Digital Models for Data Analytics and Digital Twins in Industrial Automation Applications

Introduction of a Common Interoperability Registry for linking diverse functional domains

Nikos Kefalakis
IoT Group Athens Information Technology (AIT)
Kifisias Ave. 44, Marousi, 15125 Athens, Greece
e-mail: nkef@ait.gr

John Soldatos
IoT Group Athens Information Technology (AIT)
Kifisias Ave. 44, Marousi, 15125 Athens, Greece
e-mail: jsol@ait.gr

John Kaldis
IoT Group Athens Information Technology (AIT)
Kifisias Ave. 44, Marousi, 15125 Athens, Greece
e-mail: jkaldis@ait.gr

Mauro Isaja
Research & Development
Engineering Ingegneria Informatica SpA (ENG)
Via Ferrini, 47 - Loc. San Martino
53035 Monteriggioni, Italy
e-mail: mauro.isaja@eng.it

***Abstract*— Digital representations of the physical world through the renowned "digital twin" concept, within industry 4.0 and Industrial Internet of Things (IIoT) environments, gave rise to several digital modelling approaches. This paper illustrates a complete digital model specifically focused on intensive Big Data operations, which is a common industry4.0 requirement. The model is established on patterns of world acclaimed standards-based digital models, as it was deployed successfully in one predictive maintenance manufacturing project and one project on edge computing with a block chain layer. Furthermore, the paper introduces the Common Interoperability Registry, a novel addition in the form of a standards-based, vendor-neutral method to map object entities belonging to different systems/databases. This facilitates discoverability and inserts a global unique identifier among entities from different functional domains.**

Keywords- Digital Twins; Digital Models for Distributed Data Analytics; Common Interoperability Registry; Industrial Internet of Things; Industry4.0; Manufacturing Plant Modelling

I. INTRODUCTION

Digital modelling of the physical world is one of the core concepts of industry digitization and the fourth industrial revolution (Industry 4.0). It foresees the development of digital representations of physical world objects and processes as a means of executing automation and control operations, based on digital operations and functionalities (i.e., at the cyber world). This concept is conveniently called "digital twin". The advent of Industry 4.0 (including models and standards, such as RAMI 4.0) has led to the identification of standards-based data schemas and data formats, which can be used for describing plants, automation operations, production systems and more. Usually, digital models are accompanied by a set of functions, which undertake the synchronization of these models with the physical world entities that they represent. Digital models serve three complementary objectives:

a) Semantic interoperability

By providing a uniform representation of the concepts and entities that comprise an IIoT deployment, they boost semantic inter-operability across diverse digital systems and physical devices. Indeed, the use of common data model provides a uniform vocabulary for describing sensors, Cyber-Physical System (CPS) devices, production systems and more.

b) Information Exchange

Digital models provide a basis for exchanging information across different similar deployments, which is closely related to the inter-operability objective.

c) Digital Operations

Digital models are a key prerequisite for performing automation and control operations at IT (Information Technology) timescales. Processes and devices can be configured through IT systems that configure and update digital models, which reflect the status of the physical world. However, this requires a synchronization, which can be challenging to implement. The functionalities of digital models should support:

- Factory and Plant Information Modelling
- Automation and Analytics Processes Modelling
- Synchronization of Cyber and Physical worlds
- Dynamic Access to Plant Information

Nevertheless, the above properties do not address issues, such as data intensive applications, similar to those faced within H2020 FAREDGE and H2020 PROPHESY. Moreover, existing digital models were found insufficient for this purpose. Section II explains why a novel model was required to be developed. Section III presents the guideline standards for our design. Section IV illustrates the new data model in detail. Section V introduces the Common Interoperability Registry (CIR), the merging problems that it solves, and its potential applications for concurrent use of different models & standards.

II. REASONING FOR A NEW MODEL

The reasoning behind the introduction of a new Data Model in this paper is twofold: First, to focus on data-intensive applications like data streams, analytics, digital twins on analytics, etc., and second, to provide a CIR implementation for linking with other relevant data models. Most Industry4.0 applications are data driven and hence digital modeling of Big Data operations is a cornerstone requirement. In this respect the FAR-EDGE data model is tailored to data intensive operations, rather than lightweight automation functions. Furthermore, linking and integration of other data models is made possible through the CIR, ensuring suitability for a wider class of Industry4.0 applications. The resulting digital model has been successfully deployed in two predictive maintenance cases of H2020 PROPHESY, and two factory cases of the H2020 FAREDGE edge computing with block-chain project.

