D1.1: DETAILED OVERALL MANAGEMENT AND BODIES MANAGEMENT, INCLUDING THE QUALITY ASSURANCE PLAN

Felix Sasaki

Distribution: Consortium Internal Report

MultilingualWeb-LT (LT-Web)
Language Technology in the Web

FP7-ICT-2011-7

Project no: 287815

Document Information

Deliverable number:	1.1
Deliverable title:	Detailed Overall Management and Bodies Management, including the Quality Assurance Plan
Dissemination level:	CO
Contractual date of delivery:	31st December 2012
Actual date of delivery:	19th February 2013
Author(s):	Felix Sasaki
Participants:	DFKI
Internal Reviewer:	TCD
Workpackage:	WP1
Task Responsible:	Felix Sasaki
Workpackage Leader:	Felix Sasaki

Revision History

Revision	Date	Author	Organization	Description
1	13/02/2013	Felix Sasaki	DFKI	Initial Draft
2	15/02/2013	David Filip	UL	Revised Version

CONTENTS

1. Executive Summary

This report summarizes the overall management of the MultilingualWeb-LT (LT-Web) project. The challenge of setting up the management was the three-fold nature of LT-Web: It continues the MultilingualWeb brand, it is a distinct EC project with a focus on creating reference implementations, and it is a W3C Working Group standardizing metadata for the Multilingual Web.

A reliable management of the project was set up on the basis of the W3C communication infrastructure, see LT-Web deliverable D1.2.1. for details. Having a sustainable, Web based record of meetings, decisions, plans and time line is an essential foundation. This included a clear separation of communications meant for the Multilingual Web workshop program committee, the EU project partners, and the W3C Working Group.

This report describes how the three groups of people in forehand mentioned areas interplay, how co-ordination W3C internally and with external contributors takes place, and how the quality of the overall work progress is assured.

2. Introduction

The LT-Web project conducts its work involving four areas of input:

1. The DoW as laid out in the grant agreement for LT-Web;
2. A program committee involving participants from the LT-Web project, its ancestor MultilingualWeb and additional program committee members. The main task of the program committee is the planning of the MultilingualWeb workshops. The members for the upcoming workshop 12-13 March 2013 are listed at http://www.multilingualweb.eu/rome#pc .
3. Implementers providing metadata reference implementations. Some of these are funded by the LT-Web project. Some of them provide implementations as sub-contractors or on a voluntary basis. All implementers are participating in the test suite effort and are listed in the test suite dashboard, see https://github.com/finnle/ITS-2.0-Testsuite/ and an HTML preview at http://tinyurl.com/its20-testsuite-dashboard .
4. A W3C working group charter for the MultilingualWeb-LT working group, defining the standardization process undertaken in MultilingualWeb-LT (LT-Web); see the online version at http://www.w3.org/2012/09/mlw-lt-charter . The charter is part of this report and available in the same directory as the file "MLW-LT_Working-Group_Charter.pdf". The W3C working group participants are listed at https://www.w3.org/2000/09/dbwg/details?group=53116&public=1

The three areas 2) - 4) are constituted by different groups of people. This report describes how the interplay between these groups is organised to assure the quality of the overall work progress. A key means here is the usage of Web based communication tools; see D1.2.1 "Report on Internal and External Communication Tools" for details.

3. The DoW and the three Constituencies

3.1. Overview

Figure 1 summarizes the relation between the DoW and the three groups of people or "constituencies" involved in LT-Web.

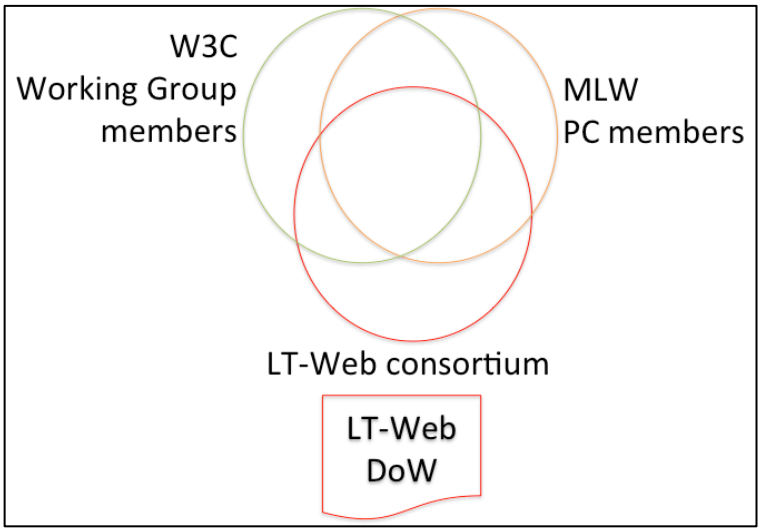

Figure 1: Relation between LT-Web DoW, the LT-Web consortium, W3C working group and MultilingualWeb

PC members.

