Enterprise Architecture modelling with ArchiMate

Kaïs Chaabouni[1], Alessandra Bagnato[1], Ståle Walderhaug[2], Arne J. Berre[2], Caj Södergård[3], and Andrey Sadovykh[4,1]

[1]Softeam R&D Department, France
{kais.chaabouni,alessandra.bagnato}@softeam.fr
[2] SINTEF, Norway
{arne.j.berre,stale.walderhaug}@sintef.no
[3] VTT, Finland
caj.sodergard@vtt.fi
[4] Innopolis University, Russia
a.sadovykh@innopolis.ru

Abstract. The Data-Driven Bio-economy project (DataBio) focuses on developing new technologies and services for agriculture, fishery and forestry by exploiting the huge potential of Big Data technologies. This Lighthouse project includes 27 pilots and 91 technological components provided by 27 of the 48 project partners. It applies a standard Enterprise Architecture modelling language: "ArchiMate 3.0". ArchiMate models are created with the tool "Modelio" which allows contributors to create ArchiMate diagrams and collaborate on a synchronized version of the models. The DataBio models cover different aspects of the project from the specification phase including requirements, goals and strategies, to the implementation phase by describing the different processes of the tasks included in the work packages and representing the technological components. This paper describes the use of ArchiMate modelling applied in the context of the DataBio research project.

Keywords: Enterprise Architecture, Modelling, ArchiMate, Modelio.

Project data

- Acronym: DataBio
- Title: Data-Driven Bio-economy
- Start date: 1 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 **Data-Driven Bio-economy** project (**DataBio**) [2] focuses on exploiting **big data technologies** for improving the production of raw materials from agriculture, forestry and fishery for the bio-economy industry to produce food, energy and biomaterials in a responsible and sustainable way. DataBio takes advantage of recent innovative big data technologies applied in bio-economy sectors and aims to develop a big data platform on top of partners infrastructures and solutions. The technologies consist of 91 technological, mostly software, components provided by 27 partners including 37 datasets of Earth Observation and sensor data as well as 13 component pipelines. The 27 DataBio pilots have been classified into three categories:

- The agriculture pilots which aim to improve precision farming based on observational data and predictive analysis.
- The forestry pilots which aim to improve forest monitoring, predict risks and optimize tree resources.
- The fishery pilots which aim to improve vessel energy economy, logistic efficiency and predictive analysis of the fishery market as well as decreasing the environmental impact.

Taking into consideration the complexity of these tasks and the big number of heterogeneous technological components involved, we chose ArchiMate 3.0 [3] as an enterprise architecture modelling language in order to represent the components and the processes of the different pilots. These models can be useful for requirements elicitation, technical solutions design, interactions between stakeholders, facilitating communications between partners and for reporting results. This paper is structured as follows: section 2 presents the ArchiMate modelling approach in the context of DataBio project, section 3 illustrates the modelling of the pilots and Section 4 contains the modelling of the technological components used in DataBio project.

2   DataBio modelling approach with "ArchiMate 3.0"

The enterprise architecture methods are used for providing organizational structure, business processes and the IT infrastructure of an enterprise. These methods can be applied to DataBio context, where we are interested in modelling the case studies business objectives, processes, requirements, data pipelines and IT components. ArchiMate language has been proven to be particularly helpful for modelling organizations having complex IT infrastructures [4]. Moreover, the ArchiMate framework provides a wide range of modelling concepts which represent different layers of the enterprise such as: strategy, application, motivation, technology, business, etc [5]. The modelling environment used for this task is the Modelio ArchiMate modelling tool [6] which allows developers to collaborate on a synchronized remote version of the models. At the first stage of the project, first six months, the modelling process was focusing on the specification phase which meant to provide diagrams to specify the objectives, the requirements and the desired outcomes of the pilots. This helps to understand the big picture of the pilot, the guidelines and the interactions between the stakeholders. In the second stage of the modelling process, the technological design of the pilots became more mature and the partners started working on diagrams that illustrate the software components, their interactions and their deployment environment. The models are structured in five ArchiMate projects described as follows: three projects (agriculture, forestry and fishery) for pilot description, one project for modelling software and IoT system components and finally one project for modelling Earth Observation data services.

3   DataBio pilot models

DataBio pilots allow to experiment with different technologies on real case studies with the objective of analyzing the feasibility, the efficiency and the economic impact. The project's goal is to have a big data platform that integrates these technologies at the end of the project. We define this platform as an environment in which a combination of software components are developed to be deployed in hardware, virtualized infrastructure, operating system, middleware or a cloud. This environment provides through the DataBio Hub (http://databiohub.eu) [7], a big data toolset which offers functionalities primarily for services in agriculture, forestry and fishery. The functionalities enable new software components to be easily and effectively combined with open source, standards-based big data, and proprietary components and infrastructures based on generic and domain specific components.

The DataBio toolset supports the forming of reusable and deployable pipelines of interoperable and replaceable components, that can integrate the technologies adopted in the pilots. Each pilot has been modelled in the specification phase with diagrams that emphasize the motivation and strategy point of view. The motivation views provide elements that motivate the choices of the pilot such as objectives and requirements. The strategy views complete the motivation views by planning long term actions to meet the specified objectives. Some of the pilots went further in their design into providing additional models such as application views, data views and business processes.

