MBSE2DL Executive Summary

Author:	ScopeSET Technology Deutschland GmbH
Contact:	Armin Müller email: amueller@scopeset.de phone: +49 8028 905658
Date:	July 5, 2022
Version:	1.0
Document-ID:	f/114238
Customer Approval:	

Version	Date	Changes
1.0	July 5, 2022	Initial Version

TABLE OF CONTENTS

1 Introduction

This executive summary introduces the Integrated MBSE Analytics Platform (IMAP) which has been developed in the course of the OSIP [OSIP] MBSE campaign's MBSE2DataLake project. It demonstrates how MBSE data can be downstreamed from early phases and integrated with Telemetry data or system operations data. Applying MachineLearning analytics and customized visualization on these sets of integrated data provides seamless analysis capabilities and data browsing experience for the user. The approach leads to linking the former separated domains of System Design and AIT/Operations. This in turn leverages the MBSE contained knowledge across later phases of system development lifecycle, as well as providing cross-domain users a single point of access to the combined data.

2 State of the Art and Challenges

We perceive the current state of the art and challenges in regard to MBSE adoption as follows:

- MBSE and telemetry/operations data are still silos today, both in terms of the people involved as well as in terms of the data and models generated or used
- Partial "point to point interfaces" of data exist, but a holistic integration has not been realised
- Serious amounts of time and effort are invested to create valid MBSE models in the early project phases. In later project phases, these models are not or hardly used or understood
- Not using the same models causes inefficiencies due to duplication of work, additional efforts in communication, increase in complexity and inconsistencies in data and documentation
- Knowledge transfer to successor projects is interpersonal rather than model based. Time and money for (new) model designs are invested repeatedly in each project
- Using a model-based approach along the lifecycle is hard to judge and evaluate from a business perspective, since business/management is too far away from systems engineering

Our working hypothesis: By integrating the engineering system model (MBSE system model) with later phases (Figure 1), systems test/operations data, the identification and elimination of errors in the context of the assembly integration test, but also in operations, could be significantly accelerated, simplified and optimised. This is additionally supported by suitable visualisations and manifests itself in less system failure, faster troubleshooting, and improved communication between the project participants.

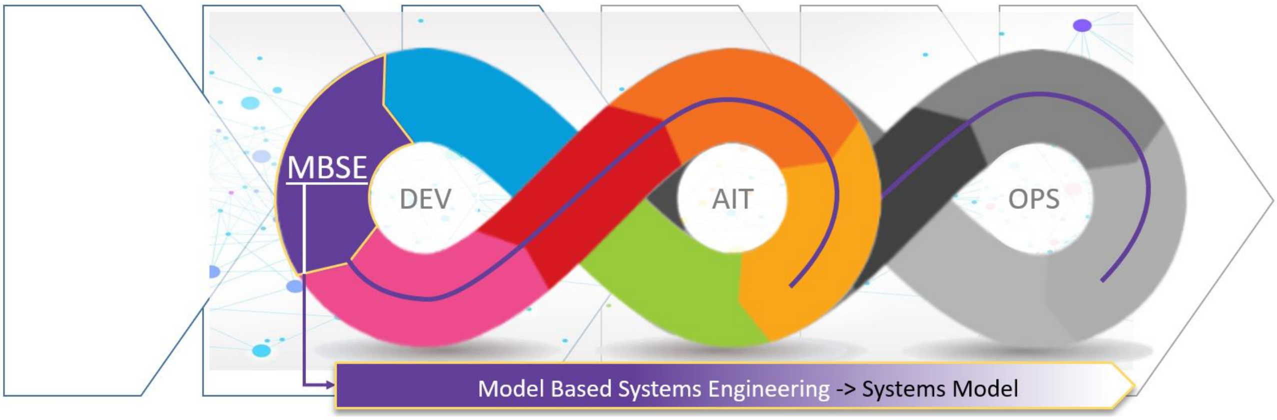

Figure 1 Integration of MBSE, AIT and Operations into one holistic Workflow -> "DevOps for MBSE"

3 Solution to the challenges

Within the framework of the IMAP demonstrator, it was shown that:

