

D5.2.3 Project Progress Report (third half year)

ModelWriter

Text & Model-Synchronized Document Engineering Platform

Project number: ITEA 2 13028

Work Package: WP5

Task: T5.2 - Project Coordination and Reporting

Edited by:

Ferhat Erata <ferhat.erata@unitbilisim.com> (UNIT)

Moharram Challenger <moharram.challenger@unitbilisim.com> (UNIT)

Date: 29-Sep-2016

Document version: 1.0.0

Apart from the deliverables which are defined as public information in the Project Cooperation Agreement (PCA), unless otherwise specified by the consortium, this document will be treated as strictly confidential.

Document History

Version	Author(s)	Date	Remarks
0.1.0	Ferhat Erata Moharram Challenger	29-Sep-2016	Draft

Table of Contents

DOCUMENT HISTORY.....	2
1. INTRODUCTION.....	4
■ <i>Role of the deliverable</i>	4
■ <i>The List of Technical Work Packages</i>	4
■ <i>Structure of the document</i>	4
■ <i>Terms, abbreviations and definitions</i>	4
2. THE PROJECT PROGRESS REPORT (2016 SEMESTER 1)	4

1. Introduction

|Role of the deliverable

This document is the first version of the project progress report covering 2016 semester 1.

|The List of Technical Work Packages

UC Code	Requirements derived from
WP2	Semantic Parsing and Generation of Documents and Documents Components
WP3	Model to/from Knowledge Base (synchronization mechanism)
WP4	Knowledge Base Design and Implementation
WP6	Architecture, Integration and Evaluation

|Structure of the document

This document is organized as follows:

- Chapter 1 introduces the document.
- Chapter 2 provides the PPR

|Terms, abbreviations and definitions

Abbreviation	Definition
RDF	Resource Description Framework
WP	Work Package
UC	Use Case

2. The Project Progress Report (2016 Semester 1)

Project key data		
Project name	13028 ModelWriter	
Full length title	Text & Model-Synchronized Document Engineering Platform	
The project envisions an integrated authoring environment called "ModelWriter" for Technical Authors (such as Software or Systems Engineers etc.) which will combine a Semantic Word Processor (= the "Writer" part), looking like a usual word processor but capable to "understand" pieces of text and transparently create models of contents out of them; and a Knowledge Capture Tool (= the "Model" part), looking like familiar information modelling tools such as UML, BPMN, ReqIF, etc. ModelWriter will allow Technical Authors to freely move bi-directionally and interactively between text and model to enhance the quality (consistency and completeness) of the technical documents.		
Call & project ID	ITEA 2 Call 8 - 13028	
Time frame	start: 01-10-2014 end: 30-09-2017	
Size	PY: 48.09 M€: 2.8	
Project leader	Ferhat Erata (UNIT Information Technologies R&D Ltd.)	
Involved countries	France, Turkey	
France	Etienne Juliot (OBEO)	14 PY
Turkey	Aydin Can Polatkan (Mantis)	34 PY
PCA status	PCA has has not been signed yet	
Project page	13028 ModelWriter	
Latest FPP	Change Request (23-09-2016)	
Latest PPR	Progress report in 2015 (semester 1)	
Latest review	ModelWriter #1 (a.m.) (24-09-2015)	
Next review	ModelWriter #2 (p.m.) (15-11-2016)	

STG evaluation	
Submitted:	
STG Reviewers	

Project acronyms

KB (Knowledge-base), UC (Use Case), MW (ModelWriter), ALM (Application Lifecycle Management), EMF (Eclipse Modeling Framework), RDF (Resource Description Framework), QDMS (Quality Document Management System), MBSE (Model Base Software Engineering), BAFLING (Back and Forth Linguistic Processing), DL (Description Logic), Req. (Requirement), FORL (First Order Relational Logic), +ROI (positive Return of Investment = Benefit), FEAD (Front End Accessory and Drive), EGR (Exhaust Gas Recirculation)