3.1 Pilot motivation views

The pilot motivation views present the reasons and factors that justify and guide the pilot choices. This step is important for introducing the pilot and for explaining the relevance of the pilot concept. The motivation diagrams contain goals which specify the main objectives of the pilots. In relation to goals, these diagrams contain certain outcomes that realize the specified goals. Moreover, these views contain the stakeholders which are the individuals, teams and organizations involved in the pilot. There are also internal and external factors represented by "driver" elements which motivate an organization to define its goals and implement the changes necessary to achieve them. In addition to these concepts, we use "requirement" elements which are functionalities that need to be implemented and "constraint" elements which are factors that limit the realization of the defined goals.
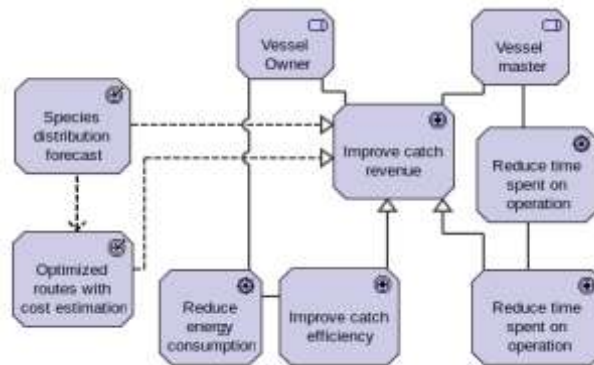

Fig. 1. Fishery pilot B1: "Oceanic tuna fisheries planning" Motivation View

An example of the motivation view can be found in the published presentation of the "tuna fisheries planning" pilot [8]. The purpose of this pilot is to improve profitability of tuna fisheries by saving fuel based on fish observation and vessels route optimization. Observe, that the catch volume cannot normally be increased as it is limited by quota. Therefore, the main goal of this pilot is "improving catch revenue". More specifically we want to "reduce energy consumption" and "improving catch efficiency" (see Fig. 1). In this pilot there are two major stakeholders: the "Vessel Owner" and "Vessel master" who both want to improve catch revenue. The vessel master is motivated by logistics on the ship and reducing the time spent on operations. Similarly, the vessel owner is concerned with reducing energy consumption and improving catch efficiency. These objectives are realized by optimizing route cost efficiency and species distribution forecasting.

3.2   Pilot strategy views

The pilot strategy views allow decision makers to elaborate a global roadmap to implement the announced objectives by defining course of actions for implementing the tasks at hand considering the capabilities of the organization and the available resources. For example, the strategy view of the fishery pilot mentioned in the previous section demonstrate how strategy elements provide leads for implementing the objective of "Improving the revenue of fish catching" (see Fig. 2). This objective can be realized by providing a decision support system on vessel operations. Given our capabilities in data collection and data analytics, we can develop the specified decision support system by collecting sensor data and extract useful information for decision makers.
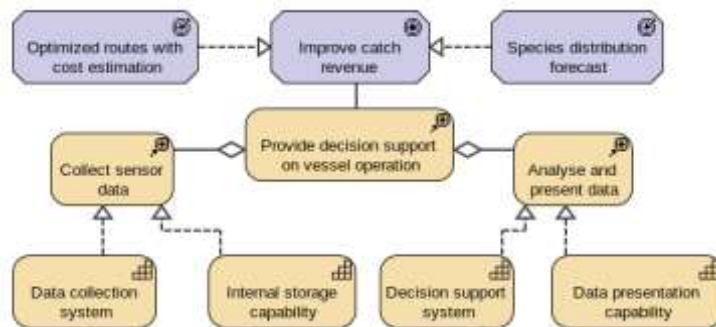

