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

RENDER

Deliverable D6.2.3

Impact and continuation plans for RENDER services and software

Editor:	Ioan Toma, UIBK
Author(s):	Ioan Toma, UIBK; Andreas Tlhammer, UIBK; Delia Rusu, JSI; Felix Leif Keppmann, KIT; Fabian Flöck, KIT; Mariana Damova, Ontotext; Angelika Adam, Wikimedia; Daniel Kinzler, Wikimedia; Enrique Alfonseca, Google
Deliverable Nature:	Report (R)
Dissemination Level: (Confidentiality)[1]	Public (PU)
Contractual Delivery Date:	M18
Actual Delivery Date:	M18
Suggested Readers:	General public, researchers, developers.
Version:	v0.1
Keywords:	RENDER software, RENDER services, impact, planning

[1] Please indicate the dissemination level using one of the following codes:
• **PU =** Public • **PP =** Restricted to other programme participants (including the Commission Services) • **RE =** Restricted to a group specified by the consortium (including the Commission Services) • **CO =** Confidential, only for members of the consortium (including the Commission Services) • **Restreint UE =** Classified with the classification level "Restreint UE" according to Commission Decision 2001/844 and amendments • **Confidentiel UE =** Classified with the mention of the classification level "Confidentiel UE" according to Commission Decision 2001/844 and amendments • **Secret UE =** Classified with the mention of the classification level "Secret UE" according to Commission Decision 2001/844 and amendments

Disclaimer

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:	Impact and continuation plans for RENDER services and software
Editor (Name, Affiliation)	Ioan Toma, UIBK
Work package Leader (Name, affiliation)	Andreas Thalhammer, UIBK
Estimation of PM spent on the deliverable:	11

Executive Summary

This deliverables reports on measures and plans that RENDER consortium members will implement in order to ensure that the services, software, corpora and ontologies developed within RENDER will remain available after the lifetime of the project. It also includes an analysis of the current state and strategic planning for the further growing of the development community around RENDER created open source software.

Table of Contents

Abbreviations

DCMI	Dublin Core Metadata Initiative
FOAF	Friend Of A Friend
HTTP	Hypertext Transfer Protocol
HTML	Hypertext Markup Language
KDO	Knowledge Diversity Ontology
LT-Web	Language Technology in the Web
NER	Named Entity Recognition
OWL	Web Ontology Language
REST	Representational State Transfer
RDF	Resource Description Framework
SIOC	Semantically-Interlinked Online Communities
URL	Uniform Resource Locator
XLike	Cross-lingual Knowledge Extraction
XML	Extensible Markup Language

1 Introduction

This deliverable reports on measures and plans in order to ensure that the services, software and corpora developed within RENDER will remain available after the lifetime of the project. For each artefact developed in the project we have derived specific plans and measures that are in the process of being implemented or will be implemented in order to ensure long term availability and impact of RENDER results. The services developed in RENDER, namely Enrycher (WP2) and the data layer infrastructure based on OWLIM (WP1) will be maintained and made available after the end of the project. They will run on the JSI and Ontotext servers. The services will be further developed in upcoming research projects.

RENDER is also developing a suite of software extensions to popular Web2.0 collaborative systems, such as Drupal, MediaWiki, Wikidata/Shortipedia, Twitter, etc. The software extensions developed in RENDER are all made available under open-source licenses and are hosted on public repositories. The extensions will be made available, used and further developed in following research project. This deliverable also describes the plans and measures for growing the development community around RENDER created open source software and ensures that corpora and ontologies developed in the project will be made available on long term.

The rest of this document is organized as follows. Section 2 describes the measures and plans for long term availability of RENDER services, software, corpora and ontologies. Section 3 discusses the measures and plans to grow the RENDER development community. Finally, Section 4 concludes the deliverable.

2 Measures and plans for long term availability of RENDER services, software, corpora and ontologies

This section describes concrete measures and plans that consortium members will implement in order to make sure that services, software, corpora and ontologies produced in the project will be available long after the end of the project.