Top 4 overall targeted innovations	
1	<p>Capability to maintain a readable textual document (using an editor) and relate its content to existing elements of models</p> <p>Main contributors: Obeo, UNIT, LORIA</p> <p>State-of-the-Art: > There are some Document annotation systems > A new capability is to annotate a text using an ontology > A new Recommendation system is addressed (synchronization links automatically proposed)</p>

Top 4 overall targeted business impacts	
1	<p>MBSE development</p> <p>Main contributors: All</p> <p>Market / competitors: Challenge in MBSE development is how to maintain the coherence between multiple distributed models or between models and documentation.</p>

2	Model / Text Synchronization Engine with iterative and interactive matching synchronization
	<div> <div>Main contributors</div> <div>Obeo, UNIT, KocSistem</div> </div> <div> <div>State-of-the-Art</div> <div> > Hand made synchronization available > Only doc generation or reverse engineering exists without interaction > There is no platform which extensively supports sync. between technical texts, models and knowledge base. Also, there is no Eclipse based platform capable of sync. text or model with a KB. </div> </div>
3	Formal Specification and Verification of Semantic Relationships between software and system artefacts
	<div> <div>Main contributors</div> <div>UNIT</div> </div> <div> <div>State-of-the-Art</div> <div> Formal specification and verification techniques have been used widely in industry, especially in the development of safe and secure systems. However, their use in the specification for the construction of semantic relationships between software/system artefacts need further investigation which will be addressed. </div> </div>
4	Semantic Annotation of Text with Model Elements
	<div> <div>Main contributors</div> <div>CNRS/LORIA, Obeo, Airbus</div> </div> <div> <div>State-of-the-Art</div> <div> Semantic Annotation is mostly restricted to the annotation of text with concepts or relations. ModelWriter will investigate how to make these approaches more generic and extend them to more complex items such as axioms and cardinality restrictions. </div> </div>

5	Reversible Semantic Processing
	<div> <div>Main contributors</div> <div>CNRS/LORIA</div> </div> <div> <div>State-of-the-Art</div> <div> Semantic parsing maps text to semantic representations. Natural language generation maps semantic representations to text. While these two processes are standardly treated independently, we aim to produce a system that both parse and generate a simple interaction between text-to-model and model-to-text conversion. </div> </div>
6	Natural Language Generation from KB Data
	<div> <div>Main contributors</div> <div>CNRS/LORIA</div> </div> <div> <div>SotA</div> <div> Existing approaches of generating from KBs generally use templates thereby yielding stilted text. There are also no approaches that can generate fluent text from arbitrary KBs. We aim to produce natural sounding text from KBs using a grammar based approach. </div> </div>
7	Capability to ensure / manage the synchronization between the artefacts (documents and models) bi-directionally
	<div> <div>Main contributors</div> <div>Obeo, UNIT</div> </div> <div> <div>State-of-the-Art</div> <div> A synchronization link management is addressed as a synchronization engine. </div> </div>
8	Design and implement of a heterogeneous knowledge representation approach, which allows combination & migration between different representations in MW
	<div> <div>Main contributors</div> <div>MANTIS</div> </div>

2	Reducing time to spend for Quality Control activities, by this way it provides manufacturers faster production
	<div> <div>Main contributors</div> <div>HISBIM</div> </div> <div> <div>Market / competitors</div> <div>QA QDMS, Microsoft Pinpoint QDMS</div> </div>
3	Synchronization of EGR and FEAD designs of engines with respect to design specifications in Ford-Otosan
	<div> <div>Main contributors</div> <div>Ford-Otosan, UNIT, KocSistem</div> </div> <div> <div>Market / competitors</div> <div> During product design phase, vehicle systems evolve according to changing market requirements. Therefore, change impact of one of these systems to other systems and compliance to design specifications must be evaluated in order to meet the quality standards. </div> </div>
4	More sell of Obeo Designer and Obeo SmartEA
	<div> <div>Main contributors</div> <div>OBEO</div> </div> <div> <div>Market / competitors</div> <div> > Support and Maintenance option in the contract "Obeo Designer" to support ModelWriter > Collaboration feature which support ModelWriter links with Obeo Designer Team > New feature in Obeo SmartEA to add documentation synchronization for strategic transformation and governance analysis </div> </div>