III. STANDARDS-BASED DIGITAL MODELS

For over a decade, various industrial standards have been developed, and a long list of relevant digital models exists. Many standards come with a set of semantic definitions, typically used for modelling and exchanging data across systems and applications. For reviews and comparative assessments, interested readers can refer to relevant literature (e.g., [1][3][4]). The most prominent ones, which have also been driving the specifications for the relevant H2020 projects PROPHESY and FAREDGE, can be found below. Several of them are referenced and/or used by RAMI4.0.

A. IEC 62264 B2MML

An XML based specification and implementation of the ANSI/ISA-95 family of standards, and a very good choice for modelling interactions across entities within MES and ERP systems and their involvement in automation operations. With reference to this hierarchy, the standard covers the domain of manufacturing operations management (i.e., Level 4) and the interface content and transactions within Level 3 and between Level 3 and Level 4. Hence, the standard is primarily focused on the integration between manufacturing operations and control, rather than on pure control (i.e., Levels 1, 2, 3).

B. IEC 61512 BatchML

An XML based implementation of the ANSI/ISA-88 Batch Control family of standards, suitable for the modelling of ISA-88 compliant systems.

C. IEC 62769 (FDI)

It includes an information model that represents automation systems' topologies, including field devices and the communication networks that interconnect them, and is hence suitable for modelling information on the field layer of the factory (devices, networks), but without provisions for data analytics.

D. ISO 15926 XMplant

It covers the structure, the geometry and 3D models about a plant, and provides support for digital modelling of plant information, based on the ISO 15926 specification. It is a good choice for modelling the static elements and behaviors of a plant.

E. IEC 62453 (FDT)

The IEC 62453 Field Device Tool (FDT) by fdtgroup.org, is an open standard for industrial automation integration of networks and devices. It provides standardized software to enable intelligent field devices that can be integrated seamlessly into automation applications, from the commissioning tool to the control system. FDT supports the coupling of software modules, which have been implemented as representatives for field devices and are therefore able to provide and/or exchange information.

F. IEC 61512 (Batch Control)

IEC 61512 – Batch control is also referenced by RAMI 4.0. It models batch production records, including information about production of batches or elements of batch production.

G. IEC 61424 (CAEX)

It provides the means for modelling a plant in a hierarchical way. It supports an XML-based representation of plant information, including all components in a hierarchical structure, and adopts an object-oriented philosophy. CAEX separates vendor independent information (e.g., objects, attributes, interfaces, hierarchies, references, libraries, classes) and application dependent information, such as certain attribute names, specific classes or object catalogues. CAEX is appropriate for storing static metadata, but it not designed to hold dynamic information. CAEX can cover the modelling of the plant elements, but is inappropriate for modelling maintenance-related information such as sensor-based datasets.

H. IEC 62714 AutomationML

AutomationML is an XML-based open standard, which provides the means for describing the components of a complex production environment through a hierarchical structure, and it is commonly used to facilitate consistent exchange and editing of plant layout data across heterogeneous engineering tools. It relies on three other standards, namely: CAEX (IEC 62424) in order to model topological information, COLLADA (ISO/PAS 17506) of the Khronos Group in order to model and implement geometry concepts and 3D information, as well as Kinematics (i.e., the geometry of motion), and finally PLCopen XML (IEC61131) in order to model sequences of actions, internal behavior of objects and I/O connections.

I. MTConnect

MTConnect provides an XML-based format for exchanging data between the shop-floor and IT applications, including data about devices, topologies and component characteristics.