2.1 Services

The RENDER infrastructure relies on a number of services to extract, process, store and manage diversity information. Two of the core services of the RENDER infrastructure are: (1) Enrycher, a service performing text analysis on Web articles, tweets, blogs or other text documents, and (2) OWLIM, a service that provides storage infrastructure for diversity data. This section discusses the plans and measures for long term availability of Enrycher and OWLIM services.

2.1.1 Enrycher

Enrycher is a service-oriented system developed by JSI, providing shallow as well as deep text processing functionality at the text document level. As such, it is part of the core text processing suite of services developed at JSI. We started its development before the RENDER project, having just a subset of the current functionality (named entity extraction and subject-predicate-object triplet extraction). During the lifetime of RENDER we are improving existing components as well as adding new functionality like entity resolution, word sense disambiguation, article template extraction, etc.

The Enrycher system will be further extended within the XLike[2] and LT-Web[3] European projects where JSI is a project partner, providing funding until the beginning of 2015.

The XLike European project 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. Within XLike, we are going to extend the Enrycher services in order to process multi-lingual textual information.

The LT-Web European project aims at establishing best practices and standards for linguistic content processing. Following this direction, we are working on Enrycher becoming a reference implementation of the upcoming W3C standard on textual enrichments (the standard will be developed with in LT-Web project).

We are also analysing the possibility of establishing potential spinoffs of the AILab department at JSI, having as underlying technology the Enrycher and news crawling systems.

2.1.2 OWLIM

Ontotext provides the service and data infrastructure layer that underlies RENDER services. The service infrastructure layer consists of an instance of the OWLIM semantic repository augmented with the capability of virtual or nested repositories. The OWLIM repository will contain the data layer of RENDER services. It will store:

- the interlinked datasets of FactForge, a segment of LOD cloud augmented with inferred facts, along with the Reference Knowledge Stack, the upper-level ontologies layer which provides an efficient access and navigation to the heterogeneous basic dataset of FactForge,

[2] http://www.xlike.org
[3] http://cordis.europa.eu/fp7/ict/language-technologies/project-lt-web_en.html

- the data coming from Enrycher text processing and opinion and sentiment mining tools, which will produce RDF from textual sources coming from social media like Twitter, news sites on the web, blogs, etc. based on KDO, PROTON and pre-determined topic models, such as Telefonica topic model, dmoz, and the like
- Google News cluster data, Telefonica data, Wikipedia data
- the data uploaded in instances of the nested repositories on top of FactForge by other users

This layer is integrated with Enrycher mining services, CLAS (CLustering and Spreading Activation)[4] toolkit, which provides a variety of methods for ranking and selecting RDF facts, and RENDER front-ends like Drupal, ensuring the availability of the content for RENDER services.

The RENDER service and data infrastructure will be maintained for one year after the end of the project. It will run on Ontotext servers, and will guarantee that RENDER project use cases will be up and running for interested parties, stakeholders and potential future clients. We will advertise the integrated technological solution within RENDER project to our existing clients, and seek additional use cases where this solution will be showcased and exploited during the one year this service will be maintained, and will aim at closing deals based on its added value, where Ontotext will continue to provide the service and data infrastructure layer for the holistic solution. The sustainability of RENDER project results will be ensured by the fact, that the data infrastructure offers a standard SPARQL endpoint.

The data infrastructure layer FactForge and the nested repositories will be packaged as DataAsAService and SoftwareAsAService offerings. FactForge along with the Reference Knowledge Stack will remain a free evaluation service and positioned in a way to gain large audience as a use case for large scale reasoning with linked data and efficient access and navigation to a LOD segment.

The capability of the nested repositories will be presented like an additional service to the already existing pubic service FactForge. Users will be allowed after subscribing to this service to upload their own data and have them interlinked with FactForge datasets. They will be able to access and use their derived datasets via the standard FactForge interfaces and other interfaces created within RENDER.

The CLAS toolkit will be maintained as services as well. It will be connectable to FactForge or other data sources and services. CLAS toolkit services will allow users to specify the parameters of their interests, and CLAS toolkit will provide them with diversity aware ranking of their queries. CLAS toolkit will be provided free of charge as an evaluation service as a feature of FactForge and the nested repositories.