- The gap between systems design models and test/operation data can be closed in an integrated manner
- Holistic analyses for fault identification, referencing the system model in the background, are possible and can be well supported by Machine Learning
- After initial training, analyses become increasingly better/meaningful due to learning effects with MachineLearning
- The created solution can form the basis of a scalable and expandable platform for fully integrated data analytics
- The solution architecture is suitable for larger satellite projects (separation into backend/core/hardcoded, "expert areas" via Jupyter notebooks and end-user visualisations)
- The solution architecture is scalable for big data by relying on established BigData technology

Platform characteristics and implementation

IMAP consists of a set of integrated tools, services and web front ends to ingest both MBSE data (in the form of SysML models) as well as Telemetry definition and sensor data. A semi-automated tracing is performed which ultimately provides the capability to link and overlay MachineLearning based Analysis Results to well-established SysML block diagrams as shown in Figure 2.

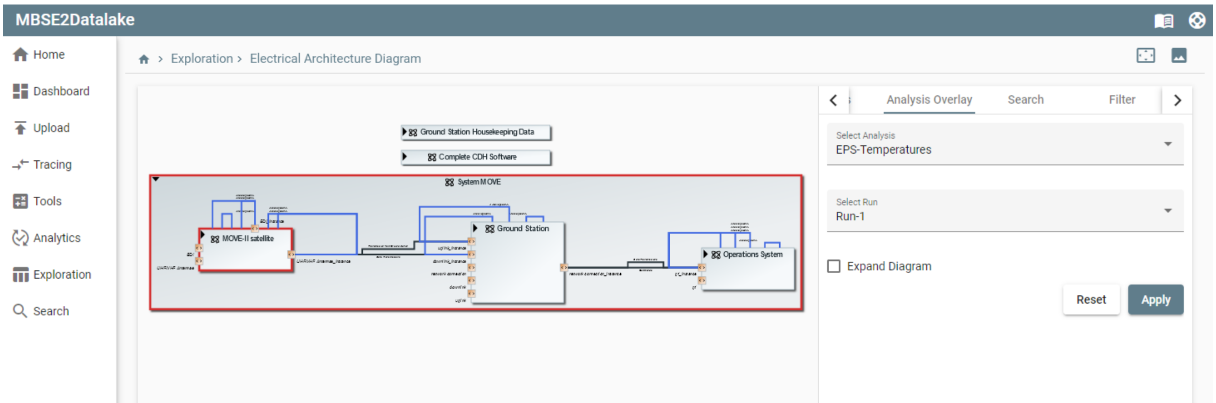

Figure 2 SysML Analysis Overlay

Those block diagrams are generated and autolayouted from the imported MBSE data as mapped to an RDF representation and stored in a scalable Hadoop file system. This file system also stores CSV files which are created by the Telemetry Importer and which are subsequently ingested into a set of extensible Analysis scripts.

Although these Analysis scripts can implement a certain pattern or MachineLearning approach (e.g. LSTM for the platform's demonstration case), every mission, model and kind of telemetry data will have their own unique characteristics metamodel. In line with a general DataLake approach ("Collect and store all data first, take care about use cases and data processing later"), these Analysis scripts are therefore not hardcoded as platform services, but are implemented in Python in Jupyter notebooks (which is the de-facto standard for data analytics).

Obviously, dealing with Jupyter Notebooks requires a kind-of expert level understanding (i.e. typically provided by a domain expert/data analyst). To allow easy access to non-expert end users, the platform provides a set of front-end components which retrieve their data either directly from the output of a Jupyter based Analysis script or the integrated and linked data in the DataLake. For most of the user, the complexity of data provision and preparation remains hidden in the background. He/She can thus concentrate exclusively on the analysis task/error elimination.

