RENDER

Deliverable D6.2.4

Report on community building activities Y2

Editor:	Fabian Flöck, Karlsruhe Institute of Technology (KIT)
Authors:	Fabian Flöck, KIT; Angelika Adam, Wikimedia Deutschland; Ioan Toma, STI
Deliverable nature:	Report (R)
Dissemination level: (Confidentiality)	Public (PU)
Contractual delivery date:	30/09/2012
Actual delivery date:	30/09/2012
Suggested readers:	RENDER reviewers
Version:	2
Total number of pages:	16
Keywords:	Deliverable, community building, external collaboration, collaboration, Social Web, Web 2.0, prototype, extensions, Drupal, MediaWiki, Shortipedia, Wikidata

Disclaimer

Full project title:	RENDER – Reflecting Knowledge Diversity
Short project title:	RENDER
Number and title of work package:	WP 6: Dissemination, exploitation, and community building
Document title:	D6.4.2 - Report on community building activities Y2
Editor (Name, affiliation)	Fabian Flöck, Karlsruhe Institute of Technology (KIT)
Work package leader (Name, affiliation)	Ioan Toma, STI Innsbruck (STI)
Estimation of PM spent on the deliverable:	2

Copyright notice

Executive summary

This deliverable reports on the activities regarding community building for the diversity toolkit and Web 2.0 platforms in the second project year. Activities started in the first project year have been continued and expanded; others have been launched during the second year. Section 2 covers extensions to Web 2.0 platforms, with subsection 2.1 describing the efforts made in regards to the Wikipedia community. Three communities are distinguished: The Wikipedia/Wikimedia community, which consists of Wikipedia users, the Wikimedia developer community and the Wikipedia/Wiki researcher community. Each of them are approached aligned to their specific characteristics. For the users, blog entries, community portals on the Wikimedia sites are used to establish a readership and encourage low-threshold participation by the readers. The RENDER toolkit on the Wikimedia Toolserver is used as showcase, together with several mailing lists, to attract users to try out the tools and put them to productive use. We took the opportunity to connect with the community directly and build a group of interested editors at the largest gathering of Wikipedians, the Wikimania 2012 in Washington D.C.. Similar efforts are made at the WikiConvention, a European gathering of similar type. In order to encourage developers to get actively involved in expanding the RENDER toolkit, communication was established via mailing lists, the RENDER/Wikidata Summit in Berlin, and the Berlin Hackathon. Researchers were approached not only with paper and poster presentations at the Web Science 2012, the Wikimania 2012, Wikipedia Academy 2012, Hypertext and Social Media 2012, and WikiSym 2012 conferences, but also through discussion rounds and break out sessions organized by the RENDER partners as well as speed geeking sessions, information booths and dissemination materials, eventually leading to the initiation of at least two promising collaboration projects.

Section 2.2 covers the TwiDiViz application, which finds its community of users primarily in the Telefónica use case with a strong interest in the tool. As the development of the tool in combination with the dashboard from the Telefónica use case just reached a stable state, the community will now be approached. The community of Microsoft Silverlight as an underlying technology will also be approached, as for Silverlight developers with any interest in data visualization techniques or opinion mining, sentiment analysis and such, it will be a very attractive candidate for reuse, customization and enhancement.

For Drupal, the focus of the second year was on phases 1 and 2 (general information, approach core groups of the Drupal community) of the plan presented in D6.2.2. Several discussion sessions with representatives of the Drupal community in Austria took place by email and also face to face at the Open Government Data Businessday 2012[1] in Vienna, March 22nd and at Open Source Day 2012 in Innsbruck, May 11th. At the Open Source Day 2012 we presented the extension. As the development has advanced, the extension and results will be further discussed and advanced together with the online communities concerned with RDF in Drupal and presented at a major Drupal event (e.g. Drupalcon or DoitwithDrupal), to engage with more communities that are involved in the Drupal development, or in the development of extensions for Drupal, finally to pass the leadership of the diversity aware extension to the community.

Section 3 finally describes what has been done in regards to establishing a user and supporter base for the KDO ontology, and lists actual and potential use cases for the semantic technologies developed in RENDER as a part of OWLIM and FactForge.