The service and data layer infrastructure of RENDER will contain data from RENDER use case partners Google, Telefonica, Wikipedia. We will look for opportunities to have the service and data layer infrastructure in continued use by them.

2.2 Software

To make the results of the project accessible to a wide audience, RENDER is developing software extensions to popular collaborative systems. This section discusses the plans and measures RENDER partners are implementing in order to ensure that software developed in the project will be available for long time.

2.2.1 Semantic MediaWiki (SMW)

Semantic MediaWiki (SMW) is an extension for MediaWiki, providing semantic annotations that allow users to collaboratively enrich a wiki and collect data about entities. This data can be queried within the wiki or from external applications. SMW development started before the RENDER project at the institute AIFB at KIT and was also partly supported by past EU projects, in particular ACTIVE (Enabling the Knowledge

[4] http://www.larkc.eu/wp-content/uploads/2008/01/LarKC_D2.4.1-Spreading-activa tion-components-v1.pdf

Powered Enterprise), NeOn (Lifecycle Support for Networked Ontologies) and SEKT (Semantically Enabled Knowledge Technologies). During the RENDER project SMW is further maintained and expanded, e.g. development of a major release.
SMW is part of the MediaWiki project. Therefore the operational infrastructure is provided by the Wikimedia Foundation, in particular the source code repositories (SVN). Further the services of Sourceforge provide the download location for official releases and the mailing lists host. Along with an active open source developer and user community the long term availability of the SMW project, and thus of contributions within the RENDER project, is secured and independent from the lifetime of this project.

Semantic MediaWiki (SMW) is an extension for MediaWiki, providing semantic annotations that allow users to collaborative enrich a wiki and collect data about entities. This data can be queried within the wiki or from external applications. SMW development started before the RENDER project at the institute AIFB at KIT and was also partly supported by past EU projects, in particular ACTIVE (Enabling the Knowledge Powered Enterprise), NeOn (Lifecycle Support for Networked Ontologies) and SEKT (Semantically Enabled Knowledge Technologies). During the RENDER project SMW is further maintained and expanded, e.g. development of a major release.
SMW is part of the MediaWiki project. Therefore the operational infrastructure is provided by the Wikimedia Foundation, in particular the source code repositories (SVN) and the website hosting. Further the services of Sourceforge provide the download location for official releases and the mailing lists host. Along with an active open source developer and user community the long term availability of the SMW project, and thus of contributions within the RENDER project, is secured and independent from the lifetime of this project.

2.2.2 Wikidata/Shortipedia

The Shortipedia wiki is based on a Semantic MediaWiki and allows the collection of facts and sources for facts about anything. It was released as a proof-of-concepts during the lifetime of the RENDER project. Using the ideas, concepts and experiences from the Shortipedia development the requirements for a diversity-enabled semantic wiki have been created as well as a plan for integration with the Wikipedia architecture. These developments lead to the Wikidata proposal, at present a new Wikimedia project starting development in April 2012. Wikidata is proposed to provide a central knowledge in a machine- and human-readable and -editable form. It will be integrated in all language versions used in the Wikipedia and supply these wikis with centralized data. Furthermore the data is available for free and public use in other use cases.
Wikidata can be seen as the successor of Shortipedia. The initial project is funded and planed for one year till March 2013. Established as WikiMedia project and after successful integration of Wikidata and Wikipedia, the long-term development and availability of contribution within the RENDER project is guaranteed by the Wikimedia Foundation.

2.2.3 Wikipedia Use Case software

2.2.3.1 Analysis tools and dashboard on the Wikimedia Toolserver

We are developing analysis tools on the Wikimedia Toolserver[5], using data from the copy of the Wikipedia wiki databases provided there as well as from XML dumps of Wikipedia content and information retrieved via Wikipedia's web API. Results are made available via a web interface and RESTful[6] API also hosted on the Toolserver cluster.
These tools will be licensed under (L)GPL and/or BSD-style licenses and made available to the Toolserver developer community via a revision control system like Subversion or Git.