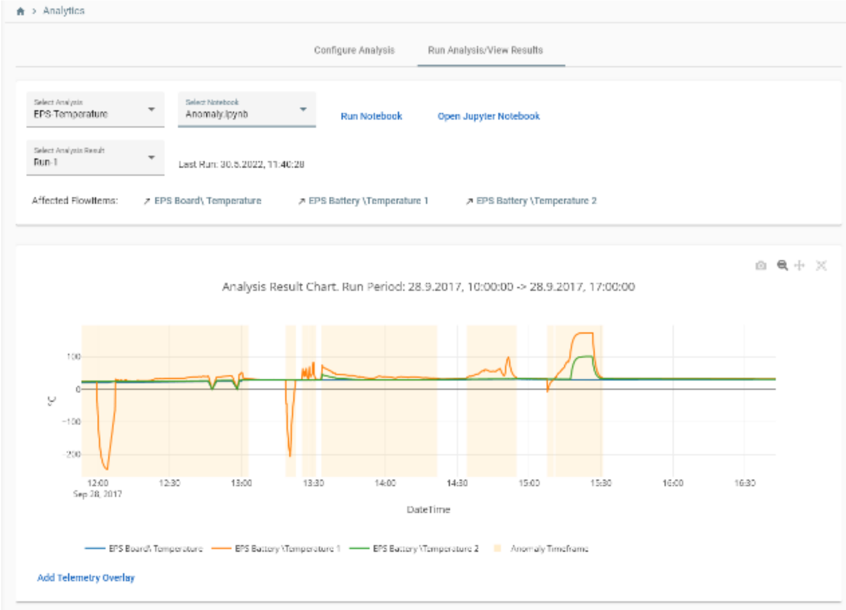

Figure 3 MachineLearning based anomaly detection results

In addition, the platform provides a search interface for all artefacts stored in the DataLake and an integration with Neo4J, which enables further graph-based analysis capabilities. Following the above approach, these graph analyses are also maintained in Jupyter Notebooks (including navigations into the Neo4J Graph browser).

The overall architecture of the platform is influenced by the BigData Europe [BDE] project , it is distributed as a set of Docker images and can basically be deployed from a single docker compose file. In the future, this stack can be scaled up into a Kubernetes cluster where different services are deployed on different hardware resources.

In a bigger context, we see the platform as an (initially) read-only consumer of an MBSE Hub, providing a broader picture and analysis capability. Eventually, results being produced by the platform could be streamed back to a MBSE Hub however.

Benefits - BizDevOps applied to MBSE

It is always a challange to judge business benefits of a digital solution. For topics in the field of digitalisation, it has proven useful to reference the BizDevOps [BDO]. The main units of the organisation, involved in value creation, work together towards a coherent solution.

Applying the BizDevOps approach to the use of MBSE in the enterprise, MBSE could act as a digital enabler. With end-to-end data integration and a model-based way of working, the following flow could emerge: The Business (= voice of the customer and strategic entity) specifies the requirements for the system to be designed. Based on those requirements, Development, across the domains and iterations involved, designs the system and creates digital models, including system components, system behaviour and system functions. Those are further used for analysis, simulation into assembly, integration, test and operations. With its flow through the various project phases, the models are charged with system integration data, are calibrated, simulated and kept synchronised with the real system until release to operation and beyond. System models are the first reference for troubleshooting the system in all phases of development and promote communication, both internally and with suppliers. Eventually, with the product handover to operations, the models are also handed over and form the digital representation of the system. Following this approach could solve several issues of today's State of the Art development processes on business level but also on level of operations.

4 Benefits provided by IMAP – referenced to a real project

Using the MOVE-II [MOV] project as a reference, an in-depth analysis of the procedures, deficiencies and lessons-learned has been performed by TUM. In addition, a potential future project (TUM MOVE-III) with an application of the IMAP platform has been compared against. As a result, the following key improvements have been identified:

- Significant reduction of time for searching/gathering data
- Improved communication across teams, based on visualizations and models
- Automated tests on available data, specified by experts, proceeded by test engineers
- End to end data browsing, independent from entry point
- Faster onboarding of people due to easy-to-use frontends
- Completely new analysis capabilities based on MachineLearning

Due to the fact that no formal criteria for efficiency measurement and performance gain have been established yet, only estimates, based on the above challenges and before and after analysis have been performed. The overall efficiency gain is estimated to be in the range of 30%-40%. This may be degraded slightly, depending on effort which will need to be spent on MachineLearning training (supervised vs. unsupervised) in relation to the confidence level of the resulting prediction models.

Literature

[OSIP] https://www.esa.int/Enabling_Support/Preparing_for_the_Future/Discovery_and_Preparation/The_Open_Space_Innovation_Platform_OSIP

[MOV] https://www.move2space.de/MOVE-II/

[BDE] https://github.com/big-data-europe