5	Supporting Traceability in ALM, e.g. by addressing impact analysis in the procedure of application development
	<div> <div>Main contributors</div> <div>UNIT, Havelsan</div> </div> <div> <div>Market / competitors</div> <div>> No similar offer in TFS / none announce</div> </div>
6	Expertise on document extraction
	<div> <div>Main contributors</div> <div>Obeo</div> </div> <div> <div>Market / competitors</div> <div> Some professional service will be proposed for the following purposes: - creation of dedicated engines to detect structured data in documentation - creation of new connectors to support designers during both modeling and installation - customization of ModelWriter core feature for specific needs </div> </div>
7	New option for Obeo Designer support
	<div> <div>Main contributors</div> <div>OBEO</div> </div> <div> <div>Market / competitors</div> <div> A new option is going to be provided in Obeo Designer to support ModelWriter features. https://www.obeodesigner.com/en/ </div> </div>
8	Creation of M2Doc thanks to MW knowledges
	<div> <div>Main contributors</div> <div>OBEO</div> </div>

State-of-the-Art	Representation of knowledge can vary in different abstraction levels. When a text like system description or req. definition is modeled as a single model that defines different aspects, there needs to be a coherence between definitions. MW will provide this coherence by an ontology used at different abstraction levels.		
Top 4 overall KPIs		Current	Target
1	Flexibility, speed, and ease of deploying and integrating ModelWriter within existing frameworks/platforms	N/A	+ROI in 3 Months
	Metric description	K1: demonstrated integration within at least one document management framework or platform. K2: agreed scenario for (automatic) synchronization launching (for a set of artefacts). K3: easy configuration of the (re)synchronization method. K4: number of commonly used doc formats supported. K5: number of commonly used modeling languages supported. K6: robustness of the synchronization with the 'real' documents managed externally. K7: robustness of the synchronization with the 'real' models managed	
2	Quality and precision of automatic synchronization update/preservation	N/A	100% correct links
	Metric description	K1: completeness of the list of modification operations that are allowed within models and within documents, without losing the synchronization. K2: warning mechanisms are implemented in case of de-synchronization detection K3: warning mechanisms in case of de-synchronization detection are easily interpretable, i.e. the user can identify easily what is synchronized / what has impacted the "desynchronization"	
3	Quality of the automatic suggestion of synchronization links	N/A	30%
	Metric description	K1: Coverage of ontology-driven automatic markers identification by the parser (i.e. if users make effort to develop terminological resources, then the terms should be retrieved - No existing term in the reference that is not retrieved). K2: Quality of automatic markers from the end user point of view: i.e. compared with a manual synchronization for a same use case and corpus).	
4	Coverage of the links/markers consistency check	X time	X/5 time
	Metric description	K1: Availability of Alloy based consistency check. K2: Diagnosis of detected inconsistency. K3: Availability of OWL axiom based consistency checks.	

5	Number of supported Domain Specific notations; Supporting various user-visible modeling languages	0	>=2
	Metric description At least two languages need to be supported such as: Ecore, OWL, Alloy and ReqIF		