[5] http://toolserver.org/~RENDER
[6] http://en.wikipedia.org/w/index.php?title=Representational_state_transfer&oldid=477673620 accessed on 02.03.2012

The tools are part of what we call RENDER toolkit for the Wikipedia use-case. The first version of the toolkit will be announced via Wikipedia important mailing lists and by publishing a blog entry (http://blog.wikimedia.de) and on http://de.wikipedia.org/wiki/Wikipedia:Kurier.

2.2.3.2 Integration into Wikipedia using JavaScript

Results from these tools mentioned in Section 2.2.3.1 are integrated into the Wikipedia website using JavaScript. MediaWiki provides several mechanisms for integrating scripts; we expect to use site-wide scripting via common.js[7] as well as optional per-user scripts via the "Gadget" mechanism.
Such scripts are stored as wiki-pages on Wikipedia and are generally maintained by the community of Wikipedia administrators. They will be licensed under CC-BY-SA and possibly also under (L)GPL and/or BSD-style licenses.

2.2.3.3 MediaWiki Extensions

If appropriate, some functionality will be implemented as a server-side extension to the MediaWiki software. Such extensions (if deployed on Wikimedia sites) are maintained by the community of MediaWiki developers. Deployment of new extensions however has to pass a lengthy review process by the Wikimedia Foundation. Because of this, we will try to avoid the dependency on server-side extensions as far as possible.

2.2.4 Drupal extension

Drupal is a free and open source content management system (CMS) and content management framework (CMF) written in PHP and distributed under the GNU General Public License. Drupal has become very popular due to its modular design as well as easy to use and extensible functionality. One can easily extend Drupal by developing add-ons to its core functionalities, known as *contrib modules*. Starting with version 7.0, Drupal is also offering support for RDF as part of its core modules. In the RENDER project, Drupal was selected as one of the platforms for which diversity-aware extensions are developed. The reasons for selecting Drupal include plug-in extensibility and modular design, very large, active and open developer community as well as community's experience.

The Drupal extension is interacting with RENDER backend services. The current implementation is using Enrycher service for shallow as well as deep text processing of Drupal nodes content as well as OWLIM as a back end storage service for storing the diversity information extracted by Enrycher and formalized in RDF using light-weight ontologies such as Knowledge Discovery Ontology (KDO)[8]. As described in D4.1.1, we are planning to integrate other RENDER enabled services such as FactForge[9]. The general approach of integrating the Drupal extension with the RENDER backend services is briefly described below. This approach is currently implemented for Enrycher, and planned to be implemented for other RENDER services.

- Drupal nodes content is sent to the RENDER backend service
- The RENDER backend service processes the information and extracts diversity information related to the Drupal node.
- The RENDER backend service also interlinks the diversity information with in the linked open data cloud
- The diversity information in RDF is return to the Drupal extension

[7] http://de.wikipedia.org/w/index.php?title=MediaWiki_Diskussion:Common.js&oldid=100355920 accessed on 02.03.2012
[8] http://render-project.eu/resources/kdo/
[9] http://factforge.net/

- The Drupal extension processes the RDF and displays this information back to the user. Furthermore, the user is able to store the information in an OWLIM store.

The RENDER Drupal extension is made available under LGPL license and will be soon accessible at GitHub[10]. A demo of the first prototype of RENDER Drupal extension is available at: http://render-project.eu/drupal/. The prototype was developed by STI Innsbruck who plans to maintain and extend it in the future. Drupal , OWLIM and light-weight ontologies (e.g. DC, FOAF, etc.), used in the RENDER Drupal extension, are core technologies that STI Innsbruck is using to develop its multi-channel dissemination approach. The multi-channel dissemination approach is part of STI Innsbruck core development for long time, and as such STI Innsbruck is planning to use the mentioned technologies, as well as to maintain and extend the RENDER Drupal extension for longer time, beyond the lifetime of the project.

