Industry 4.0: Standardization and Reference Architecture Model

Executive summary: *Future industrial systems, because of the "digitalization" of their components, will produce and consume large amounts of data. Data thus becomes itself a product. Exchanging industrial data is however a challenge. To address this challenge, two things are of particular relevance: (1) standard data representation formats that guarantee semantic interoperability and (2) requirements for interchangeable building blocks to be used in data analytics workflows. In order to conduct this standardization effort, a close collaboration between researchers and industrial stakeholders is necessary.*

Introduction

As France and Germany are shaping their strategy for industrial development in the 21st century, a general consensus has already emerged: industrial systems will be digital. Product lines will incorporate digital sensing components, as well as machines with digital commands, such that industrial control becomes more accurate, more versatile, more robust. Vast amounts of digital data are already being generated by product lines and, at an even bigger scale, entire supply chains. Data itself becomes a product, such that an entire business has settled around it. The data business comes with its own practices to design and engineer data, to produce it at a massive scale through well-defined processes and, finally, to commercialize it.

Data, however, is in many ways a peculiar product that differs from other commodities, and that on all three aspects of engineering, production and commercialization. First, individual sets of data have more value when they are combined than taken in isolation. Second, unlike any product being the result of a transformation process from raw materials, data is not physically constrained. It requires tools to be manipulated and can be anywhere, anytime. Finally, when data is sold, its property is not transferred by default.

As a result, there is a significant gap in terms of practices between traditional industries and this emerging "industry" of data. When manufacturers wish to make use of the data their machines produce, significant challenges thus arise. The present document offers a brief overview of these challenges, in order to take action in the context of French-German cooperation.

Data Analytics Workflows

The benefits of analyzing data generated in factories, like the hundreds of sensor readings embedded in an electro-mechanical drive or traces left by digital tags as items move across a product line, have already been demonstrated at various occasions. It helps reduce production times, adapt more quickly to fluctuating order books and, above all, it allows for predictive rather than reactive maintenance. As an example, the Smart Data Innovation Lab (SDIL) in Karlsruhe has conducted an intense technology transfer activity over the past three years, running more than fifty projects with companies in need for expertise on data analytics [1].

With the experience it gained through this activity, the SDIL has developed a unified workflow combining open and proprietary data to be able to quickly deliver results (as fast as every month, for most contracts). At a high level, this workflow includes three phases: preparation, realization and finalization. Despite the fact that the added value of data analytics lies in the realization phase (in which machine learning plays an important part), most efforts are generally put in the preceding phase of preparing data (where data is extracted from sources, then transformed and optionally replicated on SDIL servers). If the entire workflow were to be used at a larger scale as is envisioned in future industrial systems, that effort on data preparation should be considerably reduced.

The main issue on data preparation is the high heterogeneity of sources, using various representations and conventions to expose data. Of course, there would not be such a problem if there existed a single platform all companies could rely on to publish and consume industrial data. However, letting a dominant platform emerge, as has happened with the GAFAM for end-consumer services, is not desirable economically and also not a necessity, technologically. It would indeed suffice to standardize data representations such that consumers can browse the data exposed by publishers the same way they would browse the Web—as long as they have been granted access to that data.

Data Representation Standards

There exist many primitives to represent complex data: key-value pairs, lists, trees, graphs, tables and others. These abstract data structures can be each materialized in many more interchange formats, like JSON, XML or CSV. Even if, in practice, most data is eventually represented in one of these formats, consumers and providers have to agree on the meaning of a node in a JSON or XML tree or the meaning of a key in a key-value pair. This agreement, which defines the semantics of the data, is not trivial to acquire. Even simple (x, y) coordinates can be represented as lists, as key-value pairs or within tables.

The concept of semantic interoperability, i.e. the ability of independent systems to agree on the meaning of the data they exchange, has been extensively studied over the past decades and there already exist mature technologies to address this issue, in particular for relational databases. Such technologies have yet to be adapted to fast-changing and highly heterogeneous data, though. The most advanced approach so far has been to align existing data representations with semantic models openly available on the Web. Such models are called Web ontologies. One can cite the Semantic Sensor Network (SSN) ontology [2], the Smart Applications REFerence (SAREF) ontology and its derivatives [3], and, to some extent, the upcoming Thing Description (TD) ontology [4].

Web ontologies have been regularly criticized for their complexity. Simpler models also exist in various application domains although access to these models is often restricted, e.g. when published by ISO/IEC. In contrast, the public availability of Web ontologies is key in the perspective of a large adoption. To ease their use, public ontology portals like saref.etsi.org [3] can play an important role by the guidance they offer to the users of these semantic models. To be successful, however, these portals have to be developed in close collaboration with their direct users, be they engineers, software developers or other industrial stakeholders.

Data Ownership and Privacy Preservation

Complementary to the question of representing data in a unified way remains the question of exchanging it. Precisely on this aspect, many concerns are being raised among business people. Indeed, they who own the data also get the profits resulting from its processing. A single stakeholder can however not hold the whole value chain associated to industrial data: equipment suppliers provide means to acquire data on their machine, manufacturers acquire it while running their product lines and information technology companies have the infrastructure to analyze it. Such a configuration necessitates that data is being exchanged at some point. It is yet not clear how to solve the technological challenge of transferring data without transferring or extending its ownership. Theoretical solutions exist (homomorphic encryption and distributed ledgers like blockchains, among others) but there is room for French, German and other European companies to acquire that technological know-how.

Meanwhile, industries can start a standardization activity to develop reusable building blocks around data exchange. This is for example the goal of the International Data Space Association (IDSA) [5]. The IDSA is working on a high-level specification of data spaces, which are virtual spaces in which data consumers and data providers negotiate services, applications and contracts. The specification does not commit to given technologies. Rather, it elicitates the necessary components of data spaces, as well as their interconnection. A similar effort is being conducted at ISO on specifying buildings blocks for data analytics. Such activities should eventually empower companies to process their own data themselves, to a certain extent, and exchange it with third-parties at low cost and low security risk. Technology transfer institutions like the SDIL would then leave the floor to a network of small and medium-size enterprises selling their expertise to large manufacturers. To enable this situation and thus guarantee the success of IDSA, it is crucial that it gain members from as many industrial sectors as possible.

Conclusion

When data becomes a product, new practices have to be developed. In most cases, a significant level of standardization is required. Because exchanging industrial data necessitates recent scientific advances to turn into practical technologies, a close collaboration between researchers and industrial stakeholders is paramount. This is particularly true for the domains of machine learning and cryptography. Moreover, diversity is an important factor in making standards being adopted in practice. A standardization body is diverse if its members have various operational roles (as necessary in the development of saref.etsi.org) as well as expertise in various industrial sectors (as desired in the IDSA).

Standardization results from a fragile balance between cooperation on what is necessary and competition on everything else. In the context of industrial data, to drive future digitalized industries, there is no doubt that (1) data representation formats and (2) reusable data analytics workflows are among what is necessary.

[1] https://www.sdil.de/en/
[2] https://www.w3.org/TR/vocab-ssn/
[3] https://saref.etsi.org/
[4] https://www.w3.org/TR/wot-thing-description/
[5] https://www.internationaldataspaces.org/