Market / competitors	https://github.com/ObeoNetwork/M2Doc/blob/master/plugins/org.obeonetwork.m2doc/src-gen/org/obeonetwork/m2doc/template/Default.java		
Top 4 overall risks			
		S	P
1	Technical difficulty to deploy and integrate within existing frameworks/platforms	H	L
	Avoidance action	To consider most used technologies for both DSM and document edition	
	Back-up / mitigation plan	We have several frameworks as our target and if one of them has problem with integration, we will focus on the alternative one(s).	
2	Low performance and scalability	H	M
	Avoidance action	To create as early as possible a large data set to test the tool and its scalability, e.g. to test with real SIDP documents and models, typically more than 150 pages and 500 elements respectively.	
	Back-up / mitigation plan	We can change the architecture to client server and use background computation. Additional visualization techniques can be used to support the scalability of user interface, such as filtering, having different viewpoints and so on.	
3	Annotations and markers should be resistant to modification of input documents	M	M
	Avoidance action	This is the main technical challenge. e.g. if the user cuts/pastes a large paragraph, the link should be kept. Avoidance: This challenge needs to be as the heart of the initial design of the KB.	
	Back-up / mitigation plan	If the engine can't keep automatically the link, an interactive UI should help the user to massively reconnect these links based on previous mapping. We could use Eclipse facilities to keep previous states of documents, markers. links. etc.	
4	Lack of data inside the consortium prevents the training of high quality Natural Language Processing Tools	M	M
	Avoidance action	Use external data	
	Back-up / mitigation plan	Demonstrate usefulness of the NLP tools on data that is external to the project (e.g., RDF data from DBPedia dn text data from Wikipedia).	

5	Semantic annotation is highly ambiguous and yields many possible annotations for a single text/model fragment (LORIA)	H	M
	Avoidance action Apply disambiguation techniques (e.g., Lesk algorithm for word sense disambiguation)		
	Back-up / mitigation plan Allow for interactive annotation where the user manually provides the required disambiguation		

6	Faster QDMS reporting/generating	2 per day	10 per day
	Metric description	In HISBIM pilot factory, Quality control department personnel is able to report 1 or 2 QDMS documents in a day. With the ModelWriter system 10 QDMS documents generation is expected.	
7	Covering different textual representatives in the project	0	>=2
	Metric description	Structured texts and semi structured texts need to be supported, e.g. Java code, Mark down, and so on.	
8	Number of projects which use MW software deliverables	1	6
	Metric description	The projects can exploit MW and its deliverables. For example, in ASSUME project (another ITEA3 project), in the context of "WP3-System Engineering", UNIT, Ford-Otosan, and Havelsan is using the Tarski which is released under WP3 of MW.	
STG feedback on KPIs			

6	The definition of KB is still in early stage and the API for the semantic services is not defined		M	M
	Avoidance action	Mantis should work on this architecture and API for the next year.		
	Back-up / mitigation plan	The model / text synchronization can work without this semantic services connection.		
7	Gendoc project has some features which overlap MW		M	M
	Avoidance action	Discuss with Gendoc committer to align MW and Gendoc to avoid this overlapping.		
	Back-up / mitigation plan	Fork Gendoc or do everything in MW.		
8	MS Word plugin task is late		M	H
	Avoidance action	HISBIM needs to be speed up		
	Back-up / mitigation plan	Assigning more resource to release the first plug-in		
STG feedback on risks				

Changes in the technological and business relevance during the reporting period

- > New technology called M2Doc has been created by Obeo for document generation which could have connection with MW.
- > HISBIM intends to provide an extension for Cloud online version of ModelWriter plug-in for MS Word processor

Project statement on progress during the reporting period

- > A first version of ModelWriter Synchronization has been released. Some UIs have been create to start obtaining feedback from users.
- > The second versions of the main components (WP2 and WP3-Tarski) of the project are developed. The integration plan for the 2nd release is completed and the technical integration procedure is started.

STG recommendations

Exploitation

Updates to partners' exploitation prospects

UNIT Information Technologies R&D	sme	TUR	11 PY
Centre National de la Recherche	res	FRA	8 PY
Hisbim Bilgi ve İletişim Teknolojileri	sme	TUR	7 PY
KoçSistem	ifc	TUR	7 PY
Mantis	sme	TUR	8 PY
OBE0	sme	FRA	4 PY
Airbus Group SAS	ifc	FRA	2 PY
Ford Otosan	ind	TUR	1 PY
Havelsan	ind	TUR	0 PY

