Work Package 3

Deliverable (Delta report) for Sprint 5

Reporting from 2019-07-01 to 2019-12-31

Version 1.0

Work Package Responsible:

Markus Borg, RISE Research Institutes of Sweden AB

Task Leaders:

Task 1: Markus Borg, RISE Research Institutes of Sweden AB

(Acting) Kristian Wiklund, Ericsson, Sweden

Task 2: Frédéric Loiret, KTH, Sweden

Task 3: Karl Meinke, KTH, Sweden

Task 4: Serge Demeyer, Ericsson, Sweden

Task 5: Tanja Vos, Open University, the Netherlands

Table of Contents

WP3 Sprint 5 Executive Summary

Sprint 5 concludes and WP3 is in a position to start the sixth and final sprint. During Sprint 5, the WP3 leader Markus Borg (RISE) entered a period of parental leave that will last until the middle of Sprint 6. In the meantime, Tanja Vos (Open University) has stepped in as the operational manager running the bi-weekly pulse meetings. Risks related to this temporary change of leadership were identified in the risk management process already in Sprint 4. Thanks to frequent communication, WP3 has managed to avoid any major issues related to Markus' absence. WP3 will continue to monitor the related risks.

Sprint 5 included two summer months with limited activity (July-August) followed by preparations for the second annual ITEA review in September. Markus Borg represented WP3 at the review meeting in Berlin October 1. The assessment was positive, but WP3 acknowledges the need to incorporate the WP3 research into the WP6 TAIM model. Our work in this direction has started, and will be assigned the top priority in Sprint 6. Our primary target is to document key research findings as test automation patterns, in line with the rest of the TESTOMAT project. Examples include adoption of metamorphic requirements testing and scriptless GUI testing. Another dissemination activity that we prioritized during Sprint 5 was the production of learning videos for the TESTOMAT YouTube channel. As we now are ready with the videos, we will instead shift our focus in Sprint 6 to writing booklets.

Selected highlights from the five tasks follow:
- Task 1: Ericsson drives most of the research related to test case maintenance. Publications on test code clones and test smells with Empear and System Verification are pending. Furthermore, Verifyter has again presented related research at an industry workshop in the US.
- Task 2: KTH continued evolving the continuous integration dashboards. Furthermore, the task has prepared a workshop in Istanbul, Turkey in February, 2020.
- Task 3: Coverage studies have been carried out by two partner pairs: Ericsson-KTH and AKKA-KTH. The AKKA-KTH group has studied coverage results for machine-learning based testing using the AKKA Battery Charger case study. The KTH-Ericsson group has studied test coverage measurement and redundant test case removal using the tools SonarQube and NiCAD.
- Task 4: Mutant schemata has been investigated in parallel by VTT plus SAAB and the University of Antwerp. System Verification and SAAB have continued improving their respective mutation testing tools. SAAB works on rolling out mutation testing in different business units.
- Task 5: Focused on dissemination activities and evaluating web testing with the TESTAR extension for Selenium WebDriver.

