable:	1

Copyright notice

© 2010-2013 Participants in project RENDER

Executive Summary

During the second year of the RENDER project, we continued our collaboration with networks of excellence and started new collaborations with European projects and universities from the United States and Canada.

The networks of excellence are **Pascal**, **MetaNet** and **PlanetData**, and the collaboration points have been described in the Year 1 Report on clustering and liaison activities. In this deliverable we present a follow-up of these activities.

RENDER participated at the EU Project Networking session hosted by the Extended Semantic Web Conference (ESWC) 2012. The networking session provided the opportunity to initiate new collaboration activities with other EU projects: **XLike**, **ENVISION**, **LOD2**, **ViSTA-TV**.

XLike aims at developing technology to monitor and aggregate mainstream and social-media knowledge and enable cross-lingual services; ENVISION offers support in the process of semantic discovery and composition of environmental services. LOD2 has a number of goals such as improving coherence and quality of data published on the Web and improving the performance of RDF data management. Vista-TV is a project aiming at creating real-time TV recommendations for viewers.

RENDER also initiated a collaboration with the LT-Web project, which will propose a W3C standard on textual enrichments.

Finally, RENDER defined two new common tasks with universities from the United States and Canada, on the development of a Wikipedia tool to track edit provenance and tools for automatically updating Wikipedia based on news sources and social media.

In the third year of the project we are going to continue the collaboration established thus far with the aforementioned projects and report on the progress being made, as well as foster collaborations with other initiatives.

List of authors

Organisation	Author
JSI	Delia Rusu
KIT	Fabian Flöck

Table of Contents

Executive Summary	3
List of authors	4
Table of Contents	5
Abbreviations.....	6
1 Introduction	7
2 Collaboration Activities	8
2.1 EU Project Networking.....	8
2.1.1 XLike.....	8
2.1.2 ENVISION	8
2.1.3 LOD2	9
2.1.4 ViSTA-TV.....	9
2.2 The LT-Web Project.....	9
2.3 Other Collaborations.....	9
2.3.1 Development of a Wikipedia tool to track edit provenance	9
2.3.2 Automatically updating Wikipedia based on news sources and social media	10
3 Year 1 Follow-up.....	11
4 Conclusions and Future Work	12
References.....	13

Abbreviations

ENVISION	ENVironmental Services Infrastructure with ONtologies
ESWC	Extended Semantic Web Conference
UoW	University of Waterloo
LOD	Linked Open Data
ViSTA-TV	Video Stream Analytics for Viewers in the TV Industry
LT-Web	Language Technology in the Web

1 Introduction

During the second year of the RENDER project we started new collaborations with European projects and universities from the United States and Canada. We also continued the collaborations established in the first year with networks of excellence and several EU projects.

RENDER participated at the EU Project Networking session hosted by the Extended Semantic Web Conference (ESWC) 2012. The networking session provided the opportunity to initiate new collaboration activities with other EU projects: XLike, ENVISION, LOD2, ViSTA-TV. Section 2 of this deliverable is dedicated to describing each of the aforementioned projects, as well as the collaboration activity which will be carried out.

RENDER also initiated a collaboration with the LT-Web project, which will propose a W3C standard on textual enrichments. The terms of the collaboration are presented in Section 2.2.

Finally, RENDER defined two new common tasks with universities from the United States and Canada, on the development of a Wikipedia tool to track edit provenance and tools for automatically updating Wikipedia based on news sources and social media. We present the first steps towards defining common activities in Section 2.3.

The networks of excellence which RENDER collaborates with are Pascal, MetaNet and PlanetData, and the collaboration points have been described in the Year 1 Report on clustering and liaison activities. Pascal is a European Union funded network which aims at supporting collaboration between experts in Machine Learning, Statistics and Optimization. MetaNet aims at initiating a multilingual European information society, via a dynamic and influential community, and an open distributed facility for sharing and exchanging resources. PlanetData is focused on sharing multiple approaches to large-scale data management from different disciplines. In Section 3 of this deliverable we present a follow-up of these activities.

2 Collaboration Activities

This section describes the EU Project Networking session hosted by the ESWC 2012 conference and the participating projects that RENDER established collaboration points with: XLike, ENVISION, LOD2, ViSTA-TV. Section 2.2 provides more detail with respect to collaboration activities carried out with the LTWeb project, while Section 2.3 reports on other cooperation with universities from the US and Canada.

2.1 EU Project Networking

This year's Extended Semantic Web Conference (ESWC 2012) hosted the EU Project networking track [1]. In this track, partners in 12 European projects held presentations describing their projects. The purpose was to increase awareness within the research community, and to establish possible collaborations among various projects in an organized setting.

