Received: 31 January 2022 / Accepted: 12 May 2022 / Published online: 17 May 2022

digital twins, manufacturing systems, machining 4.0, machine tools, intelligent functions

Arno WEISS^{1*} Steffen IHLENFELDT¹

INTEGRATION OF OPC UA INFORMATION MODELS INTO ENTERPRISE KNOWLEDGE GRAPHS

Building repositories of data relevant for enterprise operations requires harmonization of formats and semantics. OPC UA's nodes-and-references data model shares basic elements with well-established semantic modeling technologies like RDF. This paper suggests the use of transformed OPC UA information models on the higher level of Enterprise Knowledge Graphs. It proposes good practice to integrate the separate domains by representing OPC UA servers as RDF-graphs and subsequently attaching them to Digital Twins embedded in Enterprise Knowledge Graph structures. The developed practice is implemented, applied to combine a server's structure with an existing knowledge graph containing an Asset Administration Shell and released open source.

1. INTRODUCTION

OPC UA is established as the dominant standard for information modeling of industrial assets. It is well accepted and maintained by an ecosystem of industrial consortia and standards. Various communication mechanisms are designed to preserve the structure of variables, objects and data. Models for individual machine types and instances are built as so-called *nodesets* that rely on the common meta-model with nodes and references between the nodes. This graph structure is leveraged for the definition of industry-specific domain-models that are defined in *companion specifications* (OPC UA CS). This way, a certain consistency of models can be guaranteed between manufacturers and vendors of control equipment. The semantics are defined bottom-up so that orchestration systems such as Manufacturing Execution Systems (MES) or Industrial Internet of Things (IIoT) platforms can exploit not only the data but also the meta-data about the relationships between assets.

Semantic Interoperability denotes the ability of systems to operate exchanging information with unambiguous meaning. It is considered essential to meet central challenges in manufacturing [1]. The necessary digital representation of an asset, process or system that

¹Cyber-Physical Production Systems, Fraunhofer IWU, Chemnitz, Germany

^{*} E-mail: arno.weiss@iwu.fraunhofer.de

https://doi.org/10.36897/jme/150008

captures attributes with regard to communication, interpretation and processing is called a Digital Twin [2]. However, consistent OPC UA information models do not suffice to achieve true semantic interoperability between industrial assets. As visualized in Fig. 1, there is no semantic consistency between Digital Twins and an asset's near-hardware representation. The reason for this is twofold: Firstly, OPC UA models are designed for a very particular phase of the lifecycle – namely the usage phase that is preceded by the design phase with its own established standards. Secondly, OPC UA does not aim to model industrial processes as much as it models production resources. Aspects of material management or maintenance are neglected as they simply lie beyond the current scope of the standard.



Fig. 1. The Reference Architecture Model Industry 4.0 visualizes integration standards and challenges. The right chori-zontal axis denotes the hierarchy known from ISA-95/IEC 62243. The left horizontal life cycle axis is derived from IEC 62890. The vertical layers portray the abstraction from the physical assets up to business processes. This paper attempts to close the integration gap between the *Station* and *Work Centers* on the *Information*-layer during the *Instances* phase. Other standardization and integration gaps are neglected in this paper.

To provide a standard modeling procedure for this purpose, a variety of norms and specifications have been drafted and released into the industry with varying degrees of adoption. On the level of enterprise software, Digital Twins are designed to formally encapsulate all relevant facets of an industrial asset [2]. They shall enable enterprise interoperability between partners on the supply chain and maintain all relevant information in a unified manner. Many of them are built upon ontologies leveraging the Resource Description Framework (RDF) – a meta-model denoting a labelled, directed graph of nodes and references. This gives users the ability to query the model with the SPARQL query language and traverse the graph using standardized templates. Despite RDF often being used in conjunction with RDFS (RDF Schema) and OWL (Web Ontology Langauge), they are non-essential extensions that define standard building blocks for an RDF-graph enabling the usage of reasoners.

Enterprise Knowledge Graphs have an even broader scope, transgressing the industrial domain and interconnecting all concepts, properties, individuals and links representing knowledge relevant for an enterprise [3]. The focus for Knowledge Graphs in general lies on the RDF-based interconnection of objects from multiple sources with multiple schemas respectively [4].

