

Semantic Data Integration for Industry 4.0 Standards

Irlán Grangel-González^(✉)

Enterprise Information Systems (EIS) Group,
University of Bonn and Fraunhofer (IAIS), Sankt Augustin, Germany
`grangel@cs.uni-bonn.de`

Abstract. Industry 4.0 initiatives have fostered the definition of different standards, e.g., AutomationML or OPC UA, allowing for the specification of industrial objects and for machine-to-machine communication in Smart Factories. Albeit facilitating interoperability at different steps of the production life-cycle, the information models generated from these standards are not semantically defined, making the semantic data integration a challenging problem. We tackle the problems of integrating data from documents specified either using the same or different Industry 4.0 standards, and propose a rule-based framework that combines deductive databases and Semantic Web technologies to effectively solve these problems. As a proof-of-concept, we have developed a Datalog-based representation for AutomationML documents, and a set of rules for identifying semantic heterogeneity problems among these documents. We have empirically evaluated our proposed framework against several benchmarks and the initial results suggest that exploiting deductive and Semantic Web techniques allows for increasing scalability, efficiency, and coherence of models for Industry 4.0 manufacturing environments.

1 Problem Statement

The Industry 4.0 vision aims at creating Smart Factories by combining the Internet of Things, Internet of Services, and Cyber-Physical Systems. To support this vision, Industry 4.0 communities have fostered the definition of standards such as AutomationML (IEC 62424) and OPC UA (IEC 62541). AutomationML is one of the core standards of Industry 4.0 for exchanging plant engineering information as specified by [1, 7, 16, 18]. AutomationML can describe plant components and their sub-components from different views such as mechanical, electrical, or software. OPC UA [4] also allows for the description of the production life-cycle in Smart Factories, but contrary to AutomationML which describes characteristics of plant components, OPC UA models machine-to-machine communication. Although Industry 4.0 standards provide the basis for data exchange in a Smart Factory, information models of these standards require being aligned to facilitate the merging of a virtual process with real production life-cycles.

Smart Factories along with Cyber-Physical concepts impose new challenges to traditional approaches of data integration. A new generation of data-centric

systems need to be developed and integrated, where data meaning, as well as data variety, veracity, and adaptivity must be managed [12]. In this context, achieving semantic data interoperability techniques suitable for these data properties, is of paramount importance for making the Industry 4.0 vision a reality.

With the aim of assessing the integration of Industry 4.0 standards, Biffi et al. [1] and Kovalenko and Euzenat [9] have characterized seven semantic heterogeneity issues among different views of an artifact. **(M1)** Value processing—same properties are not modeled equally, e.g., using different datatypes; **(M2)** Granularity—same objects are modeled at different levels of detail; **(M3)** Schematic differences—differences in the way how semantics is represented for the same object; **(M4)** Conditional mappings—relations between entities exist only if certain conditions occur; **(M5)** Bidirectional mappings—relations between entities have to be defined bidirectionally; **(M6)** Grouping and aggregation—different semantic modeling criteria are applied to group elements for the same object; and **(M7)** Restrictions on values—mandatory values for properties in the object that have to be handled in the mapping process. These semantic heterogeneity issues may occur between documents defined in the same or different standards, i.e., interoperability can be *intra-* or *inter-standard*.

Industry 4.0 standards and initiatives for identifying heterogeneity issues evidence the success of the Industry 4.0 movements. However, integration of standards are still conducted manually [5, 17], negatively affecting the effectiveness of the production process. This doctoral work attempts to achieve two research goals to solve these problems. **RG1:** Addressing *intra-standard interoperability* among multiple pieces described in one standard, e.g., AutomationML. **RG2:** Assessing *inter-standard interoperability* in documents specified in different standards, e.g., documents in AutomationML and OPC UA.

To accomplish our research goals **RG1** and **RG2**, we propose a rule-based system that combines Deductive databases and Semantic Web technologies to effectively integrate documents specified in Industry 4.0 Standards.

2 State of the Art

Related work is divided into two sub-sections: (1) approaches for solving *intra-standard interoperability* with the focus in AutomationML; and (2) approaches for addressing *inter-standard interoperability* on AutomationML and OPC UA.

