UNIVERSITY of York

This is a repository copy of *Monitoring ArchiMate Models for DataBio Project*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/212151/

Version: Accepted Version

Proceedings Paper:
Chaabouni, Kaïs, Bagnato, Alessandra and Garcia-Dominguez, Antonio orcid.org/0000-0002-4744-9150 (2019) Monitoring ArchiMate Models for DataBio Project. In: Franch, Xavier, Männistö, Tomi and Martínez-Fernández, Silverio, (eds.) Product-Focused Software Process Improvement - 20th International Conference, PROFES 2019, Proceedings. 20th International Conference on Product-Focused Software Process Improvement, PROFES 2019, 27-29 Nov 2019 Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) . Springer , ESP , pp. 583-589.

https://doi.org/10.1007/978-3-030-35333-9_42

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Monitoring ArchiMate models for DataBio project

Kaïs Chaabouni[1], Alessandra Bagnato[1], and Antonio Garcia-Dominguez[2]

[1] Softeam, R&D Department, France
{kais.chaabouni,alessandra.bagnato}@softeam.fr
[2] School of Engineering and Applied Science, Aston University, Birmingham, U.K.
a.garcia-dominguez@aston.ac.uk

Abstract. The Data-Driven Bio-economy project (DataBio) is a large scale project that aims to develop a platform that offers access to big data technologies in the domains of agriculture, fishery and forestry. This project applies the standard Enterprise Architecture language: "ArchiMate 3.0" for modelling the pilot studies and for modelling the software components in order to facilitate comprehension and communication between partners. The models are created with the modelling tool "Modelio" which allows contributors to collaborate on a shared version of the ArchiMate models. These models are monitored continuously by the monitoring tool "Measure Platform" and the model querying tool "Hawk". This paper describes the monitoring approach and specifies the metrics defined to evaluate the quality level of the models.

Keywords: ArchiMate · Enterprise Architecture · Models Metrics.

Project data

- Acronym: DataBio, Title: Data-Driven Bio-economy
- Start date: January 2017, Duration: 36 months
- Partners: INTRASOFT International S.A. Belgium (project coordinator), VTT Technical Research Centre of Finland LTD, SINTEF and 45 more partners including IT companies and research institutes [1]

1 Introduction

The DataBio project [2] aims to develop a big data platform based on existing partners' solutions and contains 27 pilot studies that fit among one of these categories:

- Improving precision farming and utilizing predictive analysis in agriculture.
- Improving forest monitoring, predicting risks and optimizing tree resources.
- Predicting fishery market and rationalising its environmental impact.

Each pilot integrates through its workflow a number of software components that are linked together and act as a data pipeline in which every component has a specific task along the data value chain from data collecting and processing (mostly satellite imagery and IoT sensors data) to analyzing and visualizing [3]. In order to facilitate the comprehension of the pilots requirements and the technological design of the components, there is a need for a common modelling language that allows people to have the same modelling conventions. Therefore we use the standard "Enterprise Architecture" language "ArchiMate 3.0" [4] which proved to be suitable for specifying requirements/strategies and has at the same time a wide range of concepts for modelling IT systems [5]. The modelling environment used for this task is "Modelio" [6] which allows partners to collaborate on synchronized SVN repositories containing the ArchiMate models. In order to maintain the quality level of the models throughout the project we defined new metrics for the models' quality and we monitor continuously the models' repositories with the monitoring tool "Measure Platform" [7] and the model querying tool "Hawk" [8]. This paper is structured as follows: section 2 presents the monitoring of the ArchiMate models, section 3 illustrates the defined model quality metrics and the final section ends with concluding remarks.

2 Monitoring of Modelio ArchiMate models

The DataBio ArchiMate models are structured in five Modelio projects described as follows :

- Three projects : Project 1, 2 and 3 corresponding to the pilots of the following domains of research: agriculture, fishery and forestry. These projects contain motivation views, strategy views and business process views.
- Project 4 for modelling software and IoT system components
- Project 5 for modelling "Earth Observation" data services.

These projects are monitored by the monitoring and analysis tool "Measure Platform" designed primarily for monitoring software projects, integrating third party analysis tools and creating a customized dashboard for visualization. Measure Platform collect periodically predefined "measures" that were developed to monitor the ArchiMate Models by interrogating the model indexing tool "Hawk" which allows to query the Modelio repositories [9]. For each metric, we add measures for the five monitored projects where we specify the query expression in **EOL** language (Epsilon Object Language) [11] which is then interpreted and executed by Hawk. Hawk optimizes the querying process by creating a graph database index that contains the different elements of the model and their relationships and thus improving the response time of the queries [10]. The collected measurements are stored in Measure Platform and can be visualized through the platform dashboard. Fig. 1 shows an example of the measurements that can be visualized by Measure Platform dashboards.