Thus, integrating OPC UA models with the Digital Twins present in Enterprise Knowledge Graphs would connect two disparate models, each with distinct scopes, as portrayed in Fig. 1. It can provide the data owner with a more detailed and nuanced view of the reality on the shopfloor. Showcasing this benefit and providing a reproducible method to archive that goal is the purpose of this paper. Chapter 2 will first discuss previous academic and standardization work, followed by considered approaches to integration in Chapter 3 and an implementation showcase in Chapter 4.

2. RELATED WORK

2.1. BUILDING GRAPHS FROM OPC UA INFORMATION MODELS

Nodes and references between them are the core of any graph-based data structure. Such a reference requires at least a source node (subject), the relation's type (predicate) and a target node (object). This structure is present in OPC UA nodesets as well as RDF-graphs. Dispite this similarity, nodesets are serialized in a specific XML-format that does not build on RDF. Thus, a mapping of both meta-models is necessary but feasible. In fact, a complete formal mapping was first described by Schiekofer et al [5] building upon OWL. This required a matching of OPC UA concepts with conceptually equivalent counterparts in the semantic web stack. OPC UA models could now be used in conjunction with the RDF/OWL toolbox to scrutinize them with regard to consistency and validity. The authors also showed how SPARQL can be used as a capable query language for nodesets [6]. Perzylo et al. [7] published a repository of ontologies along the entire scope of OPC UA. It shows the applicability of the mapping concept for the core building blocks of the standard as well as domain- and even application-specific nodesets.

These works laid the foundation for a variety of applications for OWL-transformed nodesets: Selecting aspects from transformed OPC UA nodesets via SPARQL queries, it became feasible to automatically generate domain-specific ontologies from OPC UA information models directly that could then be instantiated [8]. A relevant contribution of this work is the public release of a GPL-3-licenced Java library to automate the transformation by browsing through a live server.

SPARQL natively only supports querying of graph data, which is insufficient since manufacturing data also varies over time. To compensate for this shortcoming, Bakken [9] introduced an engine with a SPARQL-dialect that also gives access to time-series data and implements the model transformation to OWL in Python (Apache 2 License). Despite the relevant research progress, there has been no completed effort to specify a mapping from OPC UA to RDF from any relevant standardization body.

2.2. RELEVANT STANDARDS FOR DIGITAL TWINS IN ENTERPRISE KNOWLEDGE GRAPHS

As stated in the introduction, Enterprise Knowledge Graphs are comprehensive RDFgraphs. They serve as a network of entities, their semantic types, properties and relationships between entities [10]. As such, physical objects like production resources must also be portrayed.

The potential of RDF-based Digital Twin models to foster interoperability in the industrial space has been first described briefly after the technology's inception. Borgo [11] envisioned manufacturing taxonomies based on the semantic structure of foundational ontologies built on top of schema and ontology languages like RDFS/OWL. Analyzing the problem of incompatible data models in the supply chain and lifecycle, they derive a necessity for unified terminology and a formal description. Their suggested ADACOR model covers for instance materials, work orders, products and 16 other classes on the same abstraction level that are interconnected by a limited set of relationships. However, the challenge persisted. Ontologies with a similar abstraction level were suggested for the manufacturing domain by academic works [12] and in 2011, by release of a standard called "Universal Machine Connectivity for MES" (UMCM) [13], industrial consortia started suggesting frameworks modeled specifically for the discrete manufacturing domain.

Following the release of the Reference Architecture Model Industry 4.0, the German Platform I4.0 released a specification for the Asset Administration Shell (AAS) that is supposed to provide standardized modeling capabilities under a common meta-model [14]. Even though the data model at its core was initially not an ontology [15], the specification contains serialization rules as an RDF-graph with RDFS annotations. This allows to consider the AAS an industrial domain-ontology with a well-developed ecosystem, a sizable community and reliable tool support. It integrates taxonomies such as ECLASS and IEC CDD and strives to organize domain-models in so-called submodel templates.