---

[1] http://ogdb.eventbrite.com/

List of authors

Organisation	Author
KIT	Fabian Flöck
Wikimedia	Angelika Adam
STI	Ioan Toma

Table of contents

Abbreviations

CMF   Content management framework

CMS   Content management system

ESWC   Extended Semantic Web Conference

GPL   General Public License

KIT   Karlsruhe Institute of Technology

LOD   Linked Open Data

MS   Microsoft

RDF   Resource Description Framework

UI   User Interface

URI   Unique Resource Identifier

WP   Work Package

SMW   SemanticMediaWiki

1 Introduction

This deliverable reports on the activities regarding community building for the diversity toolkit and Web 2.0 platforms in the second project year. Activities started in the first project year have been continued and expanded, other have been launched during the second year. The subsections of the tools section 2 describe the efforts made in regards to the Wikipedia community, the Drupal community and the community for our TwiDiViz tool, while section 3 describes what has been done in regards to establishing user and supporter base for the KDO ontology, OWLIM and FactForge.

2 Extensions to Web 2.0 Platforms

2.1   Wikipedia Diversity Toolkit

For the Wikipedia use case we distinguish different communities. These are the Wikipedia/Wikimedia community, which consists of Wikipedia users, the Wikimedia developer community and the Wikipedia/Wiki researcher community. For all of these communities we used different kinds of information channels and events to inform and to increase the interest for the RENDER project.

2.1.1   Wikipedia/Wikimedia community

We used several publishing channels to keep and increase the interest and enthusiasm of the Wikipedia and Wikimedia community for the project and to support our work by highlighting the project objectives and development results:

Wikimedia blog entries:

At the end of each month, Wikimedia Deutschland is publishing a blog entry. These entries contain a brief presentation of the activities in all associated projects. The current tasks and the working progress of the RENDER project were also introduced in this context. These entries are online available under: http://blog.wikimedia.de/monatsberichte/

Additionally, we published several blog entries related to RENDER events:

- A blog posting about the Review Meeting in
  Luxemburg: http://blog.wikimedia.de/2011/12/08/erstes-render-review-meeting-in-luxemburg/

- An announcement of the RENDER toolkit http://blog.wikimedia.de/2012/03/27/analyse-der-diversitat-in-wikipedia-das-render-toolkit/

- About the RENDER plenary in Madrid: http://blog.wikimedia.de/2012/05/11/render-projekt-treffen-in-madrid/

- A report about the Berlin Hackathon 2012: http://blog.wikimedia.de/2012/06/11/rueckblick-berlin-hackathon-2012/

- Concerning our new team member: http://blog.wikimedia.de/2012/06/26/johannes-kroll-verstarkt-das-render-team/

- A blog entry about the Wikipedia Historical Article Data set which can be downloaded from the RENDER toolkit download page: http://blog.wikimedia.de/2012/06/28/google-stellt-historische-infobox-daten-zur-verfugung/

- A report about RENDER at the WebScience 2012 Conference:
  http://blog.wikimedia.de/2012/07/03/render-bei-der-websci-2012/

Community information websites:

During the first year, we established an information page on Meta Wiki[2], as we reported in D 6.2.2. In the second year we created additionally an information webpage in the German Wikipedia[3] to inform in

---

[2] Meta Wiki is the global community site for the Wikimedia Foundation's projects.
http://meta.wikimedia.org/wiki/RENDER
[3] http://de.wikipedia.org/wiki/Wikipedia:RENDER

particular the community supporting Wikimedia Deutschland about the project in general, the plans, events and the current development status.

• **The RENDER toolkit on the Wikimedia Toolserver**

We established the RENDER Toolkit on the Wikimedia Toolserver[4], which contains a collection of tools developed during the RENDER project and will be improved and extended as development continues. As such it provides a central access point to current and future results, data sets, which are extracted or calculated and offers interested users and researchers a participation channel for testing, assessing and commenting on these tools. The toolkit can be seen as a showcase to give insights as to which approaches to analyse diversity in Wikipedia we are taking into account. It is thereby an important display to demonstrate the results of RENDER to the community and a possibility for them to engage with the project as all users have been and are invited to contribute to the toolkit. It therefore serves as an important catalyser for the community building.

