Data Management in Cyber-Physical Work Environments

Influences on a Decision Model Derived from the Example of a Facility Management Support System

Clauß, Michael.; Müller, Egon
Department of Factory Planning and Factory Management
Chemnitz University of Technology
Chemnitz, Germany
e-mail: michael.clauss@mb.tu-chemnitz.de
e-mail: egon.mueller@mb.tu-chemnitz.de

Hofmann, Marcus
Department of Business Informatics
Dresden University of Cooperative Education
Dresden, Germany
e-mail: marcus.hofmann@ba-dresden.de

Götze, Janek.; Schumann, Christian-Andreas
Department of Business Informatics
Zwickau University of Applied Sciences
Zwickau, Germany
e-mail: janek.goetze@fh-zwickau.de
e-mail: christian.schumann@fh-zwickau.de

***Abstract*—Semantic technologies are said to have huge advantages over traditional data keeping approaches regarding flexibility and interpretability that are of increased importance in rather unstructured environments such as cyber-physical systems (CPS). But what are the parameters that influence a decision for or against its application in real-world data integration projects? Based on the findings of the ongoing research project FMstar (Facility Management with semantic technologies and augmented reality), the article derives some relevant influences on a respective decision model using the example of a facility management (FM) scenario.**

Keywords-data management; semantic technologies; factory management; facility management; decision model; parameter identification.

I. INTRODUCTION

Data is said to be the new oil both of the New or Digital Economy and, in a mere reflecting manner, of the Old Economy, too. First of all, why is that? With the development and implementation of approaches like the Internet of Things [1] or the concept of CPS [2], the real world gradually merges with its virtual counterpart. In manufacturing, this means that new insights from analyzing the data can not only be used for supplementary value-added services around the core business [3], but also have an impact on how the core business itself works inside. Approaches for logistic control systems that utilize the ubiquitous availability of data at the runtime of a manufacturing system are essential for modern flexible and changeable manufacturing systems [4]. Unfortunately, these new capabilities come with new dependencies that need an active and foresighted management in order to ensure reliability and profitability. So, the question that may arise is what technology is the best to organize the data in a certain area of application? Since the answer to that general question supposedly is a rather complex one, this paper will focus on practical experiences with data management in a well-defined area of application: facility management. Therefore, the article will follow an inductive approach and is structured as follows. First of all, a compact overview of the state of the art of data management in the factory management domain with an outline of open issues is given in Section II. Subsequently, in Section III, the facility management subdomain with a use case from the FMstar project [5] will be described and reviewed for relevant influences on the decision making process. The outcome of that will be condensed and integrated into a preliminary draft of a decision model interface description in Section IV that summarizes relevant influencing factors. A short summary and conclusion will be given in Section V.

II. STATE OF THE ART

Today's practice of data generation, storage, usage and management varies among different areas of application. To start with, personal work is often supported by tools such as Excel or individually organized file storage systems. There is no explicit semantics and the principles used to organize the data depend a lot on the personal preferences of the user. In [6], some further dedicated tools for more collaborative tasks such as computer-aided design (CAD), knowledge base engineering (KBE), product data model/ product lifecycle management (PDM/PLM) or enterprise resource planning (ERP) systems are exemplarily identified. They make use of more adequately structured data models that ensure interchangeability, at least to a certain extent. In the majority of cases, relational databases and proprietary storage solutions are used. The latter often can only be accessed through more or less lossy export mechanisms based on particular exchange standards. A full picture is hard to draw

 ISBN: 978-1-61208-420-6

at this point. Some further examples will be given in Section III. The basic distinction that can be made so far is the one between relational databases and so-called NoSQL approaches [7]. While relational databases still represent the dominating approach for data storage, the NoSQL approach refers to a larger group of database types (e.g., document-, object- or graph-based databases etc.) that become more and more important in cyber-physical application environments such as factory management. Since they offer potential for distributed architectures and in-memory operations they are more flexible and much faster in certain situations, but still lack powerful mechanisms to update and retrieve data at a large scale. There are pros and cons and further enhancements on both sides [8]. Their coexistence as well as the need for an economically justifiable integration of indeed old-fashioned designed but still indispensable legacy systems can be regarded as a given fact that has to be respected.

