The Data Value Quest: A Holistic Semantic Approach at Bosch

Baifan Zhou[1], Zhuoxun Zheng[2,3], Dongzhuoran Zhou[2,1], Gong Cheng[4], Ernesto Jiménez-Ruiz[6,1], Trung-Kien Tran[2], Daria Stepanova[2], Mohamed H. Gad-Elrab[2], Nikolay Nikolov[7,1], Ahmet Soylu[3,5,1], and Evgeny Kharlamov[2,1]

[1] SIRIUS Centre, University of Oslo, Norway
[2] Bosch Center for Artificial Intelligence, Germany
[3] Department of Computer Science, Oslo Metropolitan University, Norway
[4] State Key Laboratory for Novel Software Technology, Nanjing University, China
[5] Department of CS, Norwegian University of Science and Technology, Norway
[6] City, University of London, UK
[7] SINTEF, Norway

Introduction. Modern industry witnesses a fast growth in volume and complexity of heterogeneous manufacturing (big) data [1, 2] thanks to the technological advances of Industry 4.0 [3, 1], including development in perception, communication, processing, and actuation. Data has become the new oil for industries[8]. However, despite the effort and time invested in the data business, there still exists a big room for improvement in exploiting the value of data. In particular, data is still often scattered and stored in silos affecting its usage [4]; a lot of data generated by sensors is not used in applications; companies possess precious data but do not have a trustworthy scheme to share its value; etc. There are certainly many ways to address these issues. In this paper we discuss the dimension of meaning in data and how we address it at Bosch (Fig. 1) in a holistic semantification fashion that bestows data with meanings which has always been important for humans to perceive, comprehend, reason, and produce. We believe the emphasis, the clarification, and the promotion of the eminent and profound roles of semantic technologies in the industry should lead to considerable opportunities for advances in technology, growth of profitability, and paradigm change in the industrial practice.

Holistic Semantification at Bosch.

– **Data collection.** Semantification begins with data collection [5]. During which, vast amounts of heterogeneous data with mutli-faceted variety in locations, formats, physical equipment, customisation, etc. are annotated with precise and uniform meta-data, which sets the first corner stone for many activities that are based on the collected data.
– **Data understanding.** In big manufacturing companies like Bosch, data science projects are typically multi-disciplinary teamwork where experts with asymmetric knowledge backgrounds (e.g., engineers, equipment experts,

[8] https://blog.s4rb.com/data-is-the-oil-of-the-21st-century

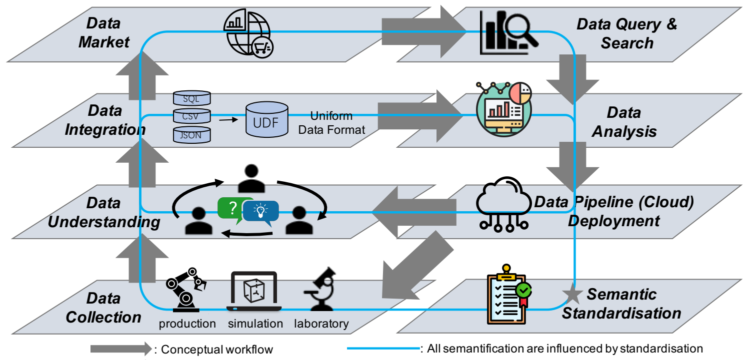

Fig. 1. An overview of our holistic semantification approach

measurement experts, data managers, data scientists, managers) need to talk to each other, to gain a mutual understanding of the process, data, solution, infrastructure, strategic interests, etc [6]. These experts with distinct backgrounds speak different technical or management languages, which tends to lead to error-prone and time-consuming communication. Thanks to their conciseness and unambiguity, semantic models play an essential role here, serving as the "lingua franca" between the experts speaking different languages [7, 8].