Mailing lists: We used several mailing lists (WikiDE-list, VereinDE-list) and the German Wikipedia news portal Wikipedia:Kurier to inform the Wikipedia and Wikimedia community about relevant events and publications concerning the RENDER.

Additionally, we presented and discussed the project at several **conferences and community gatherings**:

• **Wikimania**[5] is the annual international conference of the Wikimedia community. In 2012, it took place from July 12th to 15th in Washington, D.C., USA. There, we presented the current status of the project to the most active editor community and discussed with community members about functionality and requirements of the supporting tools as well as best practices for implementation and usage. The presentation was published on YouTube. We used the opportunity for exchanging and discussing with interested Wikipedians and different researchers during speed dating sessions and a meet up we organized, focussing on the diversity challenges tackled by RENDER. All interactions with the editor base showed high engagement with the topic, which was anyway a high priority topic for the Wikimedia foundation at this year's Wikimedia, a fact that channelled additional editor interest to RENDER and its endeavours regarding Wikipedia. Researcher present also showed heightened interest in the topic, leading to the initiation of some promising collaboration with members of KIT (for details see D6.3.2).

• **WikiConvention**[6] is the annual community conference of the German Wikipedia and her sister projects. In 2012, this conference took place from 31st August to 2nd September in Dornbirn, Austria. There we presented the prototypes of the supporting tools during a workshop and discussed further requirements and needs for the users.

Wikidata has been involved in a number of events addressing both developer and user community. A complete list is available online.[7] In the M36 version of the deliverable we will elaborate on the impact of these community building measures together with an assessment of the social media and other channels we used to inform about the progress of the project and its results.

2.1.2   Wikimedia Developer Community

Many Wikimedia developers use the Wikimedia Toolserver[8] as infrastructure for their development related to Wikimedia projects. We established communication with these developers via the Toolserver mailing list and motivated this community to acquire deeper insights into the RENDER project. We as well discussed and presented the analysis approaches with help of the RENDER Toolkit. We collected feedback on the analysis tools and further suggestions to extend the analysis set of the supporting tools.

Additionally, we approached the developer community at the following events:

---

[4] http://toolserver.org/~RENDER/toolkit/
[5] http://wikimania2012.wikimedia.org/
[6] http://de.wikipedia.org/wiki/Wikipedia:WikiCon_2012
[7] http://meta.wikimedia.org/wiki/Wikidata/Events
[8] http://toolserver.org/

- The RENDER/WikiData Summit[9] took place in Berlin at the end of May 2012. There we presented and discussed the current development status and the further ideas with developers from the Wikimedia Foundation.

- The Berlin Hackathon 2012[10] is a meeting of Wikimedia community developers and system administrators and was organised by Wikimedia Deutschland and the Wikimedia Foundation. There we had discussions with interested developers about our case study and worked together with some volunteers on analyzing tools, in particular on the approach to identify political bias within German Wikipedia articles.

2.1.3 Wikipedia Researcher Community

We used the WikiResearch mailing list to inform researchers about the toolkit or new highlights within the RENDER project.

We presented and discussed the project in the context with several conferences:

- During the WebScience2012[11] we presented the RENDER toolkit to researchers and discussed further ideas and improvements.

- The Wikipedia Academy 2012[12] conference was organised by Wikimedia Deutschland, FU und HU. There we presented and discussed the project with researchers during a speed geeking session and in the context of this event. A discussion session on the topic of knowledge diversity in Wikipedia was hosted by KIT and attended by some 15 researchers and editors. The discussion yielded very productive insights and also let to a cooperation between KIT and one of the participants, Rob Warren from the University of Waterloo, Canada (See D6.3.2 for details).

- WikiSym2012[13], the international symposium on wikis and collaboration, took place in Linz (Austria). There we shared the ideas of RENDER and exchanged about the usage of further research results which are directly related to diversity aspect or techniques for the Wikipedia use case.