WP3 Organization

In terms of effort, WP3 is the largest individual work package in the TESTOMAT project. To support information sharing and encourage *collaboration clusters*, WP3 invites all partners to bi-weekly telcos referred to as Pulse meetings. We refer to the time between two telcos as a "dash", i.e., a TESTOMAT *Sprint* comprises several WP3 *Dashes* - see Figure 1.
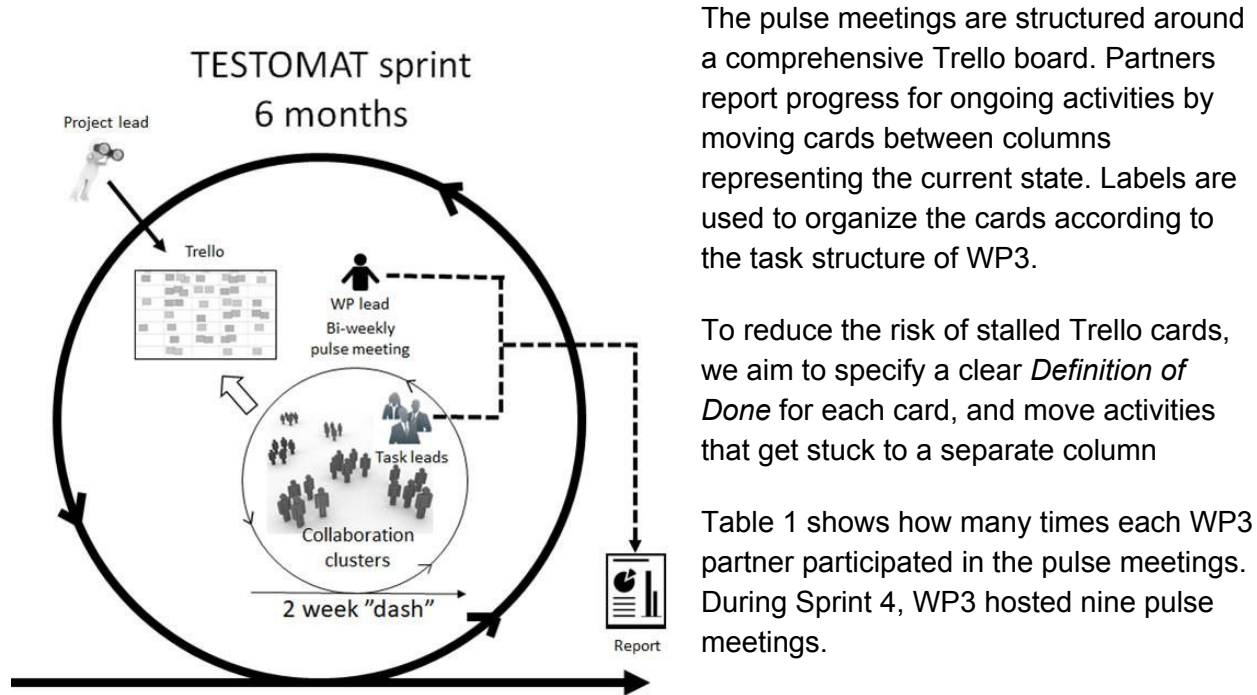
The pulse meetings are structured around a comprehensive Trello board. Partners report progress for ongoing activities by moving cards between columns representing the current state. Labels are used to organize the cards according to the task structure of WP3.

To reduce the risk of stalled Trello cards, we aim to specify a clear *Definition of Done* for each card, and move activities that get stuck to a separate column

Table 1 shows how many times each WP3 partner participated in the pulse meetings. During Sprint 4, WP3 hosted nine pulse meetings.

Figure 1: Sprints and dashes in WP3.

Effort and task allocation

Table 1 lists all contributing partners sorted by their total effort (according to the project proposal) in WP3. The table uses the following coloring scheme: problem owners, knowledge providers, and solution providers (note that some partners have dual roles, which is not presented in the table).

The five columns with the prefix 'T' show how the partners are allocated in the five WP3 Tasks. An 'x' in a cell denotes that the partner has committed effort to the Task. The final column shows the number of pulse meetings attended.

Table 1: WP3 partners, effort, and Task allocation for Sprint 4.

Partner	Total effort (PY)	T1	T2	T3	T4	T5
VTT Technical Research Centre of Finland Ltd.	4.3				x	
Comiq Oy	3.95	x	x			x
Qentinel Quality Intelligence Oy	3.9	x	x			x
Ericsson	3.2	x	x	x	x	
KTH (Royal Institute of Technology)	3.1	x	x	x	x	
Saha Bilgi Teknolojileri	3	x	x	x	x	
Ponsse Plc	2.9	x	x		x	x
Prodevelop	2.3		x	x		x
Saab AB	1.64				x	
The Open University of The Netherlands	1.51					x
Expleo	1.37					x
RISE AB	1.35	x	x	x	x	
Kuveyt Turk Bank	1.3	x	x	x		x
EfiCode	1		x	x		x
Parasoft Deutschland GmbH	0.82			x		x
TNO	0.81		x		x	x
Axini	0.8		x	x		x
Empear AB	0.8	x				
System Verification Sweden AB	0.8				x	
Verifyter AB	0.8	x			x	
Testwerk	0.7					x
Bombardier	0.5			x		
AKKA	0.3	x		x		
University of Mondragón	0.21					
Symbio	0.2		x	x	x	x