RENDER collaborated with the following participating projects: XLike, ENVISION, LOD2, ViSTA-TV and identified points of collaboration and, in some cases, agreed on the next steps. In what follows, we give a short description for each of the abovementioned projects, and highlight the collaboration points.

2.1.1 XLike

The XLike European project [2] aims at developing technology to monitor and aggregate mainstream and social-media knowledge and enable cross-lingual services for publishers, media monitoring and business intelligence. Research wise, the project proposes:

- to extract formal, structured knowledge from multilingual texts and augment it with cross-lingual knowledge bases, and
- to cope with the irregularities occurring in informal language used in social media by adapting linguistic techniques and taking advantage of crowdsourcing.

The project considers the English, German, Spanish, Chinese and Hindi as major world languages and Catalan and Slovenian as minority languages.

Within XLike, JSI is going to extend the Diversity services (Enrycher) in order to process multi-lingual textual information. Moreover, the NewsFeed services for collecting, indexing and querying news developed by JSI will be extended within XLike to cover more languages.

Therefore, via the XLike project we assure the availability of parts of the RENDER services, software and corpora beyond the duration of the RENDER project.

2.1.2 ENVISION

ENVISION stands for ENVironmental Services Infrastructure with Ontologies [3]. The project aims at offering support in the process of semantic discovery and composition of environmental services. This is achieved through a series of tools:

- the **Environmental Decision Portal** enables creating web-based applications for dynamic discovery and visual service chaining;
- the **Ontology Infrastructure** supports multilingual ontology management and visual semantic annotation tools;
- The **Execution Infrastructure** includes a semantic discovery catalogue and service mediator.

The common research topic shared by RENDER and ENVISION is in analysing the feasibility of generalized services for event detection from a generic medium. In the case of ENVISION, this medium is represented by sensors, while in the case of RENDER by text streams.

2.1.3 LOD2

LOD2 [4] is a large scale integrating project, having a number of goals such as improving coherence and quality of data published on the Web, improve the performance of RDF data management (and bring it closer to the one of relational data management), tackle the issue of trust on the Linked Data Web, enhance the experience of data publishers and users by lowering the entry barriers. Thus the project is developing:

- tools and methodologies for exposing and managing very large amounts of structured information on the Data Web,
- a testbed and bootstrap network of high-quality multi-domain, multi-lingual ontologies from sources such as Wikipedia and OpenStreetMap,
- algorithms for automatically interlinking data from the Web,
- standards and methods for detecting provenance, ensuring privacy and data security and quality,
- adaptive tools for searching, browsing, and authoring of Linked Data.

The collaboration between RENDER and LOD2 is in the lines of publishing Wikimedia and/or Google data generated within RENDER via the diversity enhancement tools in RDF in the LOD2 Cloud.

2.1.4 ViSTA-TV

ViSTA-TV (Video Stream Analytics for Viewers in the TV Industry) [5] is a research project which started in June 2012. The goal of the project is to create real-time TV recommendations for viewers. In addition, another outcome of the project will be a linked open dataset about TV, as well as highly accurate low-latency audience research.

ViSTA-TV can make use of the enrichment tool developed within RENDER, and we are going to further define the exact terms of the collaboration.

2.2 The LT-Web Project

The LT-Web European project [6] aims at establishing best practices and standards for linguistic content processing. The standards will define metadata about:

- the Web information that is relevant for language technology processing;
- localisation processes and content management workflows for creating Web content;
- language technology applications.

The LT-Web project will propose a W3C standard on textual enrichments, and we are proposing the Diversity Mining services (Enrycher) as a reference implementation.

2.3 Other Collaborations

2.3.1 Development of a Wikipedia tool to track edit provenance

Born out of the diversity discussion session at Wikipedia Academy 2012 in Berlin, which it hosted, KIT has initiated the collaboration with the University of Waterloo (UoW), Canada. PhD Rob Warren as a liaison at the UoW is working with KIT to create a tool to track the geographical provenance of edits to visualize the geographical and thereby cultural diversity of the content in an article. The partner is currently evaluating if it may acquire funding by Canadian programs to join RENDER as a self-funded external partner for the development of the aforementioned tool. If not, the collaboration will persist, if just at lower scale.

2.3.2 Automatically updating Wikipedia based on news sources and social media

During the Wikimania 2012 in Washington, D.C., USA, KIT has had talks with researchers Arkaitz Zubiaga and Taylor Cassidy from the City University of New York, USA in regard to how to automatically detect the need for updates for Wikipedia articles from Twitter and news sources. Start of a collaborative project is impending.

3 Year 1 Follow-up

In the first year of the project RENDER initiated the collaboration with three networks of excellence, namely Pascal, MetaNet and PlanetData.