2.2.5 Twitter & MS Pivot Viewer

Another software artefact that is currently developed in RENDER aims to analyse Twitter feed data and to provide intuitive and meaningful visualization of this data using technologies such as MS Pivot viewer. The idea behind this software is two-fold. On the first hand, it will provide social media consumers with an innovative tool to visualize, analyse and consume the diversity of data in currently available social media streams, namely Twitter. On the other hand, it will provide the developer community with a closed and packaged component to enrich their tools and applications with the sentiment analysis, fact coverage and diversity extraction capabilities to be provided by the RENDER services.

Telefonica and STI Innsbruck are developing this software. We are planning to make available, maintain, develop this extension after the end of the project. All these actions are prerequisites for engaging and growing the community for Twitter & MS Pivot Viewer Render extension.

2.3 Corpora

This section details the plans and measures for long terms availability of various corpora developed in the project.

2.3.1 News and Social media Crawlers

The corpora provided by JSI comprises mainstream news and social media based on the Spinn3r crawler, as well as a news crawler developed by JSI, which is correlated to the Google News story clusters. The data is available in several languages, for e.g. English, German, Spanish.

The JSI news crawler is part of the core data collection tools developed at JSI. As such, it will be further extended beyond the RENDER project. Within the XLike project (coordinated by JSI) the crawler will cover English, German, Spanish, Chinese and Hindi as major world languages and Catalan and Slovenian as minority languages.

2.3.2 Wikipedia Historical Attributes Data

Wikipedia is the most popular online encyclopedia. It is now not only used as a textual encyclopedia for human consumers, but also as a knowledge base which can be processed with automated methods. Wikipedia is used for estimating similarity scores between tokens, as a repository of entities for disambiguation, and it has been successfully applied in improving Information Retrieval and knowledge extraction systems.

[10] https://github.com/

One of the most useful sources of data for automated processing are the infoboxes: small tables containing semi-structured data about the entity defined in the encyclopedic entry. Most of the fields in an infobox are (attribute, value) pairs, although they can also contain other kinds of information such as maps, images or related entities. Parsing infoboxes has yielded useful knowledge bases such as DBPedia, a very large collection of (entity, attribute, value) triplets [1].

Some attributes of entities change over time. For example, presidents of countries succeed each other, people may work for different companies, and marriages can be broken. By studying Wikipedia over time we can learn how some attribute values have changed by observing how the Wikipedia contributors update them. Numeric values, such as the number of employees of companies or the population of countries also change over time.

Time-dependent attributes have been used to extract temporal information with which to enrich taxonomies ([2]; [3]). Relations which change over time are the object of study of the temporal slot filling task of the Knowledge Base Population (KBP) competition, organized as part of the NIST-sponsored Text Analysis Conferences (TACs; http://www.nist.gov/tac/).

As a collaboration with the Universidad Nacional de Educacion a Distancia (UNED, Spain) we have extracted, from the complete edit history of the English Wikipedia (from its beginning up to February 2012) all the different (attribute, value) pairs which appered in infoboxes.

We believe that this will be a useful resource for researchers not only part of RENDER but of the research community as a whole for studying characteristics of relations and learning to extract temporally-anchored attribute values. The Wikipedia Historical Attributes Data corpora will be maintained, extended and made available by Google after the end of the project.

2.3.3 Corpex

Developed within the RENDER project by KIT Karlsruhe, the Wikipedia corpora explorer Corpex let's you swiftly browse through all the words of Wikipedia. Select your language, and when you start typing, the system shows you two statistics in four graphs:

- the ten most frequent words that start with the typed sequence of letters (as a barchart and a piechart), and
- the most frequent letter following the already typed sequence of letters (again, as a barchart and a piechart).

Additionally, the ten most frequent following words of any input word are visualized (as a barcharts and a piechart). This can be used for many applications where the occurrence of words in different language editions of Wikipedia is of use. An API is also provided for easy use of the data.

Corpex is currently available in the following languages: German (de), English (en), Spanish (es), French (fr), Hungarian (hr), Romanian (ro), Albanian (sq), Bulgarian (bg), Czech (cs), Italian (it), Swedish (sv), Serbian (sr), Croatian (hr), Serbo-Croatian (sh), Bosnian (bs), and simple English (simple). It is further available for the Brown Corpus (brown). Further languages are being prepared.