The AAS is one of the works that Jacoby and Usländer [16] considered in their analysis of the current landscape for IoT Digital Twin standards. Their analysis however shows that of the six works considered, four support RDF as a serialization format while of the selected only OData and the SensorThings API do not. All six support JSON which will not be considered in the following analysis due to its data model relying on key-value pairs instead of graph-spanning triples suitable for a Knowledge Graph.

3. BRINGING OPC UA TO THE WORLD OF DIGITAL TWINS

3.1. PROBLEM STATEMENT

OPC UA CS – and models built from them – are attractive for models on higher levels of the automation stack. They cover a wide field of domains and have communities maintaining them. That is why (wherever possible) Digital Twin standards could profit from integrating these device-level domain-models into their architecture instead of executing the tedious work of remodeling. Transformation is a first step to make the structure available to higher-level enterprise software. Still, mapping an OPC UA AddressSpace (either crawled from the server or parsed from a nodeset file) to an RDF-representation does not suffice to integrate it with the mentioned standards for Digital Twins on the process control level. The authors have identified two challenges with this current state of technology.

The first challenge is visualized in Fig. 2. RDF/OWL just cover the MOF (Meta-Object Facility) M3 level. Consequently, the resulting RDF-graph still builds upon resources whose semantics stem from the OPC UA specification. This is portrayed on level M2 of the MOF metadata architecture: Every node still inherits its structure from one of the eight NodeClasses. The complex or primitive data types still adhere to the definition of OPC UA DataTypes. All these concepts may or may not have corresponding semantic equivalents in other meta-models. But even if that were the case, these mappings would have to be specified manually for each pairing of standards.



Fig. 2. Remaining incompatibility after UA-RDF meta-model transformation displayed in the MOF metadata architecture [17]

Secondly, OPC UA transformation to RDF exclusively considers the AddressSpace of the server. This is problematic because the relevant runtime aspects required for managing the assets in production remain unconsidered. As the AddressSpace just describes the nodes and their relations, deployment-specific properties are not a part of it. For planning and operations, it is critical to include the server's security settings or discovery-URL among other properties. As previous works mostly focused on validation and querying as self-serving purposes, the holistic display of an OPC UA server beyond its nodes was not in scope.

To the author's knowledge, these two challenges have not been solved for integration into any of the relevant Digital Twin standards or Enterprise Knowledge Graphs in general. This chapter outlines two possible solutions to the first challenge of incompatible MOF-M2-models. The more practical second challenge of enhancing OPC UA ontologies with deployment data will be solved subsequently (see Chapter 4.1).

3.2. OPTION A: TRANSFORMING OF OPC UA NODESET STRUCTURES TO DIGITAL TWIN META-MODELS

One feasible approach is to extend the mapping from the M3 level down to the M2 providing representations purely built upon a Digital Twin standard such as discussed in Chapter 2.2. This would transform the bottom-up semantic structure defined in the OPC UA

CS and instance to a semantic Digital Twin compatible with the target meta-model. The model would hold the content of the OPC UA AddressSpace, stripped of the meta-model's terminology such as *ObjectType*, *hasTypeDefinition* or *NodeId*. Integration with other assets that were modeled in the Digital Twin standard natively would be seamless. Data could even be parsed as instances of the Digital Twins structure and processed with unambiguous meaning from there. This approach assumes that all systems consuming the modeled data require adherence to the single core meta-model.

However, there are several drawbacks to this approach: The first obvious drawback is lacking clarity what Digital Twin standard shall be selected. Decisions on this matter are usually dependent on the use case domain, requirements from supply chain partners and preexisting models. Thus, one M2-level mapping would be required for each standard. Missing specification of the mapping and integration process would hinder the practical integration even more. Even the UA CS called "OPC UA for Asset Administration Shell" [18] does not address the challenge of integrating servers into the AAS as it only specifies how to serialize AAS as OPC UA nodesets. Parsing them is only possible when the UA structure strictly adheres to the types of this CS and none other. Even the approach to specify a mapping is impractical (perhaps even publishing it) since it would require extensive prior knowledge of the source- and target meta-models. It also assumes that a complete mapping is possible despite the differences in expressiveness and scope between the formats, meaning that the mapping could only cover the conceptual intersection that the two standards share.