WP3 Sprint 5 - Progress and Achievements

VTT Technical Research Centre of Finland Ltd. (4.3)

VTT has continued the development of a proof-of-concept mutation tool for C/C++. The main aim of the research is to improve mutation generation time for C/C++ programs. Hence the mutation operators of the tool has been limited to mutate binary and unary operators of expressions. This way enough mutations can be generated to assess improvements in mutation generation time.

We used OpenCV as the first evaluation software. OpenCV is an open source package for computer vision. It is written mainly in C++ but it contains also some modules written in C. It is of medium size, the source code files contains 1.5 million lines. Building the software produces 30 shared libraries, 12 static libraries, and 27 test programs.

The tool created more than 150000 mutations. The binaries containing all the mutations can be built using the standard build process of the package. The time to build all mutations binaries took less than 1.5 times the time to build the ordinary version of the software. The time to build all mutants one at a time, which is the traditional technique for C/C++, would take around 300 times the time to build VTT's all mutations binaries. The improvement is more than two order of magnitude.

Preliminary observations of all mutations technique:
+ Building mutants is much faster (speed-up factor around 300)
+ Management of mutations and mutants is simple
+ Consumes significantly less disk space
+ Easy to integrate into build process
+ Simple to use
- The executable may be a bit slower
- The executable is a bit bigger

The tool was briefly demonstrated in the second ITEA review meeting.

During the final sprint the tool will be evaluated in an industrial environment provided by Ponsse and Comiq.

Comiq Oy (3.95)

During Sprint 5, Comiq continued with utilizing computer vision and machine learning in GUI testing.

We continued development of the initial version of a tool utilizing computer vision for GUI test automation. The need for this raises especially in enterprise applications deployed with Citrix or similar solutions. The goal is to enable automated testing of applications that currently can not be automated with existing tools or the maintenance of test automation implementation would be too labor intensive. The initial results are promising and we will continue exploring this topic.

We investigated utilizing machine learning for reducing development and maintenance work of test automation implementations.

In Sprint 6, we will continue with the both above mentioned tracks. We will also in cooperation with VTT and Ponsse evaluate integrating VTT's mutation testing tool to Ponsse's CI/CD pipeline.

Qentinel Quality Intelligence Oy (3.9)

The partner has no progress to report in relation to WP3 activities during Sprint 5. The plans for Sprint 6 remain unclear.

Ericsson (3.2)

Tasks 3.1 and 3.3 are currently joined, as KTH with strong support from 2 PhD students and one engineer is trying to reproduce coverage and duplications of three different test suites. As a first step, we are reusing our experience of running the cloning tool NiCad - which also finds type 1 duplications. The next step is to see if we can also find type 4 clones - as they are semantically different. We are aiming to see if we can refactor these test suites in any way automatically.
In addition, work has commenced on investigating if we can create a coverage dashboard. A first version of coverage as a booklet writing has been discussed. We are also working in collaboration with Empear and System Verification to do publications together. One is currently being pushed for Sprint 5 on code quality.

Most work in Task 3.2 has been supporting the task leader and the Turkish consortium, to arrange a visualization workshop between the Testomat Project, and two other projects. As the two master thesis students from KTH did not complete their work, we had to completely change the Ericssons focus in this project. It has also been clear that the particular skills to aid us in this task were no longer present. Work is of course ongoing internally to improve existing dashboards, but nothing will come out of the failed master thesis, but extra work. There is a slim chance that we might pull of something for the last sprint.