Corpex is still under development. The source code is fully open source, and all the data is also freely available. This way it is ensured that the corpora are available and reproducible even after KIT should decide to switch off the Website. Corpex is available at www.render-project.eu/tools/Corpex and will be further maintained and developed after the end of RENDER project by KIT.

2.4 Ontologies

Besides services, software and corpora, the RENDER project is creating means to express and formalize diversity knowledge as ontologies. This section outlines plans and measures to maintain and make further available the ontology developed in the project, namely the Knowledge Discovery Ontology.

2.4.1 Knowledge Discovery Ontology (KDO)

A prerequisite for the creation and consumption of machine readable diversity information is the existence clear, simple and efficient terminology that can be used to express diversity. The RENDER consortium has developed the Knowledge Diversity Ontology (KDO) [4], a light-weight ontology which serves as a central knowledge model for the RENDER infrastructure. KDO re-uses existing ontologies, such as Dublin Core Metadata Initiative (DCMI)[11] for metadata specification, the Friend Of A Friend (FOAF)[12] vocabulary to enable the representation of agent entities such as persons and organisations including the connections amongst them, the Semantically-Interlinked Online Communities (SIOC)[13] ontology enables the semantic representation of all entities connected with Web forums and related containers (e.g. Wikipedia articles, blog posts and others). Using KDO, one can describe knowledge diversity by instantiating classes such as kdo:Opinion, kdo:Sentiment, kdo:Polarity and using properties such as kdo:hasSentiment, kdo:hasOpinion, kdo:hasBias, etc.

Many of the components and tools developed in RENDER are able to handle (produce, consume, or store), diversity information that is represented using KDO. The alignment in terminology and knowledge model of RENDER components and tools, through KDO, has also influenced and lead to updates in KDO (see [5]).

The KDO ontology it is available online and dereferenceable at http://render-project.eu/resources/kdo/. A proper documentation including useful examples is also made available online. STI Innsbruck, as the main developer of KDO will ensure that KDO, its documentation remains available online. Any updates and support for users interested in the ontology will be also provided.

[11] DCMI - http://dublincore.org/
[12] FOAF - http://xmlns.com/foaf/spec/
[13] SIOC - http://sioc-project.org/

3 Measures and plans for growing the RENDER development community

In Deliverable D6.2.2 [6] we have reported about community building activities performed during the first year of the project. We have also detailed measures and plans to engage with the existing communities of the software tools for which we are building diversity-aware extensions. In this section we briefly revisit these measures and plans and we define new measures and plans for growing the RENDER development community.

3.1 Semantic MediaWiki (SMW)

Community building for the SMW has been done already in earlier phases, prior the lifecycle of the RENDER project. Centred on the SMW website and wiki respectively a team of open source developers evolved as well as an active user community. Further numerous extensions are published by third party developers, covering a broad range of topics.

For communication and support different channels between developers and users exist and foster the SMW community. First, the SMW website[14], a Semantic MediaWiki itself, is the central entrance point for new interested parties. It contains general information, documentation, FAQ and further information, e.g. about third party extensions. Second, for asynchronous but direct discussion and support a mailing list[15] has been created. Last, direct interaction between members of the community is enabled via an IRC channel. These communication and support channels are used and maintained by the SMW community. Professional commercial support is not part of the SMW project, but offered by several companies or individuals. Contact details for professional support are collected in the SMW wiki[16].

The Semantic MediaWiki Conference (SMWCon) is another building block of community building for SMW. It is a twice-yearly gathering for users, developers and interested persons and is held in North America, in spring, and in Europe, in fall. Several talks, tutorials and hands-on workshops on different topics promote new ideas, foster the community and at the end push the development of the SWM itself. Contributions to community building within the RENDER project are done in this context, supporting and sponsoring the SMWCon, e.g. with talks, or by participation on the discussion and support channels.