3.3. OPTION B: APPENDING OPC UA RDF TO DIGITAL TWINS AS DOMAIN-SPECIFIC SUBGRAPHS

RDF was built upon the idea to connect resources from different ontologies together. In the "The Four Rules" for best practice data sharing [19], it is explicitly stated that drawing interconnecting links is essential to finding related data on the web. This was reiterated in the FAIR Data Principles where requirement I3 states that (meta)data should "include qualified references to other (meta)data" [20]. These guidelines motivate adherence to the basic idea of RDF – to consider all semantic data (conformant to standards or not) as part of a heterogenous knowledge graph. Apart from addressing most of the drawbacks mentioned in Chapter 3.2, this approach has one significant advantage: By not committing to a M2-meta-model and connecting data purely based on the shared foundation in RDF, enterprise architects gain significant flexibility. Considering that enterprise data is diverse, the problem of aligning data with a common meta-model will recur every time a new data source is connected. This problem is avoided by linking data so that one graph emerges without meta-model mapping. Concerns stay separated and each domain preserves their own meta-model in the knowledge graph. SPARQL becomes a unified query language that transcends M2-models.

Like before though, users must be mindful of the consequences this approach entails. If the company or community is stringent in using a single dominating standard for its Digital Twins, this approach could render the RDF-based OPC UA structure useless as tools built expecting meta-model conformance would not be able to parse the data. It also requires that the organisation is already on a path to managing all relevant metadata in graphs and thus utilizes the RDF-serialization instead of, say simple JSON or XML. As this Option B is less

complex and still yields many of the advantages of tight integration (Option A), the following chapter will focus on integrating transformed nodesets that still adhere to the OPC UA meta-model.

4. IMPLEMENTATION

4.1. MODELING A HOLISTIC REPRESENTATION OF OPC UA SERVERS

With the first challenge from the problem statement (Chapter 3.1) addressed, the second issue of incomplete representation persist as the current concepts for implementations of OPC UA to RDF transformation focus on the AddressSpace that lacks meta-data required for asset management at runtime. Drawing inspiration from other standards, adding the required metadata when offering the data on a network seems reasonable. This concept is exemplified by the *Descriptor* and *Endpoint* concepts in the AAS [21]. Because the standard OPC UA nodeset (*ModelUri http://opcfoundation.org/UA/*) includes data structures that are used for communication rather than object modeling, many of the missing elements that describe the server itself can be found there and used out of the box from the transformed standard nodeset. The following three data structures were identified as missing from OPC UA models in RDF this far but require consideration:

- The *Server* object is located beneath the *Objects* folder on every running server. When traversing a server with *Browse* requests before RDF-transformation, all this information will be present in the resulting ontology. It also holds all objects based on the *PublishSubscribeType* that indicates the server's support of the communication mechanisms of OPC 10000-14. However, the *Server* object is not usually present in nodeset.xml files as its values are dependent on the configuration and stack of the running server. This poses a problem for transformation based simply on the nodeset file.
- The DiscoveryService *FindServers* returns a data structure of DataType *Application Description* which holds a localized description of the application running as well as information on what Discovery profiles the server supports and (most importantly) the *discoveryUrl*.
- The DiscoveryService GetEndpoints yields a structure of DataType Endpoint Descripiton that contains information about the endpoints that a server offers to clients including the securityPolicy, and transportProfileUri. This structure contains a server field that is of type ApplicationDescription.

These information are critical to managing the device that the OPC UA server is running on. They are indispensable for a holistic representation in higher-level enterprise information systems.

After having identified the *Server Object* and the *EndpointDescription* from the *DiscoveryServices* as relevant data that existing transformations must be appended with for a useful representation in Digital Twins, the integration of the ontologies shall be operationalized by a software tool executing the transformation from OPC UA to RDF and connecting the models.

On the side of OPC UA however, the decision was made to restrict the range to a pure RDF representation, thus omitting RDFS and OWL. This decision was made to preserve the original semantics as much as possible and to enable users to query the graph without prior knowledge of ontology languages.