The DoW describes tasks to be undertaken by the LT-Web consortium. Many of these tasks are related to reference implementations (WP3-5). A strong relation holds to the W3C working group. The working group is creating the standard Internationalization Tag Set (ITS) 2.0, which demonstrates use cases of the reference implementations.

The working group encompasses also organizations that are not directly funded by LT-Web. These have joined the working group as W3C member organizations. At the time of writing these additional working group members are: Adobe, CNR, DERI, Ecole Mohammadia d'Ingenieurs Rabat, NCSR, Logrus, Opera, SAP, and two individuals ("invited experts")[1]. Adobe and Logrus are also participating in implementation efforts (see section 6 of this report), like Shaun McCane ("invited expert").

The program committee continues planning of the MultilingualWeb workshops. In other words, it is the public face of the MultilingualWeb brand created by the ancestor project, the thematic network MultilingualWeb.

3.2. Governance Structure

The governance relations between the three groups described above are as follows.

The LT-Web consortium has created the W3C working group charter; see the following section for details. The charter defines work items to be undertaken in the W3C working group. The input from implementers gathered in the LT-Web consortium assures that standardization is not for self purpose, but driven by implementation and use case scenarios defined in the DoW. The working group is chaired by David Filip (LT-Web partner UL), Dave Lewis (LT-Web partner TCD) and Felix Sasaki (LT-Web coordinator DFKI / W3C Fellow).

Before the W3C working group started in March 2013, the charter had to undergo a review by the W3C membership. This process step is mandatory to assure broad consensus in the Web community.

In September 2012, the charter had to be re-approved and slightly modified to assure compatibility with HTML5. All these steps were governed by the W3C process[2] and not by the DoW. The double-role of the LT-Web coordinator as DFKI employee and W3C staff assured that this change did not influence the work plan outlined in the DoW.

The LT-Web participants use the MultilingualWeb workshop series as a public forum to discuss standardization issues or to showcase reference implementations. But they do not drive the program committee of the MultilingualWeb workshop. The MultilingualWeb brand and the workshops mostly became successful because of their *neutrality*. Hence, LT-Web only fulfils a small role in the actual PC work, e.g. in discussions about speaker proposals, session planning and the overall workshop structure. This helps to assure the neutral continuation of the MultilingualWeb brand.

4. The W3C Working Group Charter and Standardization Time Line

The working group charter is part of this deliverable and attached as a separate HTML file "MultilingualWeb-LT Working Group.html". In this section main passages will be analysed.

[1] See http://www.w3.org/2004/08/invexp for more information.
[2] See http://www.w3.org/Consortium/Process/ for more information.

The "Scope" section defines the overall standardization topic: the definition of the successor of "Internationalization Tag Set (ITS) 1.0", that is: metadata items for the multilingual Web. To assure that ideas about new metadata items (ITS "data categories") do not drive the overall progress, the following sentences have been added to this section:

> "In addition to the data categories mentioned above, MultilingualWeb-LT MAY discuss data categories in the following areas. The MultilingualWeb-LT Working Group does not commit to define metadata for these data categories, but MAY do so, if this does not influence the timeline described below."

A part of the timeline from the charter (see charter section 2.3), that is the part relevant for the reporting period January – December 2012, is summarized in below table. The table compares the timeline with the actual timing[3].

Timing defined in the charter	**Actual timing in the reporting period**
February 2012: First teleconference	9. March 2012
April 2012: First face-to-face meeting	13. March 2012
May 2012: FPWD of specification about metadata for multilingual language technologies and localization processes	26. June 2012
November 2012: LCWD of specification	6. December 2012

For some items, a slight delay to the charter timeline can be observed. Nevertheless, the following approach assures that the finalization of the standard "ITS 2.0" will not be delayed.

4.1. Waterfall model versus Implementation Driven Model

The traditional standardization approach taken by W3C or other standardization forums is similar to a software development waterfall model:

1. A committee defines features.
2. After agreement about features, implementation work is starting.
3. Implementation experience may lead to changes in the feature specification.
4. Several iterations of the standardization process 1-3 may be needed.
5. The specification is finalized.

With the development of HTML5, W3C has started in various areas an *implementation driven process*: features are only discussed if there is an implementation commitment and also prototypes and sample files demonstrating the proposed feature.

The W3C working group producing metadata has adopted this approach for two reasons: First, to avoid feature creep. Second, to drive the overall process by real implementations and use cases.

The effect of this approach is that we expect a relatively short distinct implementation-testing phase. This phase is called out in the charter in section 2.1, defining the W3C working group deliverable "test suite":

[3] „FPWD" means in W3C "first public working draft"; "LCWD" means "last call working draft", that is a working draft endorsed by the working group as feature complete.