In Task 3.4, several proposals and meetings have been held to attempt to initiate mutation testing at Ericsson. But as there is currently no requirement, and also no business drive to

improve on such advanced coverage methods, we first need to do the more simple coverage approval.

We have as of now completely withdrawn from 3.5 - as we have very limited GUI tests. Therefore the conclusion is that sprint 5 has one journal paper submitted to JSS and one conference paper both are on cloning. One paper with System Verification and Empear. Plans for Sprint 6 are to

1. Evaluate the cloning and coverage in our three test suites, compare with current tool (SonarCube) and see if we can refactor and minimize these test suites (both within themselves, and possibly between themselves).
2. For Task 3.2 we aim at gathering material sufficient to present what we have done so far when it comes to test management, quality assessments and reporting in form of dashboards.
3. We still hope to get a mutation testing pilot going. We will therefore finish the mutation testing booklet as a way to disseminate the results.
4. Continue working on collaboration and publications with System Verification and Empear.
5. Discussion on creating demonstrators for Ericsson..

KTH (Royal Institute of Technology) (3.1)

During Sprint 5, KTH has initiated three new activities. KTH has begun collaboration with AKKA Germany on the AKKA Battery Charging Case Study. This case study has been addressed by several other Testomat partners, and so forms a good basis for comparison of our testing ideas and techniques. KTH has succeeded to demonstrate strong coverage results using machine learning (ML) based techniques and software-in-the-loop (SIL testing) when compared with n-wise testing.

During Sprint 5, KTH has also begun a new collaboration with the Ericsson on aggregating coverage measures and eliminating redundant test cases using the tools SonarCube and NiCad. We have also taken steps to generalise our earlier results on ML-based inference of metamorphic test requirements to larger code bases by collecting code bases from Github. At the Valencia meeting, KTH developed a simple TAIM model to show how Metamorphic testing could be rolled out in industry.

Finally, during Spring 5, KTH has initiated the development of an open-source web-based prototype for the visualization of DevOps metrics within a continuous integration environment.

Achievements in this Sprint are: (1) hiring a new PhD student and Amanuensis (Research Engineer) student to bring more labour resources into the project, and (2) successfully building a test harness between KTH testing tools and the AKKA Simulink model to make test coverage measurements.

During the final Sprint we plan to bring to a conclusion the above activities and publish our results.

Saha Bilgi Teknolojileri (3)

During sprint 5,

- Worked on systematic test case design and using for redesign our tests for testinium tool
- Working on coverage and effectiveness of systematic test case design.
- Comparison of systems designed by using gauge framework and designed by model based testing.
- Research for pitest mutation tool and making demo on open source project.

Plans for sprint 6,

- Injecting mutations manually on test automation management tool testinium and measure the effectiveness of both model based and gauge framework based tests.
- Injecting mutations by pitest mutation tool and on test automation management tool testinium and measure the effectiveness of both model based and gauge framework based tests.
- Working on journal paper on mutation testing
- Several presentations on workshops.
- Working on conference paper on Model Based Testing
- Releasing the version 2 of test visualisation for Testinium.

Ponsse Plc (2.9)

Delta:
- TESTAR is now implemented to the continuous integration pipeline of a new generation software
- TESTAR model based extraction evaluated and seems to be working now
- Qentinel Pace with the local test environment for frontend automated testing is now part of development teams daily work - 140 automated test cases implemented in less than three months
- Library for forest machine simulator test automation created and control system and UI testing is possible
- Part of the forest machine SW development teams are using simulator test automation in their daily work - around 500 test cases implemented during the sprint and tests are run automatically every night in continuous integration pipeline

Achievements:

- Test automation involved for the major part of SW development teams at Ponsse

Plan for the final sprint:
- Continuous integration for the simulator test automation environment is needed and also more unit and integration tests for the test software - always remember to test the testing software also. We want to be agile on this also.
- Qentinel Pace cloud environments for web test execution. Also mobile test execution if our applications are ready for testing in Android and iOS.
- Increasing number of TESTAR protocols to cover more areas of the SUT
- Increasing test coverage in frontend testing as more features are implemented
- Increasing test coverage in simulator test automation that test runs give us more feedback about the quality