– **Data integration.** We rely on ontologies and knowledge graphs (KG) to annotate heterogeneous welding manufacturing data from Bosch and its partners with unified vocabularies. Then, enhanced by the ontology reshaping method developed in Bosch [9, 10], we transform them into uniform data formats/databases that allow uniform access, interoperability, and unified interpretation.
– **Data market.** Bosch participates in a digital open marketplace ecosystem [11], which provides a sustainable approach to connect the data providers and the data consumers to help to connect Bosch and its partners. The ontologies and KGs make the data easier to reach from and by Bosch's production units, suppliers, and customers.
– **Data query & search.** Data like XML files, KGs [12, 13] provide an efficient foundation for querying information of interest via clearly defined formats. SPARQL queries or keywords are used to query data [14–17] for inspection, information summary, and diagnostics. Data search outputs datasets, databases, or snippets of datasets [18–21] and relies on the metadata-based query, KG summarisation, natural language-based search [22], or even the content-based search, which Bosch is researching on.
– **AI and Data analysis.** Here Bosch relies on semantics in diversified ways like scaling usability of data analysis (typically machine learning (ML)-based) pipelines [23] with user interface, which improves the adoption of ML [24], (semi-)automate the generation of ML pipelines with ontologies, templates, and reasoning [25] incorporating domain knowledge via annotation and KG embeddings, etc.
– **Data pipeline deployment (scalability).** Bosch develops semantic abstraction of cloud resources for computing, storage, and networking that

facilitate the deployment of distributed ML pipelines, thus scaling the data analysis onto the big data level [26, 27]. Adaptive rule-based reasoners help to automate the configuration of resource allocation.

– **Semantic standardisation.** Now Bosch participates in the endeavour [28] working towards addressing the long call of the standardisation of semantic artefacts [29], infrastructure, and best practice via e.g. aligning to ISO standards, existing vocabularies, achieving common agreement.

Conclusion. This work gives a panorama view of semantic technologies in the data business at Bosch that is in development. We aim at advancing the exploitation of the values of data in the manufacturing industry. We envision semantic technologies continuing to be one of the keys to unlocking the potential of the values of data.

Acknowledgements. The work was partially supported by the H2020 projects Dome 4.0 (Grant Agreement No. 953163), OntoCommons (Grant Agreement No. 958371), and DataCloud (Grant Agreement No. 101016835) and the SIRIUS Centre, Norwegian Research Council project number 237898.

References

1. S. Chand, J. Davis, What is smart manufacturing, Time Magazine Wrapper.
2. I. Horrocks, M. Giese, E. Kharlamov, A. Waaler, Using semantic technology to tame the data variety challenge, IEEE Internet Comput. 20 (6) (2016) 62–66.
3. H. Kagermann, Change through digitization – value creation in the age of Industry 4.0, in: Management of Permanent Change.
4. G. Gimpel, Bringing dark data into the light: Illuminating existing IoT data lost within your organization, Business Horizons 63 (4) (2020) 519–530.
5. B. Zhou, Machine learning methods for product quality monitoring in electric resistance welding, Ph.D. thesis, Karlsruhe Institute of Technology, Germany (2021).
6. Y. Svetashova, B. Zhou, T. Pychynski, S. Schmidt, Y. Sure-Vetter, R. Mikut, E. Kharlamov, Ontology-enhanced machine learning: A Bosch use case of welding quality monitoring, in: ISWC, Springer, 2020, pp. 531–550.
7. B. Zhou, Y. Svetashova, A. Gusmao, A. Soylu, G. Cheng, R. Mikut, A. Waaler, E. Kharlamov, SemML: Facilitating development of ML models for condition monitoring with semantics, J. Web Semant. 71 (2021) 100664.
8. Y. Svetashova, B. Zhou, S. Schmid, T. Pychynski, E. Kharlamov, SemML: Reusable ML for condition monitoring in discrete manufacturing, in: ISWC (Demos/Industry), Vol. 2721, 2020, pp. 213–218.
9. B. Zhou, D. