Towards a Knowledge Graph Enhanced Automation and Collaboration Framework for Digital Twins

Christou, V.; Wang, Y.; Zhao, Z.

DOI
10.1109/e-Science58273.2023.10254845

Publication date
2023

Document Version
Author accepted manuscript

Published in
2023 IEEE 19th International Conference on e-Science

License
CC BY

Towards a Knowledge Graph-enhanced Digital Twin Framework for Automation and Collaboration

Vasileios Christou
Multiscale Networked System
University of Amsterdam
Amsterdam, The Netherlands
christou_vasileios@hotmail.com

Yuandou Wang
Multiscale Networked System
University of Amsterdam
Amsterdam, The Netherlands
y.wang8@uva.nl

Zhiming Zhao
Multiscale Networked System
University of Amsterdam
LifeWatch Virtual Lab and Innovation Center
Amsterdam, The Netherlands
z.zhao@uva.nl

***Abstract*—The Digital Twin (DT) is a digital representation of a physical process articulated alongside it, usually matching exactly the operation of the physical process in real-time. DTs of modern systems can be characterized by heterogeneity between their subsystems and stakeholders' backgrounds, factors that result in silos during the DT lifecycle. In the current study, we employ the Knowledge Graph modeling approach to boost automation and collaboration during the DT lifecycle stages. We implement our methods in developing a what-if analysis service for a DT of an ecosystem of wetlands and its automated deployment to the Amazon Web Services (AWS) cloud.**

***Index Terms*—Digital Twin, Semantic Web, Knowledge Graph, Ontology, Cloud Computing, What-If Analysis**

I. Introduction

The Digital Twin (DT) is as a virtual replica of a real-world system [1], with a bidirectional connection between the two parts. The DT lifecycle involves stakeholders with different backgrounds, such as domain experts, computational scientists, software developers, and end-users. Semantic-Web technologies [2] and the Knowledge Graph (KG) modeling approach have been widely used in DTs, due to their benefits in terms of interoperability and extensibility. However, most of the related work focuses on modeling the physical part, omitting crucial components, such as the workflow and the computing infrastructure.

The current study contributes towards the development of a platform that promotes automation and collaboration in the various stages of the DT lifecycle, transforming the information originating from the heterogeneous data sources and the involved stakeholders, to knowledge that is available to everyone, in a standard format. More specifically, we have developed and integrated ontologies to represent all DT layers, from the physical part, to the workflow within and between services, to the cloud infrastructure hosting them. We also showcase how the knowledge provided by the KG can provide the functionality for a collaborative what-if analysis platform with a User Interface (UI) to enhance end-user experience, and how it can be combined with cloud and Infrastructure as Code (IaC) technologies to automate the deployment stage[1].

This work has been partially funded by the European Union project CLARIFY (860627), ENVRI-FAIR (824068), BlueCloud-2026 (101094227) and LifeWatch ERIC.

[1]https://github.com/Vasilis421/DTCloudThesis

II. Related Work and Problem Statement

The World Avatar project has the vision of a universal DT of the world [3], and they have published work on the development of DTs in a national scale [4]. They have based their architecture on the KG, proving its value in cases where heterogeneous subsystems must be integrated. Additionally, in [5], they present a framework for what-if analysis, where the KG enables the simulation of parallel worlds next to the base scenario. However, they do not refer to the modeling of the computing infrastructure or to the importance of collaboration between the system's stakeholders. The main research question of this study is: *How can we model a cloud-hosted DT platform that promotes automation and collaboration during the whole DT lifecycle?*

III. Our Method and Implementation

A. Ontology Engineering

Our DT ontology consists of three main ontologies. The domain ontology (OntoDomain) includes information regarding the physical part, provided by domain experts. The workflow ontology (OntoFlow) describes the processes and services, as well as the data flow within and between them, enabling automation of workflows. It also includes the scenario ontology (OntoScenario), which enables a standardized what-if analysis service. Finally the infrastructure ontology (OntoInfra) describes the computing infrastructure, and includes the containerization ontology (OntoContainer) and the cloud deployment ontology (OntoDeploy).

B. System Architecture

Every component of the system's architecture should be autonomous and able to communicate with the rest. The Knowledge Base (KB) is the center of the architecture. A triplestore technology, Apache Jena Fuseki, provides a SPARQL endpoint that can be used by other components to query and update the KB. Additionally, we have developed the simulation and visualization services as autonomous containerized REST APIs that other services, such as the what-if analysis service, can call. This approach of decoupled services promotes scalability and and contributes in the extensibility requirement.

IV. EcoDTwin: A DT of an Ecosystem of Wetlands

EcoDTwin is a DT platform that enables the simulation, visualization, and sharing of scenarios in an ecosystem of wetlands in the northeastern Netherlands to predict the presence of bird species based on vegetation parameters [6]. Fig. 1 depicts how the KG enchances the DT lifecycle for EcoDTwin. Collaboration between stakeholders in the initial modelling phase is important for the encapsulation of the combined knowledge in ontologies. Knowledge retrieval and KG updates are achieved with SPARQL queries to the KB in all stages. The lifecycle is augmented with automation in processes such as data ingestion and IaC script generation. Additionally, the queried knowledge assists in the development and operation of DT services, such as the collaborative what-if analysis service.