Intra-standard interoperability issues for AutomationML. In the literature, many different approaches are proposed for integrating AutomationML documents. In [17], a tool to map two AutomationML files is presented. It allows for the integration of AutomationML documents, their respective descriptions, and the modified parts of one file into the other. Further, a mapping algorithm for AutomationML files is presented. Nevertheless, the process of mapping is performed manually. Himmler [8] presents a framework to create standardized application interfaces in plant engineering based on AutomationML. The work provides a function-based based standardization framework for the plant engineering domain. Persson et al. [13] utilize an RDF-based approach to integrate

robotized production information modeled AutomationML. Kovalenko et al. [10] explore how AutomationML can be represented by means of Model-Driven Engineering and the Semantic Web. A small part of an AutomationML ontology is developed, based on the main concepts of the language. Also, the use of rules for consistency checking is proposed, using the Semantic Web Rule Language (SWRL), but no explicit definition of the role of Semantic Web technologies on the integration problem is presented. The *AutomationML Analyzer* [14] is an on-line tool to browse, query and analyse different AutomationML data by means of Semantic Web technologies. A conceptual design to overcome integration problems in AutomationML is described. All these approaches have the potential to solve specific integration problems for AutomationML. However, they solve rather isolated problems, and a general method capable to automatically integrate AutomationML information from different perspectives is not provided.

Inter-standard interoperability issues for AutomationML and OPC UA. The current integration approach is performed by analyzing common information elements and manually describing the AutomationML objects in the OPC UA language [5, 7, 15]. In addition, in these works are documented the mappings between the common objects, i.e., their common structures and datatypes. To date, the possibility to semi-automatically integrate the information models of these two languages exploiting their semantic descriptions is still missing.

3 Proposed Approach

We propose ALLIGATOR [6], a rule-based framework for the semantic integration of Industry 4.0 standard documents. ALLIGATOR relies on Datalog and RDF to *accurately* represent the knowledge that characterizes different types of semantic heterogeneity for these documents. Further, by utilizing Datalog as well as Semantic Technologies, ALLIGATOR will be able to provide explanations for the alignments that occurred among the elements of the Industry 4.0 standard documents.

Figure 1 depicts the main components of the proposed solution. In the following we describe each component. The input for ALLIGATOR are Industry 4.0 standard documents. Next, these documents are translated into a canonical representation of a production life-cycle named the ALLIGATOR data model. RDF, RDFS, and domain-specific vocabularies are used to represent the core concepts of the ALLIGATOR data model. Further, standards documents are modeled as facts in an extensional database (EDB) of a Datalog program. Datalog rules comprise an intensional database (IDB), and state different types of semantic heterogeneity and *intra-* and *inter-standard interoperability* problems. The ALLIGATOR *Deductive System Engine* performs a bottom-up evaluation of ALLIGATOR Datalog programs following a semi-naïve algorithm that stops when the least fixed-point is reached [2]. The intensional predicates inferred in the evaluation of a ALLIGATOR Datalog program correspond to problems of semantic heterogeneity, and *intra-* and *inter-standard interoperability*. AutomationML, which