3.2 Wikidata/Shortipedia

The Wikidata project as successor of the Shortipedia proof-of-concept has not been started as the time of writing. First preliminary work will start in March 2012, the main project starts in April 2012 and lasts for at least one year. Therefore no detailed work on community building has been done. Currently wiki pages[17] provide general information, documentation and FAQ. A mailing list[18] provides a platform for discussions and information dissemination.

More communication channels, e.g. Blog and IRC channel, are planned and will be established after project start. In the future intensive work has to be done also on the attraction of Wikipedia editors, who will form the main editor group in Wikidata. Beside for example the development this editor community is one of the important build blocks of the Wikidata concept.

[14] http://semantic-mediawiki.org
[15] https://lists.sourceforge.net/lists/listinfo/semediawiki-user
[16] http://semantic-mediawiki.org/wiki/Professional_support
[17] http://meta.wikimedia.org/wiki/Wikidata
[18] https://lists.wikimedia.org/mailman/listinfo/wikidata-l

3.3 Wikipedia Use Case Study software

In the case of the Wikipedia use case study we have to handle and contact different developer groups which are concerning software development used for Wikipedia or its sister projects. Wikimedia Deutschland is organising an annual developer meet up – the Berlin Hackathon 2012[19] - which takes places in Berlin on June 1-3, 2012. We invite among others MediaWiki and extensions developers, MediaWiki systems administrators, Toolserver users, Gadgets makers, Bots maintainers, and template writers to collaborate and share ideas at this meeting. Additional to the main meeting days, there will be an invite-only meeting on Wikidata/RENDER on May 31, 2012.

3.3.1 Toolserver developers and bot owners

On the Wikimedia Toolserver[20], there exists a community of people writing web-based tools as well as command-line based bots to aid with administrative tasks on Wikipedia. We will make the tools we develop available to this community, both as a Web-API based service and as source code.
Currently we are using the infrastructures of the Wikimedia Toolserver for the development of our analysing respectively supporting tools. By announcing new versions of these developments and explicit request for testing and assessing, we will get in closer contact with interested developers using the Wikimedia Toolserver structure.
We will specifically engage this community at the Wikimedia Hackathon in Berlin in June. The goal is to make our tools known to the community, and gather feedback and insight.
In a later step, we are going to involve other Toolserver tools, which are analysing different aspects of diversity in Wikipedia in our supporting tools. Furthermore there are current developments, which will enable a better performance or better searching and analysis results, which will be useful and necessary for our future workflow. For example a current project initiative - the CatGraphApi[21] - will enable a very fast and efficient way to browse the category structure of a wiki.

3.3.2 Wikipedia administrators and power-users

On Wikipedia, there exists a community of power users and tech-savvy administrators who write and maintain JavaScript-based utilities for working with Wikipedia. Any JavaScript-based integration code we develop for integrating our tools into Wikipedia will have to be revised and deployed by this community.

Again, we plan to specifically engage this community at the Wikimedia Hackathon in Berlin in June. The goal is to make our tools and APIs known to the community, and gather feedback and insight.

3.3.3 MediaWiki Developers

In as far as we will develop server-side extensions for the MediaWiki software, and to the extent of these extensions are going to be deployed on the Wikimedia server farm, these extensions will be reviewed and maintained by the MediaWiki developer community and the technical staff of the Wikimedia Foundation.

However, because of the strict demands on code quality especially with regards to performance, and the slow deployment cycles, we will likely avoid this technique in the beginning. We will however use the opportunity of the Hackathon in Berlin to get in touch with this community as well, and gather insight and feedback.

[19] http://www.mediawiki.org/wiki/Berlin_Hackathon_2012
[20] http://toolserver.org
[21] https://toolserver.org/~daniel/CatGraph/help.html

3.4 Drupal extension

As described in D6.2.2 [6] we are implementing a 4-phase strategy to approach and work with the Drupal developer community. We revisit this strategy and describe additional measures and plans for growing the community around RENDER extension of Drupal. The strategy introduced in [6] includes the following phases:

- Phase 1: Give out general information, ask for input, discussion.
- Phase 2: Approach core groups, lay out development roadmap, ask for input, establish collaboration.
- Phase 3: Deepen collaboration, have community develop their own extensions and take leader roles, install "Knowledge Diversity in Drupal" initiative and present extensions at Drupal events and conferences.
- Phase 4: Pass leadership to community.