However, complexity in CPS design rises due to an increasing variety of requirements that have to be met [9]. Data has to be capable of inter-domain operation and this will have consequences for the scope of rather local or dedicated data management solutions [10]. The integration of data along the industrial value chains and its persistence throughout the whole product life cycle is necessary for legal or automation purposes [11]. The aforementioned PDM/PLM solutions offer part of that functionality, but their capabilities are limited to structures known at build-time. CPS in changeable environments such as modern manufacturing systems require capabilities for an adaption of data structures at run-time of the system. So, instead of or at least complementary to dedicated PLM solutions, an additional integration layer covering all relevant data sources for a specific use case seems like a Swiss army knife-like solution that meets all thinkable global integration demands. That is the point where semantic technologies usually come into play. Moreover, it is exactly the point where it is necessary to evaluate and decide whether a solution based on semantic technologies is really the best option or if traditional data integration approaches that rely on human insights and manual adaptations are the better choice.

III. FACILITY MAINTENANCE USE CASE

For this decision, two basic assumptions that are typical for a realistic scenario have been made in the FMstar project: First of all, the scope of the integration task is limited to the scenario at hand. Higher-level integration is only of theoretical interest unless there are practical guidelines to be used during integration that are supposed to enable such capabilities and usually cause additional, otherwise avoidable efforts. Secondly, the respective data sources cannot be modified in any way without violating their designated application. The consequence of these assumptions is that the syntactic and semantic interoperability can only be provided by some kind of mapping between the different data sources [12] that rely on appropriate schema management mechanisms [13]. A rough distinction between general data integration approaches is illustrated in Figure 1 that shows three alternative solutions for a mapping of the database schemas.

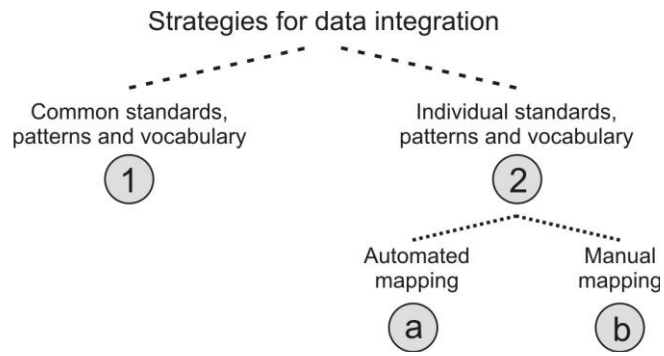

Figure 1. Alternative approaches for data integration [5].

The initial situation in the domain of the project can be best described with this apt quotation from [14]: “The building industry is a collaboration environment that requires repeated, iterative data exchanges and communication among different domains and applications in a high frequency. To automate information processing, standardized and qualified data is necessary for efficient working processes.” What the project team found was a mix of proprietary solutions for the management of 3D model data, maintenance task descriptions and dependency models for the technical infrastructure. If explicit schema models are provided, schema languages such as Unified Modeling Language (UML)/ Extensible Markup Language (XML), Resource Description Framework Schema (RDFS) and Data Definition Language (DDL) could be mapped using mapping languages such as XQuery, SPARQL, TRIPLE or Structured Query Language (SQL) [12]. Unfortunately, they were not available and that is also the reason for the solution approach that was selected: referring to Figure 1, the different databases were integrated manually, so the option 2b was chosen. The disadvantages are obvious; this repetitive process is slow, expensive, causes redundancies and does not meet the domain specific requirements described in the quotation above, since the schemas have to be analyzed manually through time-consuming interviews and workshops. Nevertheless, the solution in the project integrates the data sources using Apache Jena, an open source framework for the semantic web [5]. One goal of the project is to utilize semantic technologies for data integration, which is satisfied by implementing the backend of the FM support system based on Jena. A Google Nexus 10 tablet with Android KitKat (4.4) and libGDX for rendering the user interface (UI) is used as frontend. Figure 2 shows the overall architecture of the prototype. The bottom layer shows the UI of the prototype with a pump from the heating system. With that, the maintenance staff can access relevant information by selecting an object of the 3D model. The visualization approach is the second main focus of the project and aims at an intuitive interaction with the data.