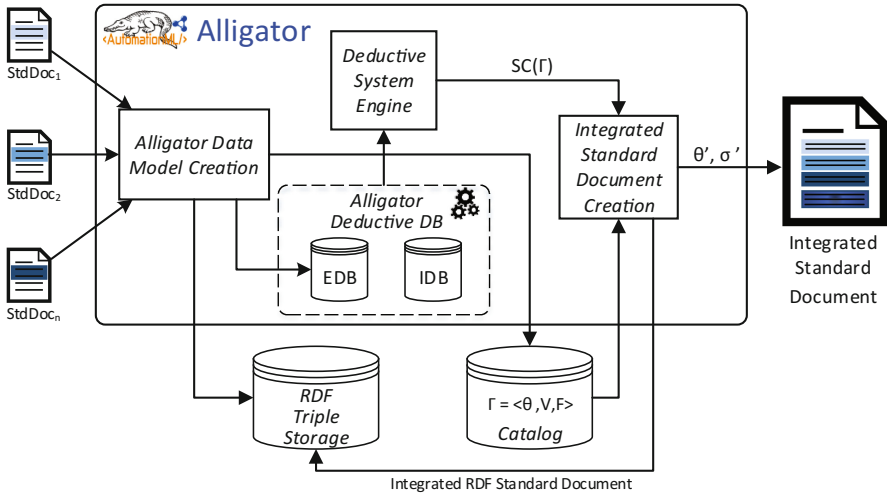


Fig. 1. The Alligator Architecture. ALLIGATOR receives documents in different Industry 4.0. standards, and creates an integrated document. Input documents are represented as RDF graphs and Datalog predicates (EDB) in the ALLIGATOR data model. Datalog intentional rules (IDB) characterize semantic heterogeneity types. A bottom-up evaluation of the Datalog program identifies semantic heterogeneity inconsistencies among input documents

is an XML-based standard, is translated into RDF using Krexlor [11], an XSLT-based framework for converting XML to RDF. In addition, the mapping rules for the conversion using Krexlor have to be created according to the AutomationML vocabulary, and an RDFS vocabulary¹ and the mapping rules² for the AutomationML standard have been defined. Similarly, we plan to define more rules to cover other standards and to semantically integrate them.

We present an example of one of the designed Datalog rules that states when two AutomationML elements are the same. Based on the AutomationML specification, this condition is met depending on links with external standards such as eCI@ss [3], which semantically identify elements. Accordingly, if two elements contains the same eClassIRDI as a value, then they are semantically equivalent. Listing 1.1 is a Datalog rule that represents this knowledge.

Listing 1.1. Rule 1.1: Semantic equivalence of two eClassIRDI AML attributes

```

sameEClassIRDI(X,Y) :- hasAttributeName(X, 'eClassIRDI') &
                        hasAttributeName(Y, 'eClassIRDI') &
                        hasAttributeValue(X,Z) &
                        hasAttributeValue(Y,Z).
    
```

¹ <https://w3id.org/i40/aml/>.

² <https://raw.githubusercontent.com/EIS-Bonn/krexlor/master/src/xslt/extract/aml.xml>.

4 Results

Testbeds. With the aim of testing the effectiveness of ALLIGATOR, we generated 30 Testbeds for *intra-standard interoperability* issues focusing in AutomationML. Testbeds are based on the semantic heterogeneity types M2 (granularity), M3 (schematic differences), and M6 (grouping and aggregation); ten testbeds per each type of heterogeneity. First, a seed (AutomationML document) was manually created for each testbed according to the type of semantic mapping. Next, we automatically generated two AutomationML documents derived from this seed containing a random number of semantic equivalent AutomationML elements³. The generation was performed following a uniform distribution. Testbeds corresponded to pairs of AutomationML documents, and thirty testbeds were evaluated in the study⁴. Further, a Gold Standard was manually generated computing the elements that were semantically equivalent as well as those different ones. For the compilation of a Gold Standard, we relied on the generated testbeds.

Table 1 reports on the values of these metrics for each type of semantic heterogeneity, i.e., M2, M3, and M6. We observed that for these semantic heterogeneity types, the value for precision is 1.0, i.e., ALLIGATOR correctly detected all the

Table 1. Effectiveness of Alligator. Per semantic heterogeneity type, effectiveness of ALLIGATOR is reported. In all testbeds, precision is 1.0. ALLIGATOR exhibits the highest performance in the testbeds of type M2 (F-measure is always 1.0), while in M3 and M6, the F-measure values are at least 0.8.

Granularity (M2)										
	TB1	TB2	TB3	TB4	TB5	TB6	TB7	TB8	TB9	TB10
Precision	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Recall	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
F-measure	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Schematic (M3)										
	TB1	TB2	TB3	TB4	TB5	TB6	TB7	TB8	TB9	TB10
Precision	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Recall	1.0	1.0	1.0	1.0	1.0	1.0	0.83	1.0	0.88	0.75
F-measure	1.0	1.0	1.0	1.0	1.0	1.0	0.90	1.0	0.94	0.85
Grouping (M6)										
	TB1	TB2	TB3	TB4	TB5	TB6	TB7	TB8	TB9	TB10
Precision	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Recall	1.0	1.0	1.0	0.66	1.0	1.0	1.0	1.0	1.0	0.83
F-measure	1.0	1.0	1.0	0.80	1.0	1.0	1.0	1.0	1.0	0.90

³ <https://github.com/i40-Tools/AMLGoldStandardGenerator>.

⁴ <https://github.com/i40-Tools/HeterogeneityExampleData>.

semantically equivalent elements. Further, recall and F-measure are also 1.0 in the testbeds of semantic heterogeneity M2. These results suggest that ALLIGATOR rules capture the knowledge required to *accurately* solve the *AutomationML semantic equivalent elements Identification* problem. For the semantic heterogeneity types M3 and M6, ALLIGATOR rules are not completely covering all possible semantic equivalences generated between *nested structures* of AutomationML elements. Thus, ALLIGATOR could not identify at most two semantic equivalent elements in five out of 20 testbeds of type M3 and M6.

5 Methodology

The methodology adopted in this doctoral work comprises the following tasks:

1. Investigation of state-of-the-art approaches relevant to the problem of integrating Industry 4.0 standards.
2. Formalization of the problem of integrating Industry 4.0 standards and proposed solutions; definition of research questions and hypotheses of our formal and empirical study are stated. ALLIGATOR is designed under the following *hypothesis*: an approach combining Datalog rules and Semantic Web technologies, for the integration of Industry 4.0 standards will exhibit better performance than the state of the art approaches. In addition, we identified the following research questions: **(RQ1)** Is ALLIGATOR able to identify pairs of semantic equivalent elements in Industry 4.0 documents? **(RQ2)** Does ALLIGATOR exhibit equal behavior whenever different types of semantic heterogeneity occur during the integration of Industry 4.0 documents?
3. Empirical evaluation of our hypothesis to measure ALLIGATOR performance with respect to state-of-the-art approaches.
 - (a) Implementation of state-of-the-art or baselines approaches.
 - (b) Definition of benchmarks to evaluate the proposed solutions.
 - (c) Design and execution of experiments and statistical tests to validate or falsify our hypotheses. The result section was a first attempt to measure the effectiveness of ALLIGATOR. We designed a controlled experiment where some of the heterogeneity types were measured. Based on this idea we plan to extend the Testbeds to cover the following:
 - i. All the heterogeneity types, i.e., from M1 to M7 for the *intra*- as well as for the *inter*-standard interoperability issues.
 - ii. A validation with domain experts to evaluate results of the integrated documents, as well as the generated explanations. This validation will be conducted for both problems, *intra*- and *inter*- standard interoperability issues, and for all the heterogeneity types.
 - (d) Analysis and discussion of the observed results.

6 Conclusions and Future Work

The problem of assessing *intra*- and *inter-standard* interoperability is addressed in this work, and we propose the combination of Deductive databases and Semantic Web technologies to effectively solve these problems. As a proof-of-concept,

we present ALLIGATOR, a deductive framework for the integration of Industry 4.0 Standard documents. ALLIGATOR relies on Datalog and RDF to *accurately* represent the knowledge characterizing different types of semantic heterogeneity for documents described in Industry 4.0 standards. Currently, the main focus of our approach is to solve semantic heterogeneity issues that may occur among documents defined in the same, i.e., *intra-standard interoperability* problem. Nevertheless, we additionally plan to extend ALLIGATOR to integrate different standards and assess the *inter-standard interoperability* problem.

The results of the empirical evaluation for the *intra-standard interoperability* problem using AutomationML standard, suggest that ALLIGATOR is able to effectively solve the problems of *AutomationML semantic equivalent Identification element* and exhibits similar behavior for the three studied semantic heterogeneity types, i.e., granularity (M2), schematic (M3), and grouping (M6). In the future, we will empower the ALLIGATOR Deductive System Engine with the expressiveness of Datalog with negation and built-in predicates. Thus, ALLIGATOR will be able to represent all the other types of semantic heterogeneity in AutomationML, i.e., from M1 to M7. The heterogeneity types will be also defined for the *inter-interoperability* problem, i.e., between AutomationML and OPC UA. Further, we will extend ALLIGATOR to create explanations of the aligned elements of the integrated Industry 4.0 standard documents. Finally, we envision to develop a more general framework, capable of semantically integrate documents expressed in different Industry 4.0 standards.

Acknowledgments. The author would like to thank to Sören Auer and Maria-Esther Vidal for their guidance and fruitful discussions during the development of this doctoral work.