J. PERFoRMML

The H2020 PERFORM project is devoted to the development and validation of a plug-n'-produce infrastructure. Following a comprehensive review and evaluation of various data models, the PERFORM consortium has selected AutomationML as the base for building its own common data model, conveniently called PERFoRMML. It makes provisions for modelling/representing the following:

1) Machinery and Control Systems

They provide the means for modelling the topology, data types and interactions of production systems at physical machinery level. The attributes of these entities enable capturing and modelling of parameters for configurations and skills, as well as for shop-floor data to be extracted from various sources such as PLCs and databases. In particular, the following sub-entities are also modelled through proper subclasses:

- Skills (e.g., pick, place, move, weld etc.) refer to abilities, functions or tasks performed by shop-floor elements. They may possess and certain values that are relevant to be extracted (e.g., cycle time, energy consumption, and sensor data).
- Configurations provide a high-level description of a possible configuration to execute a given skill, according to a set of specified parameters.
- Products, which correspond to abstractions of given product variants, along with their core-defining characteristics to enable a process-oriented description of the product.
- Processes, which present the ordered steps required for the production of an associated product.
- Connectors, which encapsulate and abstract the information required to communicate with components in the shop-floor. The abstraction property enables to support communications regardless of the actual communication protocols (e.g., OPC-UA, MQTT) used.
- Events, modelling certain occurrences in production that require the attention of the system or its users.

2) Data Backbone entities

These model the elements necessary for interactions with the tools connecting to the PERFORM middleware. These entities can acquire data and information from the lower-level and act based on it, and they include:

- System, which is an entity representing entire production systems and therefore comprises systems information in terms of topology, products and possible simulations.
- Simulation Results, which support the representation of some simulation outcome (usually a KPI: Key Performance Indicator) in-line with the PERFORM's digital twins requirements.
- Schedules, which model the allocation of the (end-to-end) steps that need to be executed in order for the production of certain product.

For each of the two sets of entities (machinery & control, data backbone), the project specified standard based interfaces for accessing instance data of the various entities. These interfaces form the basis for an API (Application Programming Interface) as well.

IV. THE FAR-EDGE DATA MODEL

The root element of the FAR-EDGE Digital Models is the "FAR-EDGE DM" and at the next hierarchy level, a set of further XSD Schemata are designed. The FAR-EDGE Digital Models' factory data and metadata are based on the entities:

a) For factory data description:

- Data Source Definition (DSD): Defines the properties of a data source in the shopfloor, such as a data stream from a sensor or an automation device.
- Data Interface Specification (DI): It is associated with a data source and provides the information needed to connect to it and access its data (e.g., network protocol, port, network address).
- Data Kind (DK): This specifies the semantics of the data of the data source. It can be used to define virtually any type of data in an extensible way.
- Data Source Manifest (DSM): Specifies a specific instance of a data source in line with its DSD, DI and DK specifications. Multiple manifests are therefore used to represent the data sources that are available in the factory.
- Data Consumer Manifest (DCM): Models an instance of a data consumer, i.e., any application that accesses a data source.
- Data Channel Descriptor (DCD): Models the association between an instance of a consumer and an instance of a data source. Keeps track of the established connections and associations between data sources and data consumers.
- LiveDataSet: Models the actual dataset that stems from an instance of a data source that is represented through a DSM. It is t is associated with a timestamp and keeps track of the location of the data source in case it is associated with a mobile edge node. In principle, the data source comprises a set of name–value pairs, which adhere to different data types in line with the DK of the DSM.
- Edge Gateway: Models an edge gateway of an edge computing deployment. Data sources are associated with an edge gateway, which usually implies not only a logical association, but also a physical association as well.

Based on the above entities, it is possible to represent the different data sources of a digital shopfloor in a modular, dynamic and extensible way. This is based on a repository (i.e., registry) of data sources and their manifests, which keeps track of the various data sources that register to it.

b) For factory analytics description, analytics workflows and pipelines:

- Analytics Processor Definition (APD): Specifies processing functions applied on one or more data sources. Three types of processing functions are supported, including data preprocessing, data storage, and data analytics functions. These can be combined in various configurations over the data sources in order to define analytics workflows.

- Analytics Processor Manifest (APM): Represents an instance of a processor that is defined through an APD. Each instance specifies the type of processor and its actual logic through linking to an implementation function (like a Java class).
- Analytics orchestrator Manifest (AM): Represents an analytics workflow as a combination of analytics processor instances (i.e., APMs). It is likely to span over multiple edge gateways and to operate over their data sources.

V. THE CIR COMMON INTEROPERABILITY REGISTRY

A novel addition to the proposed digital model, is the Common Interoperability Registry (CIR) that enables the merging of the data models from the different functional domains. Specifically, CIR provides a standards-based, vendor-neutral method to map object entities belonging to different systems/databases that share common business context. Additionally, it:

- Enables the discoverability and relation of the registered objects and helps third party applications to combine the information provided from these systems/databases.
- Provides a global unique identifier (in a UUID format) for the registered objects.