4.2. IMPLEMENTING AN ENTERPRISE KNOWLEDGE GRAPH OF OPC UA ENHANCED DIGITAL TWINS

The established approach handles these challenges by not restricting usage to a particular ontology on the side of the Digital Twin. A review of Digital Twin standards commonly used in Enterprise Knowledge Graphs (Chapter 2.2) has identified several candidates that could benefit from transformed OPC UA instance models. The approach shall be validated on a Digital Twin modeled as an AAS. Due to its focus on industrial assets and growing proliferation, it is well suited to serve as a testbed. In case of the AAS, instances of the *AssetAdministrationShell* class seem appropriate as an adaption point in the Digital Twin since they are the core elements that the specification is built around. The structure beneath the AAS object such as *Submodels* or *AssetInformation* will remain intact and coexist with the appended OPC UA model. While an *Asset* instance could be a good fit as well as it describes the production asset itself, modeling it in the AAS meta-model is optional. To illustrate, Fig. 3 beneath displays an example query that can be used to retrieve all AAS objects with links to an OPC UA AddressSpace objects.

Fig. 3. Example query to retrieve all AssetAdministrationShell objects that have a transformed OPC UA server representation attached. The *rdfs* namespace is used only in context of the AAS standard and the connection between the AAS and OPC UA is realized via a domain-ontology

5. CONCLUSION

In this work, a good practice to integrate OPC UA servers into a generic Enterprise knowledge graph was presented. Leveraging the RDF-transformation of the *AddressSpace* and appending it with data specific to a server deployment provides a holistic picture for higher-level systems to consume. The choice of Chapter 3.3 to integrate models with heterogeneous meta-models is a pragmatic decision to enable an easy-to-implement

integration of the disparate ontologies. Both showcases demonstrate this as they enhance the Digital Twin model with the semantics already present on the field level.

However, users must be aware of the overall information architecture in their organization as the preferred pure-RDF connection requires an Enterprise Knowledge Graph to exist in the first place. That is why further research on meta-model level integration (Chapter 3.2) is critical, especially for manufacturing-focused models like the AAS. The uncertainty concerning the long-term support of Digital Twin meta-models however makes the initial alignment effort risky. Chapter 2.2 has only mentioned only a small subset of models that were suggested over the multiple decades of research in this area, most of which have failed to proliferate. Thus, the proposed RDF-level integration of OPC UA can serve as a flexible, intermediate compromise between applicability and semantic rigor.

ACKNOWLEDGEMENTS

The work in this paper was partly supported by the "Fraunhofer-Gesellschaft" by funding on the basis of the "Leitprojekt EVOLOPRO".

REFERENCES

- NILSSON J., SANDIN F., 2018, Semantic Interoperability in Industry 4.0: Survey of Recent Developments and Outlook, Proceedings IEEE 16th International Conference on Industrial Informatics (INDIN), Faculty of Engineering of the University of Porto, Portugal, 18–20 July 2018, IEEE, Piscataway, NJ, 127–132.
- [2] Industrial Internet Consortium, 2020, Digital Twins for Industrial Applications: Definition, Buisiness Values, Design Aspects, Standards and Use Cases, White Paper, https://www.iiconsortium.org/pdf/IIC_Digital_Twins_ Industrial_Apps_White_Paper_2020-02-18.pdf., Accessed 12 Aug 2021.
- [3] GALKIN M., AUER S., VIDAL M-E., et al., 2017, Enterprise Knowledge Graphs: A Semantic Approach for Knowledge Management in the Next Generation of Enterprise Information Systems, Hammoudi S., Smialek M., Camp O et al. (eds), Proceedings of the 19th International Conference on Enterprise Information Systems, 1, 26–29 April, Porto, Portugal, SCITEPRESS – Science and Technology Publications, 88–98.
- [4] PAULHEIM H., 2016, *Knowledge Graph Refinement: A Survey of Approaches and Evaluation Methods*, SW 8, 489–508, https://doi.org/10.3233/SW-160218.
- [5] SCHIEKOFER R., GRIMM S., BRANDT M.M., et al., 2019, *A Formal Mapping Between OPC UA and the Semantic Web*, IEEE 17th International Conference on Industrial Informatics (INDIN), Aalto University, Helsinki-Espoo, Finland, 22–25 July, Proceedings, IEEE, Piscataway, NJ, 33–40.
- [6] SCHIEKOFER R., WEYRICH M., 2019, *Querying OPC UA Information Models with SPARQL*, 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), IEEE, 208–215.
- [7] PERZYLO A., PROFANTER S., RICKERT M., et al., 2019, *OPC UA NodeSet Ontologies as a Pillar of Representing Semantic Digital Twins of Manufacturing Resources*, 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), IEEE, 1085–1092.
- [8] STEINDL G., KASTNER W., 2021, Transforming OPC UA Information Models into Domain-Specific Ontologies, 4th IEEE International Conference on Industrial Cyber-Physical Systems (ICPS), Online, 10–13 May, IEEE, Piscataway, NJ, 191–196.
- [9] BAKKEN M., 2021, Quarry: An Open Source Tool for OPC UA SPARQL Queries Over Hybrid Architectures Using Query Rewriting, 26th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), IEEE, 1–7.
- [10] EHRLINGER L., WOLFRAM W., 2016, Towards a Definition of Knowledge Graphs, SEMANTICS 2016, Posters and Demos Track, September 13–14, Leipzig, Germany.
- [11] BORGO S., LEITAO P., 2004, *The Role of Foundational Ontologies in Manufacturing Domain Applications*, Hutchison D., Kanade T., Kittler J., et al. (eds), On the Move to Meaningful Internet Systems 2004, CoopIS, DOA, and ODBASE, 3290, Springer Berlin Heidelberg, 670–688.