Pascal [7] is a European Union funded network which aims at supporting collaboration between experts in Machine Learning, Statistics and Optimization. The Pascal Network of Excellence was chosen for the following application domains which are highly relevant for the RENDER project (more exactly the diversity mining related tasks of fact extraction and opinion mining): *Natural Language Processing, Information Retrieval, Textual Information Access*.

MetaNet [8] aims at initiating a multilingual European information society, via a dynamic and influential community, and an open distributed facility for sharing and exchanging resources. Language data (written corpora) and language processing tools developed within RENDER can be made available to a broader community via the MetaNet network.

PlanetData [9] is focused on sharing multiple approaches to large-scale data management from different disciplines. The solutions to large-scale data management developed within RENDER can be disseminated via the PlanetData network.

As the services, software and corpora developed within RENDER are being finalized to a greater extent at the end of the second year of the project, we can start disseminating the results via these three networks of excellence.

During the first year of the project RENDER also started collaboration activities with other EU projects: Living Knowledge, ALERT and CoSyne. The collaboration with LivingKnowledge was in the form of a jointly organized event, and we refer the reader to the Year 1 Report on clustering and liaison activities [10] for more details regarding the collaboration.

ALERT [11] is an Open Source project with a focus on the Open Source developers' collaborative environments, where it aims to improve the overall bug resolution process. The common task defined with ALERT was to test the feasibility of using an automatic text annotation tool to link bug resolution specific terms to OpenCyc ontological concepts [10]. The collaboration with the ALERT project was finalized, as the task was redefined and the requirement to link bug resolution specific terms to OpenCyc was dropped.

CoSyne [12] is a project aiming to automate the dynamic multilingual synchronization process of Wikis, which contain a combination of dynamic user-generated content and multilingual aspects. The collaboration with RENDER was defined in terms of sharing research results on fact coverage from a multilingual perspective. During the WikiSym in Linz, Wikimedia Germany met with Amit Bronner from the University of Amsterdam; with this occasion he presented the CoSyne demo system¹. Wikimedia Germany discussed the usage and possible ways to connect this system with Wikimedia's supporting tools developed in RENDER.

¹ http://prototype.cosyne.eu/demo/en/index.php/Main_Page, last accessed on 18.04.2013

4 Conclusions and Future Work

In this deliverable we presented a follow-up of the collaboration activities initiated during the first year of the RENDER project, as well as a number of new collaborations started during the second year.

In the third year of the project we are going to continue the collaboration established thus far with the aforementioned projects and report on the progress being made, as well as foster collaborations with other initiatives.

The main EU projects that JSI is going to continue collaborating with during the third year of the project are xLike and LT-Web. Two of the main RENDER tools – the Diversity Mining Toolkit and Newsfeed – are extended within the xLike project, in order to be able to process multi-lingual information. Additionally, we are proposing the Diversity Mining Toolkit as a reference implementation for textual enrichments in the LT-Web project. Both collaborations are very important as they allow RENDER technology to be further expanded and maintained after the finalization of the RENDER project. Additionally, JSI is going to continue disseminating RENDER results via two of the three networks of excellence: Pascal and PlanetData. This will be achieved by participating at Pascal events and using the Newsfeed web crawler in the PlanetData large-scale data management setting.

Wikimedia met with the CoSyne project team in October 2012 and attended CoSyne's Final presentation Day which took place in Amsterdam on the 1st of February 2013. Wikimedia would integrate CoSyne technology for analysing fact coverage and providing suggestions within the ASQM supporting tool in the third year of the project. This integration is conditioned on the CoSyne team finding a way to offer a version of their development which can handle live data from Wikipedia.

KIT will be advancing its collaboration with Luca de Alfaro from the University of California, Santa Cruz in developing a scalable, accurate detection of authorship in Wikipedia articles. A joined research paper is planned.

Some EU project and network of excellence collaborations (e.g. ENVISION, MetaNet) will not be further pursued because the respective projects will be finalized by the third year of RENDER.

References

- [1] EU Project Networking track at ESWC 2012: <http://2012.eswc-conferences.org/program/eu-project-networking>
- [2] XLike: <http://www.xlike.org/>
- [3] ENVISION: <http://www.envision-project.eu/>
- [4] LOD2: <http://lod2.eu>
- [5] ViSTA-TV: <http://vista-tv.eu/>
- [6] LT-Web at Cordis: http://cordis.europa.eu/fp7/ict/language-technologies/portfolio_en.html#L
- [7] Pascal: <http://www.pascal-network.org/>
- [8] MetaNet: <http://www.meta-net.eu/>
- [9] PlanetData: <http://www.planet-data.eu/>
- [10] D Rusu. RENDER Deliverable 6.3.1 - Report on Clustering and Liaison Activities (Year 1). 2011
- [11] ALERT: <http://www.alert-project.eu/>
- [12] CoSyne: <http://www.cosyne.eu>