Prodevelop (2.3)

Prodevelop has been adding functionality to the TATAT (Traceability and Test Automation Tools) toolset. The current version of the TATAT contains integrations with JIRA (Task management tool), Testlink (Test management tools) and Jenkins (Devops orchestrator). On the other hand, two testing tools have been integrated. The first one is Katalon studio, that is used to automate User Interface and API testing. The second one is Test*, that is used to test User interface without scripts.

MGEP in collaboration with Prodevelop, are taking steps to facilitate the location of the source of the error. Bearing in mind the information gathered in the previous sprint, it has been considered necessary to carry out a state of the art of industrial logging practices to determine the log structure that will enable the synchronization between the Posidonia and test logs. In addition, a tutorial about the Katalon GUI testing tool has been given in the VLC testing conference.

Major achievements
- Integration of several testing tools into the TATAT toolset
- Proposal of log structure.

Future plans

- Adding a coverage testing tool and static testing tool into the TATAT toolset
- Redefine the logs produced by Posidonia UC and adapt them to the new proposed log structure.
- Final integration of the TATAT with Jenkins.

Saab AB (1.64)

Delta:

- We held a [guest lecture about Mutation Testing](#) and [helped creating a lab](#) that is now part of the TDDD04 software testing course at Linköping University.
- We [improved the parallel mutation testing algorithm](#) for the mass deployment.
- We added functionality to [show test cases that kill mutants uniquely](#). This can be used to indicate if a test case is redundant.
- We added [functionality for incremental mutation](#), where a changeset is mutated instead of the entire code base.

Achievements:
- Dextool Mutate is now part of a software testing course at Linköping University.

Plans:
- Integrate a mutation schemata solution from WP3.4 to speed up mutation testing.
- Identify strengths and weaknesses of mutation testing when applying it to the avionic software domain. The goal is to identify possible approaches to automate test code inspection activities.
- Initiate a thesis project on [detection of equivalent mutants using code coverage](#).
- Initiate a thesis project on [visualization of mutation test results](#).

The Open University of The Netherlands (1.51)

During Spring 5, OU continued the collaboration with Testwerk, Axini, Ponsse, Kuveyt Turk and Prodevelop. The Open UI Data initiative was continued, working especially with Qentinel on validating and improving the UI data produced by TESTAR. A joint research paper with Prodevelop was written on the collaboration around TESTAR. Also, a journal paper about scriptless GUI testing and TESTAR tool was written during Sprint 5.

In addition to collaboration within TESTOMAT, a lot of effort was used for disseminating the results and TESTAR tool outside the project:
- 13.11.2019 Pekka Aho defended his PhD thesis "Automated State Model Extraction, Testing and Change Detection through Graphical User Interface" in University of Oulu
- 14.-15.11.2019 Scriptless GUI testing and TESTAR tool was presented in various public events and companies in the Netherlands
- 20.11.2019 Pekka Aho presented Scriptless GUI testing and TESTAR tool in TestDag http://www.testdag.nl/
- 29.11.2019 Pekka Aho presented TESTOMAT Project, Scriptless GUI testing and TESTAR tool in Kesko test expo in Helsinki

During Sprint 6 OU plans to finish the joint paper with Kuveyt Turk and 3 Master's thesis on:
- Inferring state models in TESTAR
- Mining potential temporal test oracles with TESTAR
- Using genetic programming to find fittest action selection strategy for TESTAR

Expleo (1.37)

Expleo had no goals for this sprint and therefore concentrated its efforts onto other work packages.