We successfully finalized phase 1. We have set up a sub-site of the RENDER website[22] explaining the planned extensions, giving up-to-date information about the latest advancements and the ways to take part in the project. A first prototype of the RENDER Drupal extension was also released. In phase 2, we will approach those groups from the Drupal community that are working on integrating semantic technologies into Drupal, namely the "Semantic Web" group[23] and the "RDF in Drupal 7 initiative"[24]. We will give the members full access to our current developments, incentivize them to make own developments and take their considerations, ideas and critique into serious account. We will also install mailing-lists and code-repositories (with prototype code) at this stage to coordinate the beginning cooperation and to answer questions and requests about our technology in time. In phase 3 we will deepen the collaboration with the community, create and install the initiative on using knowledge diversity into Drupal and showcase the RENDER Drupal extension(s) at Drupal conferences and events. Finally in phase 4 we will give the lead for the "Knowledge Diversity in Drupal" project gradually to the community, we will completely step out. Ideally, the project will be maintained and advanced by the Drupal community.

3.5 Twitter & MS Pivot Viewer

For the Twitter & MS Pivot Viewer software we are going to follow a similar plan for growing the community as described in Section 3.4.

We successfully finalized phase 1. We have set up a sub-site of the RENDER website[25] explaining the planned extensions, giving up-to-date information about the latest advancements and the ways to take part in the project. A first prototype of the RENDER Twitter & MS Pivot Viewer extension will be released at the end of year 2. In phase 2 we will target Microsoft Silverlight online communities for developers and end-users. Microsoft Silverlight is the framework which includes Microsoft Pivot Viewer. We will thus focus on explaining the possibilities of implementing this tool for exploring information diversity on any website embedding MS Silverlight. In phase 3 we will deepen the collaboration and encourage developers to expand the tool's capabilities to process other diversified data. Finally in phase 4 we will give the lead of the initiative to the community. Ideally, the project will be maintained and advanced by the Microsoft Silverlight online community.

[22] http://render-project.eu/render-extensions-and-tools-for-collaborative-web-platforms/render-extensions-drupal/
[23] http://semantic-drupal.com
[24] http://groups.drupal.org/semantic-web
[25] http://render-project.eu/render-extensions-and-tools-for-collaborative-web-platforms/render-extensions-twitter/

4 Conclusion

In this deliverable we reported on measures and plans for long term availability of services, software, corpora and ontologies developed within RENDER. This deliverable also includes an analysis of the current state and strategic planning for the further grow of the development community around RENDER created open source software.

For each service, software, corpora and ontology artefact produced in the project, the institutions that were driving the development, are also the main drivers for implementing long term availability and growing the development community around these artefacts. It is planned that the services, software, corpora and ontologies will be used and further developed by these institutions in upcoming research projects. There is a real interest of making the services, software, corpora and ontologies known and used in the relevant communities. A prerequisite for doing this is to maintain and make available in the future these services, software, corpora and ontologies, an objective that all RENDER partners commit to achieve.

References

[1] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives, "DBPedia: A nucleus for a Web of Open Data," in 6th Int'l Semantic Web Conference, Busan, Korea.

[2] Q. Zhang, F.M. Suchanek, L. Yue, G. Weikum, "TOB: Timely Ontologies for Business Relations," WebDB, 2008.

[3] Y. Wang, M. Zhu, L. Qu, and M. Spaniol, and G. Weikum,"Timely YAGO}: harvesting, querying, and visualizing temporal knowledge from Wikipedia", EDBT, 2010.

[4] A. Thalhammer, I. Toma, R. Hasan, "D3.1.1. Initial models for diversity-rich information", RENDER deliverable.

[5] A. Thalhammer, I. Toma, F. Flöck, "D3.1.2. Evaluation of models for diversity-rich information ", RENDER deliverable.

[6] F. Flöck, A. Adam, D. Vrandečić, "D6.2.2. Report on community building activities Y1", RENDER deliverable.