Since the solution for the data integration is not suitable for practical use, the idea of using a vendor-neutral, open Building Information Model (BIM), such as the Industry Foundation Classes (IFC) [14], was raised. This could be used as an intermediary language and as a standard reference for mapping between different data schemas [12].

 ISBN: 978-1-61208-420-6

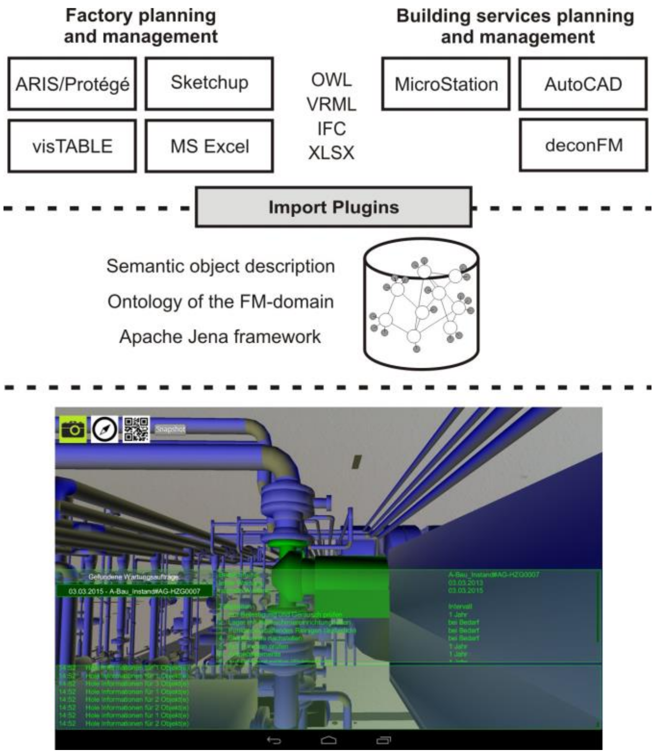

Figure 2. FMstar system architecture [5].

This way the manual mapping efforts could be reduced if each data source would provide a self-description for its schema. Referring to Figure 1, this would enable option 2a and possibly tend to option 1 as a complementary long-term development. However, option 1 represents a rather ideal solution that is very unlikely to be realized since it aims at consistent, interoperable data models at all partners. So, option 2b is probably the desired one, which also provides capabilities for non-redundant, bidirectional data flows between and not only from flexibly integrated systems. The selection of an appropriate data integration strategy depends on a couple of situational influences that will be pointed out based on the introduced use case in the next Section.

IV. CPS Data Management Decision Model

According to [15], data management "is a corporate service which helps with the provision of information services by controlling or coordinating the definitions and usage of reliable and relevant data." So, apparently reliability and relevance of data are central concerns of an appropriate data management strategy. Both concepts depend on the use case that is supposed to be supported. Referring to the central concept of data integration, this would be something related to the target context of the data. Another thing to look at from this definition is the corporate boundaries that more and more lose their limiting character. Data sources and application scenarios can be distributed over several organizations. In other words, the qualities of the source context also influence the decision. As a third area of influences, the transformation process and its capabilities determines what strategy fits best for the project at hand.

A. *Influences*

The description of the influences will be divided into the three areas just mentioned: source context, target context and transformation process. According to the assumptions described in Section II, the source context cannot be influenced by the project. The target context can be regarded as a kind of run-time environment of the desired solution and can be influenced within the project. Following this logic, the transformation process can be interpreted as the build-time environment.

The first area of influences is the source context. It is determined by the structure of the data, the standardization, the roadmap for further development and its dynamics.