RISE Research Institutes of Sweden AB (1.35)

During Sprint 5, RISE has almost exclusively focused on WP management, including preparations and participation at the second annual ITEA review in Berlin in October and editing this delta report in December. The only exception was in August, when RISE presented a paper at the 3rd International Workshop on Machine Learning Techniques for Software Quality Evaluation (MaLTeSQuE) in Tallinn, Estonia. In Sprint 6, Markus Borg will return from parental leave and resume working on WP management and dissemination, e.g., the TAIM book.

Kuveyt Turk Bank (1.3)

During sprint 5 Kuveyt Türk worked on
- Testar together with Open University. We have worked together with the Open University to make Testar work for our internet banking application.
- Testinium together with Saha BT. We worked on different backlog items created with Saha BT to enhance and improve Testinium.
- Virtualize together with Parasoft. With Virtualize we experienced service virutalization.
- Model Based Testing (GraphWalker) together with Saha BT. We created models for our internet banking application and created automatically test cases. Test case execution is also implemented and an experience paper is written, presented at a conference.

During sprint 6 Kuveyt Türk plans to work on
- Testar together with the Open University. We have some issues and will work on these issues to make a complete run of Testar for internet banking application.
- Testinium together with Saha BT. We will work on the next backlog items to enhance and improve Testinium. We will do this also for Loadium.
- We will work further on model based testing, will also try to add mutation testing to it and plan to write an experience paper.

EfiCode (1)

The partner has no progress to report in relation to WP3 activities during Sprint 5. The plans for Sprint 6 remain unclear.

Parasoft Deutschland GmbH (0.82)

In Sprint 5, Kuveyt Türk Bank tried the Virtualize tool for test environment related problems. Some problems occurred such as recording the service as https and these problems are solved. Then, regarding to service virtualization, we had the results for one of the sample test scenarios. They used service called BES (personal pension system) to virtualize, the original login service of the BES is recorded first and virtual login service is created. Then, it is tested with various requests and used in different test scenarios. A bigger test scenario has been selected and necessary preparations have been done. Therefore, this study will continue with a bigger scenario in Sprint 6.

Major Technical Achievements:
- No particular technical achievements were planned for Sprint 5 but Parasoft's solution for code coverage and requirements traceability provides critical feedback about the completeness and thoroughness of the testing process, by correlating tests and coverage results with requirements. Presenting test results in this way prevents users from spending time and effort testing non-important functionality therefore Kuveyt Türk Bank focus on measuring the amount of code covered by the test automation by using Parasoft tool.

Plans for Sprint 6:
- Our collaboration will continue in Sprint 6 with the bigger scenario on virtualize tool with Kuveyt Türk Bank. In Sprint 5 they started, but some problems occurred such as recording the service as https and they have some authorization problems. Therefore, in Sprint 5 first we solved these problems and in sprint 6 we will continue our collaboration.
- Further, in Sprint 6, we will collaborate with Kuveyt Türk Bank with the use of SOAtest to execute Appium-tests to combine API, and Web apps in one framework.
- We may collaborate with Prodevelop in Sprint 6 using SOAtest/Virtualize tools. One of the main objectives is to integrate the analysis tools for non-functional requirements in the development process. This collaboration is under discussion right now.

TNO (0.81)

No activities have been carried out during sprint 5.

Axini (0.8)

Contact: Machiel van der Bijl
- Release documentbox prototype update. Model-driven engineering and model-based testing of document generators
- Release of AML 1.5 prototype. The new version of our Axini Modeling Language. It is not entirely finished yet, but we have a new language definition with parser and compiler.

- Release of data-definition and data-generator prototype. Complex data generation with constraint solving and constraint programming
- An update of the visualization of complex state machines.
- We worked on model-checking symbolic transition systems, including a prototype for easy to read LTL/CTL properties. Unfortunately this turned out to be more complex than we thought. We will continue in the next sprint.
- We worked on cloud-native testing in the Azure environment.

Furthermore we had several collaboration and outreach activities:
- Several presentations at workshops
- We created an MBT course including practical assignments for the University of Amsterdam. This is a continuation of the course that we created together with the Open University.
- We created several movies to explain our results in the Testomat project.
- There were several meetings for the Dutch consortium.