A. *Ontology-Based Automated Data Ingestion*

Querying and mapping algorithms have been developed that can retrieve the knowledge encapsulated in ontologies, and use it to extract data from the various sources, transform them to RDF based on the ontology schema, and load them to the KG. This ontology-based ETL process is combined with technologies, such as cloud serverless functions, e.g. Amazon Web Services (AWS) Lambda, that are triggered whenever new data are available.

B. *Collaborative What-If Analysis*

OntoScenario enables the what-if analysis service. It also imports parts of the PROV-O ontology [7] to keep track of the provenance of scenarios, enabling sharing scenarios with other users and configuring scenarios based on existing ones to promote collaboration. The what-if analysis service UI is shown in Fig. 2. Querying the KB enables the rendering of standardized input forms so that the end-users can configure, execute, and share scenarios, utilizing the knowledge already encapsulated in the ontologies during the modeling phase.

C. *Automated Deployment*

OntoInfra captures information related to the containerization and the cloud resources required for the deployment of the DT platform. This information is used to generate Infrastructure as Code (IaC) templates, such as Terraform scripts, that automate the provision of cloud infrastructure. For instance, for the deployment of EcoDTwin on the AWS cloud, a container querying function retrieves information related to the configuration of the containers of each service, and uses this information to generate the Terraform script for the configuration of the Amazon Elastic Container Service (ECS).

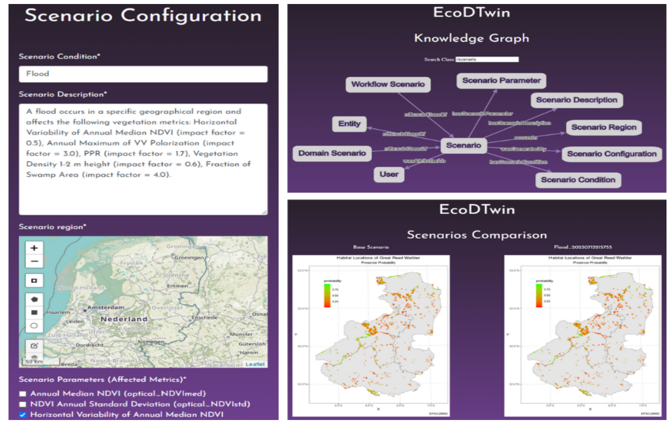

Fig. 2. What-if Analysis UI: Input forms are rendered based on OntoScenario. The map widget enables handpicking the scenario region, which is then transformed to RDF, based on OntoDomain. Scenario execution and visualization are performed by calling the respective API services.

V. Conclusion and Future Work

In the current study, we have employed a KG, to tackle the challenges of heterogeneity between the various subsystems of a DT and the system's stakeholders. Modelling accurate ontologies is crucial, since the standardized knowledge encapsulated in the KB, combined with technologies such as cloud computing and IaC, boosts automation and collaboration during the development, deployment, and operation stages. As a next step, we are aiming for the automation of all processes within the DT workflow and the provision of user friendly interfaces for all the DT services, based on the underlying ontologies. Furthermore, we are planning for the integration of EcoDTwin with a fire spreading model, to showcase the value of the KG approach in scenarios where multiple systems cooperate.

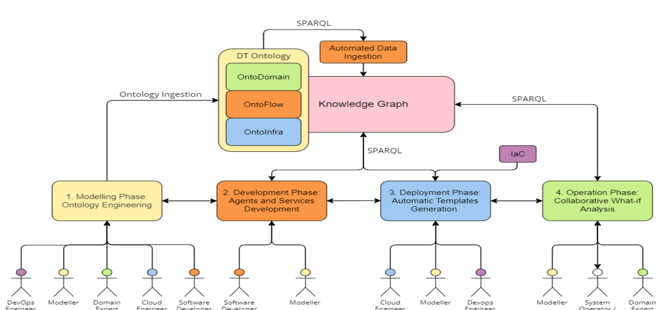

Fig. 1. KG enhanced DT lifecycle. Collaboration is important in every stage, and especially in the modelling phase to build accurate ontologies that can then be used to automate processes and enable services.

References

[1] M. Grieves, "Digital Twin: Manufacturing Excellence through virtual factory replication," March 2015.

[2] T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web," Scientific American, vol. 284, no. 5, pp. 34–43, 2001.

[3] J. Akroyd, S. Mosbach, A. Bhave, and M. Kraft, "Universal digital twin - a dynamic knowledge graph," Data-Centric Engineering, vol. 2, 2021. doi:10.1017/dce.2021.10

[4] J. Akroyd, S. Mosbach, A. Bhave, and M. Kraft, 11The national digital twin of the UK – a knowledge-graph approach," 2021.

[5] A. Eibeck, A. Chadzynski, M. Q. Lim, K. Aditya, L. Ong, A. Devanand, G. Karmakar, S. Mosbach, R. Lau, I. A. Karimi, E. Y. S. Foo, and M. Kraft, "A Parallel World Framework for scenario analysis in knowledge graphs," Data-Centric Engineering, vol. 1, p. e6, 2020.

[6] Z. Koma et al., "Better together? assessing different remote sensing products for predicting habitat suitability of wetland birds," Diversity and Distributions, vol. 28, no. 4, pp. 685–699, 2022. doi:10.1111/ddi.13468

[7] K. Belhajjame et al., "PROV-O: The PROV Ontology," 2012.