Fig. 1. Percentage of unrepresented elements in monitored Modelio projects

3 Metrics for evaluating models quality

The metrics that we use for evaluating the quality level of the models are inspired in part from literature review such as the 6C quality goals described by *Mohagheghi et al.* [12] for model driven software development. In addition, these metrics are inspired from our experience with monitoring DataBio models and evaluating their added value regarding to the purpose they serve. Hence, in this context, the models are evaluated by how much they provide understanding and clarity for users while having at the same time an efficient modelling process that makes it well worth the effort.

3.1 Metrics for optimizing the modelling process with Modelio

We present here metrics that reflect how optimal is the usage of Modelio in a manner that guarantees completeness and efficiency in the modelling process. Table 1 gives a summary of the collected measurements by Measure Platform according to the following defined metrics.

Table 1. Metrics for optimizing the modelling process with Modelio

Projects	**Proj.1**	**Proj.2**	**Proj.3**	**Proj.4**	**Proj.5**
Percentage of unrepresented elements	40%	52%	49%	15%	31%
Percentage of duplicate elements	49%	54%	44%	12%	10%
Percentage of empty diagrams	13%	6%	15%	29%	16%
Median diagram importance score	0	6	0	16	17.73

Percentage of unrepresented elements. Unrepresented elements are elements that have been created in the ArchiMate model and located in the Modelio explorer but are not displayed on any diagram. This is due to the deleting

of the element representation instead of the element itself or could be an element created in the model for future use but was never used afterwards. Having a big percentage of unrepresented elements implies having inefficiencies in the modelling process. On one hand these unrepresented elements are considered as a wasted effort because they add no value in the final diagrams. On the other hand, the presence of these unrepresented elements would result in a crowded project explorer which would increase complexity and decrease needlessly the visibility for the modellers. The monitored ArchiMate models for DataBio contain many unrepresented elements, averaging 50% in some projects (see Table 1 and Fig. 1).

Percentage of duplicate elements. Duplicate elements are different ArchiMate elements created in the models but represent the same concept. This redundancy can be the result of uncoordinated creation of elements by the different collaborators or a simple misuse of the modelling tool. The presence of duplicate elements add complexity for Modelio users and cause confusion in managing different copies of the same concept. Furthermore, these redundancies prevent Modelio users from identifying shared elements across diagrams and recognizing all relations associated to the same element. The first three DataBio models contain many redundancies (see Table 1) which is explained by the lack of experience of modellers freshly introduced to Modelio who are duplicating shared elements between pilots to use them in different diagrams instead of referencing the same element across different diagrams.

Complete diagrams. We define two metrics for ensuring the completeness of the ArchiMate diagrams. The first metric is concerned with the percentage of empty diagrams as it is self evident to assume that an empty diagram is a sign of incomplete work. However, when applying this metric in DataBio models we noticed the presence of "almost empty diagrams" that can contain for instance a few not related elements and therefore should also be considered as incomplete or not having a mature enough design. Hence, we introduce the second metric that measures the maturity level or the "importance score" of diagrams. The importance score was introduced by *Singh and van Sinderen* [13] as an attempt to formalize Enterprise Architecture metrics for measuring of the criticality and the impact of an element in an Enterprise Architecture model. The importance score is calculated based on assigned scores to elements and their outgoing relationships and therefore the more the elements inside a diagram are connected together, the more the importance score is bigger. This measure could also be considered as a indication of the maturity level of the diagrams if we assume that any thing of value must be important and should have a certain minimum defined importance score as opposed to "almost empty diagrams" which have very low importance score. In the DataBio monitored projects (see Table 1) we can see that there is still many empty diagrams and that the median importance score for diagrams is still very low especially for the first three projects.

3.2 ArchiMate comprehensibility metrics

The comprehensibility metrics evaluate the complexity to read ArchiMate diagrams by distinguishing the different elements and recognizing the connections between them. Moreover, these metrics entail also the ability to understand the concepts represented by ArchiMate diagrams such as the services, the components and their interactions. Table 2 gives a summary of the collected measurements by Measure Platform according to the defined comprehensibility metrics.