Other updates

Top 8 overall partners' Exploitation Related Achievements

1	Dissemination	Workshop	The 7th International ModelWriter Workshop		Realised
2	Exploitation	New product	Release of WP3 (Tarski platform) Components		Realised
3	Exploitation	New product	Sync. of EGR & FEAD designs with specs		Planned
4	Exploitation	New product	Creation of M2Doc thanks to MW knowledges		Planned
5	Exploitation	New product	Synchronization Engine and UI - v1		Realised
6	Exploitation	New product	Detecting Model Inconsistencies		Realised
7	Exploitation	New product	Implementation of the MW plug-in for MS Word		Planned
8	Exploitation	Enhancement	Component ontology_v1		Realised

Realized Exploitation Related Achievements statistics

Dissemination	Exploitation	Standardisation	New company	Patent	Human capital
Total: 13	Total: 16	Total: 5	Total: 0	Total: 0	Total: 0

Work progress during the reporting period

Top 4 technical achievements

1	Release of a semantic parser in WP2
Details	A semantic parser which parses text (normalized System Installation Design Principles) into OWL formulae
2	Initial traceability model for Airbus use case
Details	A model in Alloy language is provided that describes traceability links and constraints between System Installation artefacts (documents, model elements) and that can be exploitable using Tarski plugin to verify the consistency of the concrete links instances
3	WP3 Tarski platform development
Details	The 4 components which constitute WP3 is released (namely Tarski) in a stand-alone way to be used independently from MW platform. Tarski provides services for automated analysis of dynamically configurable semantics of traceability (trace-locations and links).
4	Synchronization Engine and UI v1 in WP6
Details	The first version of Synchronization Engine and User Interface (UI) is released. A common knowledge base meta model and its corresponding API to be used by the components of WP2 and WP3 have been proposed and developed by OBE0.

Top 4 next technical targets

1	A large training corpus
Details	Use generation and data expansion to create a large training corpus on which to train a statistical parser
2	Recognition of compound words by the parser
Details	Compound terms are numerous in all technical domains, thus terminological ontology may describe many compound forms. The next version of parser plugin needs to handle compound forms. This feature is important from usability point of view.
3	Synchronization of EGR and FEAD designs of engines with respect to design specifications in Ford-Otosan
Details	With this technical target we claim that ModelWriter can be able to be applied on review process of engineering design in order to increase the productivity of cross functional engineering teams by performing the impact analysis of design changes.
4	Complete the support for complete scenario
Details	Currently, the WP2 and WP3 are working separately. As the next target they will be working together to realize the complete scenario using the proposed knowledge-based metamodel.

5	Development of a consistency checker (Writer part - WP2)
Details	A consistency checker which permits identifying and removing sources of inconsistencies in an existing knowledge base.
6	Automated consistency checking (Model part - WP3)
Details	UNIT finalized the component that checks the consistency between trace links and locations of different work products such as code, model, and text elements with regards to the formal specification uploaded by the user.
7	Extended SIDP ontology to improve the automatic suggestion of markers/links by the parser

5	Use deep learning techniques to learn a semantic parser and a generator from training data
Details	
6	Cloud online version of ModelWriter plug-in for MS Word processor
Details	HISBIM has researched to make available MW plugin for cloud-based online Microsoft Word. Development activities of the new MW plug-in online version have already begun.
7	How to manage effective links to documents/artefacts living "outside" ModelWriter?