For the next sprint we focus on integrating our results. A new and exciting topic is that we will work on a model-driven engineering approach for pension administrations. This approach should result in an agile and easy to change framework for pension administration systems. It should be hard to make errors (they should be prevented by our approach).
We expect several MSc students to do their graduation project on Testomat topics.

Empear AB (0.8)

- In Sprint 5 we have been working on a paper with System Verification and Ericsson on test quality. We have delivered an open CodeScene workshop at GOTO Copenhagen in November. We also contributed to the Test Prioritization Tools booklet.
- In Sprint 6 we plan to continue our cooperation with System Verification and Ericsson with more research papers. We are also planning more open CodeScene workshops in India, Belgium and Netherlands during March, April and June.

System Verification Sweden AB (0.8)

- In Sprint 5 we have been working more on the joint paper between Empear, Ericsson and System Verification, regarding quality of automation test code and how this can be identified.
- We now have a draft of the paper that will be submitted to EASE'20, when all comments have been handled.
- In Sprint 6 we will start a master thesis that will work on improving the selection of mutations for our mutation test tool. We have finally found two students that are interested. The mutation test tool will also be presented at a SAST meeting the 3rd of March. In April the TARI roadshow will be conducted in Malmö and we will plan for and facilitate that. The plan is also to write another paper together with Empear.

Verifyter AB (0.8)

Presented "Instant Automatic Debug" at the Microprocessor Test and Verification Workshop in Austin, Texas. We described the MALAMUTE method where we use mutation testing to generate bug data which we subsequently use as training data to produce a model, using machine learning, that can point to the exact lines of code that causes tests to fail. This is allows instant debug of down to the line of code with 72% accuracy. So far only used on small examples. We will continue the work one larger code bases.

We also continue to work on predicting regression bugs using machine learning. For the final sprint we hope to be able to show it working in real customer scenarios.

Testwerk (0.7)

- In Sprint 5 we have been evaluating Testar in a live environment with application Wise of the OCLC company. We tested parts which have not been touched yet due to some issues.
  For a new customer QFSoft we started testing desktop application PoultrySoft with Testar. PoultrySoft is developed with WINDEV. Testar isn't able yet to test is. On hold.
  For a new customer Hanzehogeschool we started testing a web application OnStage with Testar. On hold. Waiting for a good test environment.
  For a new customer Carpenter Group we started testing web application Esira Food with Testar.
  All issues are reported on Github.
  Demo of Testar for the employees of Testwerk.
- In Sprint 6 we continue evaluating Testar with the above applications and retesting solved issues and new functionality.

Bombardier (0.5)

The partner has no progress to report in relation to WP3 activities during Sprint 5. The plans for Sprint 6 remain unclear.

AKKA (0.3)

In Sprint 5, AKKA started a collaboration with KTH. They apply their machine learning based test methodologies to our use case and analyzed the model coverage results of different input vectors and parameter sets. This will be finalized during the next sprint. Having a sufficient coverage is the basis to find and apply prioritization techniques, which will be done in the scope of work package 4.

University of Mondragón (0.21)

In Sprint 5, MGEP in collaboration with Prodevelop, are taking steps to facilitate the location of the source of the error. Bearing in mind the information gathered in the previous sprint, it has been considered necessary to carry out a state of the art of industrial logging practices to determine the log structure that will enable the synchronization between the Posidonia and test logs. In addition, a tutorial about the Katalon GUI testing tool has been given in the VLC testing conference.

Both MGEP and Prodevelop have collaborated in the organization of the last general assembly in Valencia.

Major Technical Achievements

- State of the art of logging practices

In the sprint 6, a log structure will be proposed that will allow synchronizing the Posidonia logs with the test logs and information of the requirements located in Jira

Symbio (0.2)

The partner has no progress to report in relation to WP3 activities during Sprint 5. The plans for Sprint 6 remain unclear.