- Structure: As mentioned before, the structure of the data sources is determined by the database schema. Even though it cannot be modified in the project, the more detailed the available schema is, the easier it is to utilize semantic technologies for mapping.
- Standardization: Standard-based languages for schema description or domain-specific standards for data organization support the utilization of semantic mapping technologies.
- Roadmap: A roadmap outlining the further development of a source database helps to evaluate its suitability for automated mapping approaches.
- Dynamics: The more often the structure of the source databases changes, the more difficult manual integration will be. Semantic technologies pay off in high volatile environments.

The second area of influences is the target context. It is determined by the use case and, derived from that, by the required quality of the data and the focus of the solution.

- Use Case: The use cases' complexity in terms of inter-domain operation determines how much it would make sense to provide an integrated view on the data. The more complex the use case, the less it pays off to go for manual integration and semantic approaches are very likely the better choice.
- Quality: The importance of data quality is a central aspect that is referred to in many publications in the context of data integration [6]. The more quality matters, the more the consequences of low data quality should be minimized by auxiliary means if semantic technologies are used.
- Focus: The focus of application of the desired solution can reach from local to global. The wider the focus, the more probable is it to have the need to flexibly integrate new data sources and schema information provided in foreign languages.

Influences from the transformation process, the third area of influences, are the competencies of the people involved.

- Competence: The people involved in the data integration project determine the conceptual power and the technology stack that the solution will be built on. Even if it sounds like a platitude, if no experts with profound competencies for semantic technologies are available, the project is better realized with traditional approaches.

B. Towards a Decision Model for CPS Data Management

In Figure 3, the influences that were just identified are summarized in a structured way. The domain gap between the source and the target of the data plays an important role when it comes to automated mapping approaches. Especially in multi-domain application environments, where the use case requires a combination of technically rather unrelated domains, the integration efforts may increase in a non-linear way if the inhomogeneity exceeds certain limits.

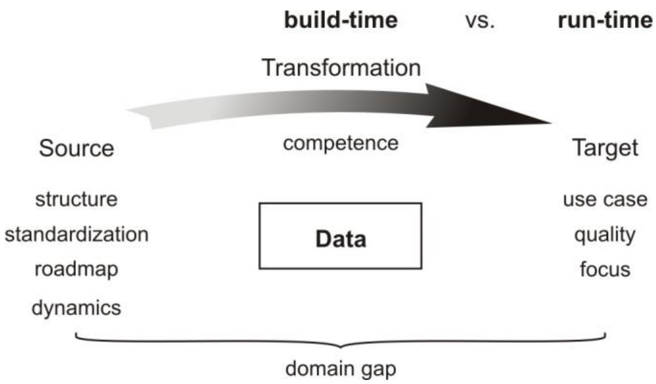

Figure 3. Influences on data management strategy.

However, this first draft of a model provides a very basic structure of the influences on the decision making process. It may support an early stage decision process in a CPS project by providing some relevant influences that should be considered. But for all that, it is supposed to serve as a starting point and to be subject to further refinement.

V. SUMMARY & CONCLUSION

This paper focuses on influences on a decision model for using semantic technologies for CPS engineering projects. While semantic technologies offer a wide range of opportunities to map schemas and integrate data automatically, this does not come for free and even the results are not necessarily satisfying. Influences on the data sources, the capabilities of the transformation process and the desired outcome including the specific use cases have to be analyzed at an early stage of a project in order to make a profound decision of what system architecture and respective data management approach to choose. This pays off in many ways, even though technology evolves away from local data management to cloud-based solutions that provide more powerful capabilities for integration right from the start [16]. However, depending on who is asked, this might be a promising perspective for data integration in the future. In the meantime, a conscientious decision model that balances requirements and specific environmental conditions is inevitable for economically successful CPS projects.

ACKNOWLEDGMENT

The work presented in this paper has been supported by the German Federal Ministry for Economic Affairs and Energy (Bundesministerium für Wirtschaft und Energie – BMWi, VP2333105DB2). We want to thank all the partners that have contributed to the results so far.