Table 2. Comprehensibility metrics for ArchiMate diagrams

Metrics	**Proj.1**	**Proj.2**	**Proj.3**	**Proj.4**	**Proj.5**
Number of diagrams	59	36	33	312	141
Number of elements per diagram	13	17	13	7	7
Relations to Elements ratio	0.89	0.80	0.82	0.80	0.86
Percentage of documented elements	15%	19.2%	23.7%	57.9%	61.3%

Number of diagrams. The number of diagrams reflect the size of the whole model. Having a big sized model increases the complexity for readers. For example, as wee can see in the table 2, the first three projects have reasonable number of diagrams, but project 4 and 5, which represent mostly the technological components in DataBio, have a big number of diagrams which can be disorienting for readers to grasp all the concepts represented by these models.

Number of elements per diagram. The number of elements per diagram metric is complementary to the previous metric because it highlights the density of diagrams and thus showing the real size of models in terms of total number of ArchiMate elements. The monitored DataBio models contain a reasonable number of elements per diagram averaging from 7 to 17 elements per diagram (see Table 2) which means that diagrams are not crowded and are easy to read.

Relationships to elements ratio. The relationships to elements ratio reflects the congestion of associations between elements and shows the number of different connections associated to the same element. The number of associations per element should be between 1 and 4, so that the resulted diagram would be neither congested too much nor sparse too much.

Percentage of documented elements. Modelio allows modellers to attach notes to the ArchiMate elements in order to describe the intended concepts represented in the diagrams. Although most of the elements have self evident names that do not require more explanation, other elements require more explanation for the readers especially if their names contain abbreviations, very technical terms or terms that describe different purpose from the intuitive and most common perception.

4 Conclusion

This paper outlines the adopted approach for monitoring ArchiMate models contained in Modelio repositories and the defined metrics that are used for collecting measurements on the monitored projects. ArchiMate models provided clarity and understanding throughout the DataBio project and therefore we needed to maintain a good quality level for the models. For this purpose, we defined metrics for model quality based on our experience with DataBio and inspired by other literature metrics. This has led to interrogate the models with a model indexing tool "Hawk" and a monitoring tool "Measure Platform" in order to evaluate models quality according to the defined metrics. We defined two sets of metrics: the first type was for optimizing the modelling process with Modelio and the second type for evaluating the ArchiMate diagrams.

For future work, we look forward to experiment with these metrics in other projects and analyse modellers feedback in order to adjust these metrics for ArchiMate modelling or for other modelling languages or methodologies.

Acknowledgements

This work is partially funded by "DataBio project" (No. 732064) under European Commission's Horizon 2020 research and innovative programme and "Measure project" (No. 14009) under the EUREKA ITEA 3 Programme.

References

1. DataBio partners, https://www.databio.eu/en/consortium. Accessed 27 May 2019.
2. DataBio homepage, https://www.databio.eu. Accessed 27 May 2019.
3. DataBio public deliverable: DataBio_D4.2 Services for Tests.
4. Josey, Andrew. ArchiMate 3.0. 1-A pocket guide. Van Haren, 2017.
5. FRITSCHER, Boris et PIGNEUR, Yves. Business IT alignment from business model to enterprise architecture. In : International Conference on Advanced Information Systems Engineering. Springer, Berlin, Heidelberg, 2011. p. 4-15.
6. Modelio Business Architecture, https://www.modeliosoft.com, (18 Apr 2019).
7. Measure Platform, http://measure-platform.org. Accessed 21 May 2019.
8. Hawk, https://github.com/mondo-project/mondo-hawk. Accessed 21 May 2019.
9. Al-Wadeai, Orjuwan, et al. "Integration of Hawk for Model Metrics in the MEASURE Platform." MODELSWARD. 2018.
10. Garcia-Dominguez, Antonio, et al. "Integration of a graph-based model indexer in commercial modelling tools." Proceedings of the ACM/IEEE 19th International Conference on Model Driven Engineering Languages and Systems. ACM, 2016.
11. Epsilon Object Language, https://www.eclipse.org/epsilon/doc/eol/, 14 June 2019
12. Mohagheghi, P., et al. (2009). "Definitions and approaches to model quality in model-based software development-A review of literature." Information and Software Technology 51(12): 1646-1669.
13. Singh, P. M., & van Sinderen, M. J. (2015, June). Lightweight metrics for enterprise architecture analysis. In International Conference on Business Information Systems (pp. 113-125). Springer, Cham.