For the last project year we plan to continue the community building activities in particular by announcing and spreading the supporting tools to get feedback and suggestions from users and developers.

2.2   TwiDiViz

TwiDiViz is a web-based application, developed by STI Innsbruck, which enables the analysis and visualization of diversity in Twitter data. This tool uses core RENDER technologies (such as the Knowledge Diversity Ontology (KDO) ontology and the Enrycher service) to process Twitter datasets for the purpose of analyzing the impact of products on basis of sentiment and topic mining. KDO[14] aims at providing a vocabulary that describes different dimensions of knowledge diversity of the Web, providing the knowledge exchange format. Enrycher[15] is a service-oriented system, providing shallow as well as deep text processing functionality at the text document level.

Using TwiDiViz, users can explore and get familiar with diversity aspects of the Twitter data in two different ways. We offer a static printable report summarizing the results of our diversity-oriented analysis. In addition, we provide an interactive visualizer based on Microsoft Pivot Viewer.

Within the RENDER project, TwiDiViz is used to support decision makers at Telefónica, by providing a consistent view on the social media data produced by Telefónica customers and published via Twitter. TwiDiViz has been already integrated into the Telefónica dashboard (see D4.1.2 and D5.3.4). The tool can

---

[9] http://meta.wikimedia.org/wiki/RENDER/Berlin_summit_2012/
[10] http://www.mediawiki.org/wiki/Berlin_Hackathon_2012
[11] http://www.websci12.org/
[12] http://wikipedia-academy.de/2012/wiki/Main_Page
[13] http://www.wikisym.org/ws2012
[14] http://render-project.eu/resources/kdo/
[15] http://enrycher.ijs.si/

be equally applied to any social media monitoring scenario described in terms of a list of topics of interest, language.

TwiDiViz was developed during the 2nd year of RENDER. Our focus during this period was on the development of the tool functionality and less on building a community of developers and end-users for the tool. We are planning to focus on community building aspects during the 3rd year of the project. As TwiDiViz is integrated into the Telefónica dashboard, we have from start access to a rather small but interested community of end-users (i.e. Telefónica decision makers). We plan to extend the end-user community by reaching out to end-user Silverlight online communities, which we are monitoring and interacting with. We will also focus on creating and growing a developer community. We will encourage developers to expand the tool's capabilities to process other diversified data. To our knowledge, this will be the first free, customizable tool available that combines the elaborated data visualization and browsing techniques offered by MS PivotViewer with the ability to directly consume online data from RENDER's unique diversity-mining techniques and can be integrated easily into websites. Thus, for Silverlight developers with any interest in data visualization techniques or opinion mining, sentiment analysis and such, it will be a very attractive candidate for reuse, customization and enhancement.[16] As it will be also an integrated part of the Telefónica internal Twitter analysis tool and will be used by a community of Telefónica business decision makers, especially in marketing. These will help to further refine the tool and TwiDiViz.

2.3   Drupal

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. It is used as a back-end system for a large number of websites, blogs and forums. It is also used for knowledge management and business collaboration.

The diversity aware extension for Drupal, built by STI Innsbruck, is a Drupal module that provides a diversity-aware view on posts provided with Drupal. It is able to show extracted topics and named entities and provides links for these. The extension uses core RENDER technologies, namely Enrycher for text processing of Drupal nodes[17] and OWLIM[18] as a backend storage service for storing the diversity information extracted by Enrycher and formalized in RDF using the Knowledge Discovery Ontology (KDO).

Drupal is a very popular platform with a very large end-user and developer community. During the last year the user community has increased by 50%, nowadays counting more that 900,000 user accounts, while the developer community has doubled in last year, counting over 20,000 developers[19]. More the 730,000 sites are currently running Drupal and there are more that 3,000 commits each week. We are targeting both the very large end-user community as well as the fast growing and active developer community. The semi-annual Drupal Conference DrupalCon hosted over 3,000 attendants in March 2011. There is a wide variety of online community channels we can and will use.