REFERENCES

[1] H. Kopetz, Real-Time Systems, 2nd ed. New York: Springer, pp. 307-323, 2011, doi: 10.1007/978-1-4419-8237-7.

[2] M. Broy, Cyber-Physical Systems. Berlin: Springer, pp. 17-32, 2010, doi: 10.1007/978-3-642-14901-6.

[3] Z. Bi, L. D. Xu, and C. Wang, “Internet of Things for Enterprise Systems of Modern Manufacturing,” Industrial Informatics, IEEE, vol. 10, Jan. 2014, pp. 1537-1546, doi: 10.1109/TII.2014.2300338.

[4] H.-P. Wiendahl et al., “Changeable Manufacturing: Classification, Design and Operation,” CIRP Annals Manufacturing Technology, vol. 56, Elsevier, 2007, pp. 783-809.

[5] M. Clauß, J. Götze, E. Müller, and C.-A. Schumann, “Real-time Data Access through Semantic Technologies and Augmented Reality in Facility Management Processes”, Proceedings of International Conference on Innovative Technologies (IN-TECH 2014), Sep. 2014, pp. 45-48, ISBN: 978-953-6326-88-4.

[6] J. Krogstie, „Capturing Enterprise Data Integration Challenges Using a Semiotic Data Quality Framework,“ Business & Information Systems Engineering, vol. 57, Feb. 2015, pp. 27-36, doi: 10.1007/s12599-014-0365-x.

[7] P. J. Sadalage, and M. Fowler, NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence. Upper Saddle River, NJ: Addison-Wesley, 2013, ISBN: 978-0-321-82662-6.

[8] H. Voigt, Flexibility in Data Management. Dresden, 2014, [Online]. Available from: http://nbn-resolving.de/urn:nbn:de:bsz:14-qucosa-136681 [retrieved: June, 2015].

[9] M. M. Bezemer, Cyber-Physical Systems Software Development. Enschede: The Netherlands: University of Twente, 2013, doi: 10.3990/1.9789036518796.

[10] M. D. Ilić, L. Xie, U. A. Khan, and J. M. F. Moura, “Modeling Future Cyber-Physical Energy Systems,” IEEE Power Engineering Society General Meeting, Pittsburgh, PA. Jul. 2008, pp. 1-9.

[11] A. Denger et al., „Organisationaler Wandel durch die Emergenz Cyber-Physikalischer Systeme: Die Fallstudie AVL List GmbH (English title: Organizational Change through the Emergence of Cyber-Physical Systems: The AVL List GmbH Case Study),“ HMD, vol. 51, Oct. 2014, pp. 828-837, doi: 10.1365/s40702-014-0090-4.

[12] H. Adametz, A. Billig, J. Einhaus, and J. Gottschick, Whitepaper Semantische Interoperabilität (English title: Whitepaper Semantic Interoperability). Berlin: Fraunhofer-Institut für Software- und Systemtechnik ISST, 2013.

[13] H. Glatzel, „Schema-Management ohne Schema? – Schema-Verwaltung in NoSQL-Datenbanksystemen (English title: Schema Management without Schema? – Schema Management in NoSQL Database Systems),“ 44. Jahrestagung der Gesellschaft für Informatik, Big Data – Komplexität meistern, Sep. 2014, pp. 2461-2472, ISSN: 1617-5468, ISBN: 978-3-88579-626-8.

[14] C. Zhang, J. Beetz, and M. Weise, “Interoperable Validation for IFC Building Models using Open Standards,“ ITcon – Journal of Information Technology in Construction, vol. 20, 2015, pp. 24-39, ISSN: 1874-4753.

[15] K. Gordon, Principles of Data Management: Facilitating Information Sharing, 2nd ed. Swindon, UK: BCS Learning & Development Limited, 2013, ISBN: 978-1-78017-185-2.

[16] L. Zhao, S. Sakr, A. Liu, and A. Bouguettaya, Cloud Data Management. Cham: Springer, 2014, doi: 10.1007/978-3-319-04765-2.