Fig. 2. Fishery pilot B1: "Oceanic tuna fisheries planning" Strategy View

4   Technological components models

DataBio pilots include various technological components with different interfaces, different data formats (sensors data, satellite imagery, etc) and different deployment environments. For each component, ArchiMate diagrams were created for modelling the interface view, the deployment view and the data view. The interface view shows the external interfaces of the component which are designed for interactions with users or with other components through various communication protocols. For example, the component **"C07.04: Data Manager"** is used for downloading and preprocessing earth observation data in forestry and fishery pilots. Its identification code **"C07.04"** is formed from component (=C), DataBio partner number (=07) and component number of that partner (=4). This component offers three interfaces for downloading data via Java API, REST API and command line interface (CLI). It also offers three interfaces for consuming data from other components (see Fig. 3) [9]. The deployment view describes the application executables and the software and physical environment required for running the application. The data view describes the format, the source and the content of the data processed by the component.
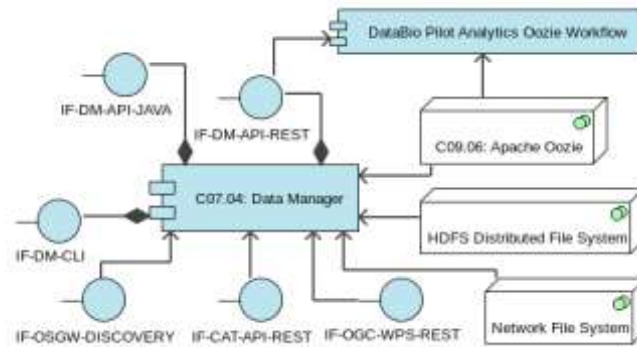
Fig. 3. Component C07.04: Data Manager Interface View
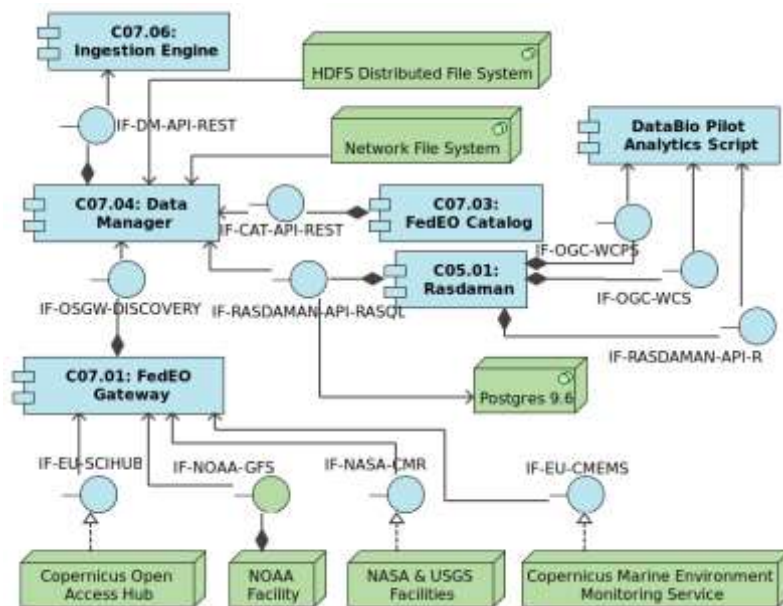

Fig. 4. Oceanic tuna fisheries planning - Pipeline View

In addition to components models, more ArchiMate diagrams were created in order to represent the data pipelines adopted by each pilot. As mentioned before, the different pilots are working on integrating several components into their workflow processes. These processes are referred to as "pipelines", as they integrate different tasks along the data value chain from data collecting to analysing and visualizing. The pipelines models represent the pilot lifecycle, the integrated components and the data flow between components. For example, the component "*C07.04: Data Manager*" mentioned previously is integrated in the pipeline of the fishery pilot "Oceanic tuna fisheries planning" (see Fig. 4) [10]. "*Data Manager*" acquires data from different components such as the component "*C07.01: FedEO Gateway*" which is a unique endpoint that retrieves geographical data from several backend providers such as "*Copernicus Open Access Hub*". Moreover, "*Data Manager*" stores data in network file systems or HDFS distributed file systems. Finally, the component "*C07.06*

Ingestion Engine" consumes the collected data by "*Data Manager*" via REST API interface.

5   Future work and concluding remarks

In this paper, we outlined the contribution of ArchiMate models in DataBio pilots specification and technological design. These models helped to provide productivity and clarity in the project by contributing to the analysis of the case studies and to the production of design documentation. This approach of ArchiMate modelling is currently being adopted by DataBio partners in national and European projects. On the other hand, based on our experience with this project, we have identified a potential for improvement with regard to both modelling process and model quality:

- Providing a holistic view of the project in addition to the pilot centered views and adding "Business Process" views to illustrate platform exploitation by end users.
- Ensuring component reuse in models to avoid duplication.
- Establishing quality metrics for "Enterprise Architecture" based on modelling experience from DataBio project and from other projects [11,12] and taking into consideration existing metrics described in the literature such as the "6C quality goals" described by Mohagheghi et al. [13].

Acknowledgments

References

1. DataBio Consortium, https://www.databio.eu/en/consortium. Accessed 18 Apr. 2019.
2. DataBio Summary,https://www.databio.eu/en/summary Accessed 18 Apr. 2019.
3. Josey, Andrew. ArchiMate® 3.0. 1-A pocket guide. Van Haren, 2017.
4. 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.
5. Desfray, P., & Raymond, G. (2018). TOGAF, Archimate, UML et BPMN-3e éd. Dunod.
6. Modelio products, https://www.modeliosoft.com/en/. Last accessed 18 Apr 2019.
7. DataBio Hub, https://www.databiohub.eu. Last accessed 18 Apr 2019.
8. DataBio public deliverable D3.1: Fishery Pilot Definition, v1.0, 2017/10/20.
9. DataBio public deliverable D5.1: EO Components Specifications, v1.0, 2017/12/29.
10. DataBio public deliverable D5.2: EO Components and Interfaces, v1.0, 2018/05/30.
11. Internet of Food & Farm Project 2020 (IoF2020) ,https://www.iof2020.eu . 18 Apr. 2019.
12. Big Data Benchmarking project (DataBench), https://www.databench.eu . 18 Apr. 2019.
13. 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.