In the previous version of this deliverable (i.e., D6.2.2) we have introduced a four phase plan to approach and work with the Drupal community. The four phases are:

- Phase1: Give out general information, ask for input, discussion

- Phase2: Approach core groups, lay out development roadmap, ask for input, establish collaboration

- Phase3: Deepen collaboration, have community develop their own extensions and take leader roles; Install "Knowledge Diversity in Drupal" initiative, broaden scope; Present extensions and results at Drupal events.

- Phase4: Pass leadership to community

---

[16] Target communities are for example http://www.silverlight.net/community, http://www.silverlightshow.net/, and http://www.microsoft.com/SilverLight/community/

[17] "All content on a Drupal website is stored and treated as "nodes". A node is any posting, such as a page, poll, article, forum topic, or blog entry." (see http://drupal.org/documentation/modules/node)

[18] http://www.ontotext.com/owlim

[19] http://www.drupal.org

During the 2nd year of the project we focused on Phase1 and Phase2 of the plan. We gave out general information about the activities related to the development of diversity-aware extension and we asked for feedback. We approached core groups of the Drupal community and we had several talks with representatives of the Drupal community in Austria by email and also face to face at the Open Government Data Businessday 2012[20] in Vienna, March 22nd and at Open Source Day 2012 in Innsbruck, May 11th. At the Open Source Day 2012 we presented the diversity aware extension for Drupal with a talk entitled "Vielfältige Informationen für Posts in Drupal" by Simon Hangl (STI).

For the last year of the project, the community building process will move to Phase3 and 4. We aim to present the extension and the results at major Drupal event (e.g. Drupalcon or DoitwithDrupal), to engage with more communities that are involved in the Drupal development, or in the development of extensions for Drupal, finally to pass the leadership of the diversity aware extension to the community.

---

[20] http://ogdb.eventbrite.com/

3 Diversity-enabling technologies and standards

3.1   The Knowledge Diversity Ontology (KDO)

The Knowledge Diversity Ontology, shortly KDO, is a light-weight ontology for representing diversity-related information on the Web. It includes concepts and properties for representing and linking opinions, viewpoints, sentiments, emotions, and biases. KDO heavily uses terminology provided by other well known existing vocabularies such as SIOC and FOAF. The ontology is currently used by core services in RENDER, which are able to process, generate, store and query data represented in KDO. The ontology is available online[21] and is also listed as a Linked Open Data vocabulary[22]. The ontology has been advertised at top scientific conferences (e.g. ESCW and WWW) by means of posters and workshop papers. We have additionally used Web2.0 channels (Facebook and Twitter posts) to advertise and spread the word about KDO. It was also presented at the poster and demo session at ISWC 2011 in Bonn. A substantially sized audience was interested in the way the ontology is applied and what kind of inferences can be made. Additionally, at the EU project networking session of ESWC 2012 the ontology was presented as part of the RENDER project. Interested conference participants asked questions about the application and for further reference. We also advertised KDO at the vocabulary index LOV.[23] As a consequence of our efforts, we encountered significant a uptake by the European project ALICE (Conesa & al., 2012). Inquiries and feedback about KDO have been made to STI on a regular basis.