	Details	The ontology is used by the parser (WP2) to suggest markers/links. The v1 describes equipment, parts, fasteners and attachment tools, etc. New vocabularies (zone, environment, etc.) have been added to be used for the same intent.		Details	The goal is to enable links to external authoring tools and to stored documents.
8	ModelWriter plug-in for MS Word processor			8	Writer part enhancement
	Details	HISBIM already developed a MW plugin for MS Word (draft version, some requirements not fulfilled yet) which will enable a seamless integration with MW on Eclipse platform. User interface is ready, and now working on integration with an API developed by UNIT.		Details	Supporting structural creation of trace locations on Java, C, and C++ programs as a writer part enhancement (in WP6)
Top 4 next technical targets from latest PPR					
1	Extension of BAFLING to Airbus Data				
2	Using BAFLING for Synchronization				
3	Development of the knowledge base meta model				
4	Automated consistency checking				
Top 4 issues			Impact		
1	Turkish Use Case: HISBIM QDMS use-case will be applied in Turkish language.		HISBIM is expecting MW will easily penetrate Turkish market by Turkish QDMS ability.		
	Details	HISBIM and MANTIS will develop Turkish version of MW for Turkey market. HISBIM have provided QDMS materials, MANTIS have already begun to model these materials.	Mitigation action HISBIM aims to develop Turkish version of MW which is independent from international version. Also, it would be independent from QDMS system by initial configurations.		
2	We need to clarify how to integrate the ModelWriter product into an industry/business environment		Difficult to envision the technical needs and scenarios for a concrete deployment		
	Details	Depending on contexts, documents are managed in a Document Management System or stored in repositories. There is a need to demonstrate how ModelWriter tool can be integrated with these "external" environments.	Mitigation action Specify, define and illustrate a solution to integrate ModelWriter tool with these "external" environments.		
3	Scalability of state of the art reasoner engines		Reasoning on big models can be in-efficient		
	Details	The recent experiments show that the models with elements around 10000 items can be analyzed in a reasonable time. Based on the provided use cases, it seems that the supported size is acceptable. However, the bigger models might be an issue for future MW cases.	Mitigation action To mitigate this issue, we are planning to propose a new decision procedure which will incrementally use underlying backend reasoners.		
4	Lack of training data		This prevents developing a more robust statistical approach.		
	Details	For increased robustness and coverage, a statistical supervised approach would be more appropriate. This requires large amount of training data.	Mitigation action We plan to design and implement data extension methods which will allow the automatic creation of large quantities of training data		

5			
	Details		Mitigation action
6			
	Details		Mitigation action
7			
	Details		Mitigation action

8						
Details				Mitigation action		
Deliverables (overall status)				Due	Total	
Number of deliverables (due / total)				68	112	
Already finalized				56	82%	50%
Delayed (> 2 months) on due				12	12	100%
Details				The project has a large number of deliverables which are inherited from the former leader of the project. As the STG also recommended to reduce the number of deliverables in the previous PPR, we have submitted a change request. However, this CR has not been approved yet). 27 deliverables are reduced by merging them to the other deliverables and total number of deliverables are now 85. Also, some of the deliverables are omitted due to the withdrawal of Belgium consortium. The details and justifications of these merges are elaborated in Annex F of updated FPP. On the other hand, 12 deliverables are behind schedule. Among them, 7 deliverables are merged to the others as discussed in the change request. Also, "D4.7.1 Future ModelWriter-Enabled Use Cases" is postponed (in change request) to the later phases of the project. And 2 deliverables are omitted due to withdrawal of Belgium consortium. As the result, 2 deliverables are delayed due to integration challenges of technical WPs (2, 3, 4, and 6).		
Delayed (> 2 months) on total				56	12	21%
Details				We are not expecting any cascading delay for the other deliverables.		

Actual-vs-planned (overall status)				Current	Total	%	
Time elapsed (months)				21	36	58%	
Planned effort consumption (PY)				26.9	48.1	56%	
Reported actual effort (PY)				0%	25.2	48.1	52%
Discrepancies explanation (partners)				> Mantis (-25%): Difference mainly occurred due to synchronization and late start of the project. > Airbus Group SAS (-71%): Late availability of the platform to start tests and implementation. Missing partner at WP2 due to sick leave of the PhD who worked on the parsing.			
Project technical progress (%)				94%	100%	60-70%	
Comments				> 1st release of platform is issued > The second versions of the main components of the project are developed and the integration plan for the 2nd release is completed.			

STG feedback on work progress	

Feedback to previous STG remarks	STG responses
----------------------------------	---------------

Based on the comments from STG in the previous PPR, the following items are updated:

- > The KPIs are more clarified, e.g. by describing their measurement methods
- > Belgium consortium is removed (since it is not funded)
- > The number of deliverables are adjusted which is proposed in a change request.
- > The MS-Word processor plugin is under development. Also, the cloud online version is planned to be developed.