CIR can be viewed as the infrastructure for linking objects and their information that reside in different databases, and enables the enrichment of the underlying datasets based on additional data and metadata residing in the linked databases. For instance, considering different functional domains, such as the Data Sources, the Virtualization/Simulation and the Automation, the need arises for information sharing among them. The obvious upside is that an IIoT application will be able to consult and access (through the CIR) information about the full context and observations that are related to an object, regardless of the repository they reside. Likewise, a flexible extension of the digital models' infrastructure with information (data/metadata) stemming for additional repositories and databases is possible. Nevertheless, note that this will require each new repository to be linked to objects of the project's database at the time of their deployment. The CIR provides an XML schema and a relational DB describing the specification. The OpenO&M (CIR) is open source and the latest version can be found in the MIMOSA organization GitHub. As implemented, the CIR [8] includes:

- Registry: The container object for a set of categories.
- ID: The user-defined identifier of the registry.
- Description: Description and expected use of CIR
- Category: Categories define sets of potentially related entries, such as equipment, which have alternate names on different systems.

VI. CONCLUSION AND FUTURE WORK

In the complex landscape of various standards for digital modelling in Industry 4.0, there exists no "one size fits all" solution that will prevail, until the present day. Standards are tailored to different applications, e.g., automation, simulation, digital twins, Big Data analytics, supply chain management, etc. Our needs dictated the design of a new model focused on data collection, routing and analytics i.e., typical data-intensive applications. It is based on several world-renowned, standards-based digital models. The future vision of a "Fully Digital Shopfloor" (i.e., for all production processes) will require the concurrent use of different models & standards. Hence, there is a need for more mechanisms to link those standards (like the proposed CIR), to digitally reflect the shopfloor consistently.

ACKNOWLEDGMENT

Part of this work (namely the FAREDGE DM and its further schemata) has been carried out in the scope of the H2020 FAR-EDGE project (contract number 723094), while another part (namely the CIR design and implementation) has been carried out in the scope of H2020 PROPHESY project (contract number 766994). Both projects are funded by the European Commission. The authors acknowledge valuable help and contributions from all partners of both projects, including the plants where the DM and CIR infrastructure have been validated.

REFERENCES

[1] H. Lasi, P. Fettke, H.G. Kemper, T. Feld, and M. Hoffmann, "Industry 4.0", Business & Information Systems Engineering, vol. 6, no. 4, pp. 239, 2014.

[2] A. W. Colombo, T. Bangemann, and S. Karnouskos, "IMC-AESOP Outcomes: Paving the way to Collaborative Manufacturing Systems", Proceedings of the 12th IEEE International Conference on Industrial Informatics (INDIN'14), pp. 255-260, 2014.

[3] W. Lepuschitz, A. Lobato-Jimenez, E. Axinia, and M. Merdan, "A survey on standards and ontologies for process automation" in Industrial Applications of Holonic and Multi-Agent Systems, Springer, pp. 22-32, 2015

[4] R. S. Peres et al, "Selection of a data exchange format for industry 4.0 manufacturing systems," IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society, Florence, 2016, pp. 5723-5728, doi: 10.1109/IECON.2016.7793750.

[5] "IEC 62714 engineering data exchange format for use in industrial automation systems engineering - automation markup language - parts 1 and 2" in , International Electrotechnical commission, pp. 2014-2015.

[6] J.Soldatos, O.Lazaro, and F.Cavadini "The Digital Shopfloor: Industrial Automation in the Industry 4.0 Era -Performance Analysis and Applications" ISBN: 9788770220415 River Publishers

[7] S. Faltinski, O. Niggemann, N. Moriz, and A. Mankowski, "AutomationML: From data exchange to system planning and simulation," IEEE International Conference on Industrial Technology (ICIT), 2012, pp. 378–383.

[8] A. Mathew, K. Bever, and D. Brandl, "Web Service Common Interoperability Registry 1.0", OpenO&M Candidate Standard 19 June 2015, available at: http://www.openoandm.org/ws-cir/1.0/ws-cir.html [retrieved: September, 2019]