"The Working Group will produce a test suite intended to promote implementation of the Candidate Recommendation, and to assess interoperability between these implementations."

In the traditional standardization approach, the test suite is built after completing the feature definition. In the W3C working group defining LT-Web metadata, that is, the "ITS 2.0" standard, feature discussion, implementation development and testing are being conducted in parallel. In the following section we will describe how this approach helps to manage the overall project process.

4.2. Test Suite Milestones and Standardization Progress

The W3C working group deliverable "test suite" is driven by the LT-Web deliverable D2.3. This is due in M21 (= September 2013), but the partner leading the underlying task 2.3, TCD, started the work already in July 2012[4]. In coordination with the overlapping constituencies "W3C working group participants" and "LT-Web project partners", several milestones were defined:

- 15 November 2012 **M1**: review of test suite files provided by TCD and contributors.
- 17 December **M2**: Implementers have run at least one test file for each type of ITS 2.0 usage (local, global) and input files applicable for the implementer (XML, HTML).
- 31 January 2013 **M3**: Implementers have run all test files.
- 8 March 2013 **M4**: Implementers use all data categories in applications.
- 15 March 2013 **M5**: Implementers have run 80% of the test files successfully.

Both milestones M1 and M2 relevant for this reporting period have been reached in time.

As described in section 3.3 and 3.4 of this report, this approach of running feature definition and implementing / testing in parallel is crucial in assuring the quality of the project progress.

5. Separation and Interconnection of Communication Instruments

The communication instruments for the three constituencies are described in D1.2.1 and summarized briefly below:

- For LT-Web and W3C working group: 1) mailing lists for feature discussions, testing and external feedback: public-multilingualweb-lt@w3.org, public-multilingualweb-lt-tests@w3.org, public-multilingualweb-lt-comments@w3.org. 2) a dedicated issue and action item tracking system 3) a wiki. 4) IRC channel on the W3C public IRC server #mlw-lt. All communication is in public.
- For the MultilingualWeb program committee: 1) mailing lists for workshop planning and logistics. 2) a dedicated issue and action item tracking system 3) a wiki. 4) IRC channel on the W3C public IRC server #mlw. All communication (e.g. wiki read / write) is visible only to program committee members and to the W3C staff.

The clear separation of the communication instruments is complemented by interconnection via sharing personnel. E.g. the MultilingualWeb-LT working group chairs are also part of the program committee. An important aspect of this relation is governance: the program committee is chaired by Arle Lommel from DFKI, who is also participant in the MultilingualWeb-LT working group. But the review process and planning of the program does not put a focus on MultilingualWeb-LT work. It is the aim of MultilingualWeb-LT (LT-Web) to contribute in this way to the sustainability of the independent & unique MultilingualWeb brand.

[4] See the test suite mailing list at http://lists.w3.org/Archives/Public/public-multilingualweb-lt-tests/ for more information.

6. Budget Considerations and Quality Assurance

The LT-Web DoW defines in section B2.4 “Resources to be committed” a “budgetary reserve”, with the following purposes:

1. Support the outreach of the LT-Web metadata.
2. Enable the creation of new reference implementations and data sets.

In the first area and during the reporting period, LT-Web has undertaken the following efforts:

- Sponsoring of the MultilingualWeb ancestor project workshop in Luxembourg March 2012
- Sponsoring of a TC37/SC4 event in Madrid June 2012 http://www.tc37sc4.org/ and the MLODE workshop http://sabre2012.infai.org/mlode September 2012.
- Press releases translations for the creation of the W3C MultilingualWeb-LT working group in March 2012, undertaken by partners LINGUASERVE and VistaTEC http://www.w3.org/International/multilingualweb/lt/#press-releases .
- “Videolectures” editing of videos for the MultilingualWeb workshop June 2012; see online videos at http://www.multilingualweb.eu/en/documents/dublin-workshop/dublin-program .
- Support of travel for non-funded participants in the workshop or face-to-face meetings (8 trips).

In the second area, LT-Web has used the budgetary reserve to bring in the following new implementations:

- Adobe to create ITS 2.0 metadata support in the Apache Jackrabbit open source content repository.
-]init[to create an open source LibreOffice plugin for authoring content including ITS 2.0 metadata.
- Logrus to create browser-based meta-data display to end users.
- Tilde to create an Enriched Terminology Annotation Showcase, demonstrating the usage of ITS 2.0 metadata.

Most of these implementations will be finalized in 2013. However it was important in the 2012 reporting period to allocate the money so that enough time for actually creating the implementations will be available.

The quality assurance for these and LT-Web partner’s implementations relies on the milestone based test suite driven approach described in section 4.2 of this report and the action and issue tracking system described in D1.2.1. This setup, together with a transparent project communication approach conducted nearly completely in public, has proven useful in moving work items successfully and in a timely manner forward.