3.2   The OWLIM repositories

OWLIM is the most scalable semantic repository in the world. It comes in three flavours, OWLIM-Lite, OWLIM-SE, OWLIM-Enterprise. OWLIM 5.0 has been released in 2012, being partially funded by the RENDER project, and employed right away for several high profile commercial projects, in which OWLIM is the underlying infrastructure for data and diversity intense solutions. OWLIM 5.0 is a pure Java implementation, compatible with Sesame 2, and JENA. It supports customizable reasoning, patent pending optimized owl:sameAs handling, clustering, geo-spatial extensions, full-text search, high performance retraction, RDF rank and priming and a notification mechanism, allowing reaction in the update stream. This OWLIM version has been the repository underlying the BBC Olympics 2012 website.[24] It powered the automatic generation of 800 websites at once, and a load of 15 million visitors per day. OWLIM was chosen after a thorough evaluation of high end semantic repositories as the semantic repository of FAO[25], the Food and Agriculture Organization of the United Nations, for a large scale data management solution comprising more than 300 various technologies. OWLIM was also selected as the technological infrastructure basis for RJ Lee Group, a developer of information management systems for law enforcement, to use it in their solutions for forensic investigations[26]. This testifies for the ability of OWLIM to support large amounts diverse data and diversity intense solutions, as developed by the RENDER project. Further, the Dynamic Semantic Publishing platform adopted by BBC Olympics and BBC Sports, using OWLIM 5.0 has also been employed by Press Association, the second largest media company in the UK, by TSO (The Stationary Office) UK, an organization that helps people to create, manage and publish their information, and by ParliamentaryICT, ensuring the smooth running of the UK parliament, KoreaTelecom of South Korea, etc. OWLIM has been regularly featured in the most important Semantic Web conferences like SemTech US, SemTech UK, where Ontotext participates as a Platinum sponsor. OWLIM is also at the core of the data infrastructure of several EU FP7 funded projects, such as MOLTO, RENDER, AnnoMarket, Europeana

---

[21] http://render-project.eu/resources/kdo/

[22] http://lov.okfn.org/dataset/lov/details/vocabulary_kdo.html

[23] http://lov.okfn.org/dataset/lov/

[24] http://www.bbc.co.uk/blogs/bbcinternet/2010/07/the_world_cup_and_a_call_to_ac.html and
    http://www.bbc.co.uk/blogs/bbcinternet/2010/07/bbc_world_cup_2010_dynamic_sem.html

[25] http://comingsoon.data.fao.org/

[26] http://www.ontotext.com/news/rj_lee_group_and_ontotext_sign_oem_deal

Creative, Bulgariana, etc. OWLIM is the underlying infrastructure of two data services LinkedLifeData (http://linkedlifedata.com) and FactForge (http://factforge.net), which is co-developed within the RENDER project, that have been visited by more than 1 million visitors per month.

3.3   FactForge

FactForge (http://factforge.net) is a free data service developed and maintained at Ontotext. Its development is partially funded by RENDER. It is in the core of the RENDER data layer, providing the data pool of generic knowledge enriching the entities recognized in texts with additional knowledge. FactForge is the largest body of heterogeneous common sense knowledge of linked data on which inference is performed. It is a compound dataset of 9 of the most popular datasets of the LOD cloud, including DBpedia, Freebase, Geonames, Wordnet, etc. FactForge is visited 5000 times monthly from users around the world. It has been used as the knowledge or data infrastructure in several FP7 projects, like MOLTO and RENDER, and regularly serves as a data source for semantic annotation tasks of Ontotext commercial projects related to media and cultural heritage such as Open Buildings, Israel; Publicis, Siemens AD, Germany; NDP, the Netherlands, top 5 media company in the US, the Semantic Knowledge Base of TNA (The National Archive of Great Britain), ResearchSpace of the British Museum, UK etc. providing the proof of the viability of the technology and the scientific insight of RENDER project.

4 Conclusions

The various efforts to build communities for the different tools can be deemed successful for the cases of the extensions that are already in a developmental state suitable for approaching an external community audience with them. Firstly, for Wikipedia, where an active user and developer base took up the RENDER toolkit and the underlying idea quickly and with great enthusiasm, as it addresses the problem of (systematic) bias lingering for quite some time now in the space of big, unsolved issues of the World's biggest Encyclopedia. Secondly for Drupal, where first encounters with the active user base show very promising collaboration potential, which will be nurtured in the following project months by engaging more closely with the core online and offline Drupal groups that can be the drivers for a successful advancement of the extension. TwiDiViz will be first introduced to its main user community in the coming project months and revised with the following feedback before we will make it available to the wider Silverlight community. KDO, OWLIM and FactForge already showed to have their user communities, which we will continue to approach to support a high profile and resulting usage of these diversity-enabling technologies.

References

[1] Conesa, J., & al. (4-6. July 2012). Towards the Representation of Emotional Information from On-Line Collaborative Learning Sessions. *Sixth International Conference on Complex, Intelligent and Software Intensive Systems (CISIS)* , S. 923-928.