References

1. Biffi, S., Kovalenko, O., Lüder, A., Schmidt, N., Rosendahl, R.: Semantic mapping support in AutomationML. In: ETFA, pp. 1–4. IEEE (2014)
2. Ceri, S., Gottlob, G., Tanca, L.: What you always wanted to know about datalog (and never dared to ask). IEEE Trans. Knowl. Data Eng. **1**(1), 146–166 (1989)
3. eClass e.V.: eCI@ss standardized material and service classification (2016)
4. Enste, U., Mahnke, W.: OPC unified architecture. Automatisierungstechnik **59**(7), 397–405 (2011)
5. e.V. AutomationML, OPC Foundation: OPC UA information model for AutomationML. Status report (2016)
6. Grangel-González, I., Collarana, D., Halilaj, L., Lohmann, S., Lange, C., Vidal, M.-E., Auer, S.: Alligator: a deductive approach for the integration of Industry 4.0 standards. In: Blomqvist, E., Ciancarini, P., Poggi, F., Vitali, F. (eds.) EKAW 2016. LNCS (LNAI), vol. 10024, pp. 272–287. Springer, Cham (2016). doi:[10.1007/978-3-319-49004-5_18](https://doi.org/10.1007/978-3-319-49004-5_18)
7. Henßen, R., Schleipen, M.: Interoperability between OPC UA and AutomationML. In: Procedia CIRP 8th International Conference on Digital Enterprise Technology DET, vol. 25 (2014)

8. Himmler, F.: Function based engineering with automationml - towards better standardization and seamless process integration in plant engineering. In: 12th International Conference on Tagung Wirtschaftsinformatik, WI (2015)
9. Kovalenko, O., Euzenat, J.: Semantic matching of engineering data structures. In: Biffl, S., Sabou, M. (eds.) *Semantic Web for Intelligent Engineering Applications*. Springer, Cham (2016)
10. Kovalenko, O., Wimmer, M., Sabou, M., Lüder, A., Ekaputra, F.J., Biffl, S.: Modeling AutomationML: semantic web technologies vs. model-driven engineering. In: 20th IEEE Conference on Emerging Technologies & Factory Automation, ETFA, pp. 1–4 (2015)
11. Lange, C.: Krextor - an extensible XML→RDF extraction framework. In: *Scripting and Development for the Semantic Web, SFSW*, vol. 449. CEUR Workshop Proceedings, Aachen, May 2009
12. Panetto, H., Zdravkovic, M., Jardim-Gonçalves, R., Romero, D., Cecil, J., Mezgár, I.: New perspectives for the future interoperable enterprise systems. *Comput. Ind.* **79**, 47–63 (2016)
13. Persson, J., Gallois, A., Björkelund, A., Hafdell, L., Haage, M., Malec, J., Nilsson, K., Nugues, P.: A knowledge integration framework for robotics. In: 41st International Symposium on Robotics and ROBOTIK 2010 (2010)
14. Sabou, M., Ekaputra, F., Kovalenko, O., Biffl, S.: Supporting the engineering of cyber-physical production systems with the AutomationML analyzer. In: 1st International Workshop on Cyber-Physical Production Systems, CPPS, pp. 1–8. IEEE (2016)
15. Schleipen, M., Damm, M., Henßen, R., Lüder, A., Schmidt, N., Sauer, O., Hoppe, S.: OPC UA and AutomationML-collaboration partners for one common goal: Industry 4.0. (2014)
16. Schleipen, M., Gutting, D., Sauerwein, F.: Domain dependant matching of MES knowledge and domain independent mapping of AutomationML models. In: 2012 IEEE 17th Conference on Emerging Technologies & Factory Automation, ETFA, pp. 1–7. IEEE (2012)
17. Schleipen, M., Okon, M.: The CAEX tool suite - user assistance for the use of standardized plant engineering data exchange. In: 15th IEEE International Conference on Emerging Technologies and Factory Automation, ETFA (2010)
18. Schmidt, N., Lüder, A., Rosendahl, R., Ryashentseva, D., Foehr, M., Vollmar, J.: Surveying integration approaches for relevance in cyber physical production systems. In: ETFA, pp. 1–8. IEEE (2015)