- [12] LEMAIGNAN S., SIADAT A., DANTAN J-Y., et al., 2006, MASON: A Proposal for an Ontology of Manufacturing Domain, Marik V. (ed), Proceedings DIS 2006, IEEE Workshop on Distributed Intelligent Systems, Collective Intelligence and its Applications, June 15–16, Prague, Czech Republic, IEEE, Piscataway, NJ, 195–200.
- [13] Verein Deutscher Ingenieure, 2011, Manufacturing Execution Systems (MES): Logic Interfaces for Machine and Plant Control, ICS 35.240.50 (VDI 5600, Sheet 3), Accessed 10 Oct 2021.
- [14] Plattform Industrie 4.0, 2020, Details of the Asset Administration Shell Part 1: The Exchanges of Information Between Partners in the Value Chain of Industrie 4.0, https://www.plattformi40.de/IP/Redaktion/EN/Down loads/Publikation/Details_of_the_Asset_Administration_Shell_Part1_V3.html, Accessed 04 Dec 2021.
- [15] BADER S.R., MALESHKOVA M., 2019, *The Semantic Asset Administration Shell*, Acosta Deibe M., Cudre-Mauroux P., Maleshkova M., et al. (eds), Semantic Systems: The Power of AI and Knowledge Graphs: 15th International Conference, SEMANTICS 2019, Karlsruhe, Germany, September 9–12, Proceedings, 11702, Springer Open, Cham, Switzerland, 159–174.
- [16] JACOBY M., USLÄNDER T., 2020, Digital Twin and Internet of Things Current Standards Landscape, Applied Sciences, 10, 6519, https://doi.org/10.3390/app10186519.
- [17] Object Management Group, 2016, Meta Object Facility Specification: Version 2.5.1, https://www.omg.org/spec/ MOF/, Accessed 06 Dec 2021.
- [18] OPC Foundation, 2021, OPC UA for Asset Administration Shell (AAS) OPC 30270, https://reference.opcfoundation. org/I4AAS/docs/. Accessed 06 Dec 2021.
- [19] BERNERS-LEE T., 2010, *Linked Data Design Issues*, https://www.w3.org/-DesignIssues/LinkedData.html, Accessed 04 Dec 2021.
- [20] WILKINSON M., DUMONTIER M., AALBERSBERG I., et al., 2016, *The FAIR Guiding Principles for Scientific Data Management and Stewardship*, Sci Data 3, 160018, https://doi.org/10.1038/sdata.2016.18.
- [21] Plattform Industrie 4.0, 2021, Details of the Asset Administration Shell Part 2: Interoperability at Runtime Exchanging Information via Application Programming Interfaces, Version 1.0RC02, https://www.plattformi40.de/IP/Redaktion/EN-/Downloads/Publikation/Details_of_the_Asset_Administration_Shell_Part2_V1.html, Accessed 25 Nov 2021.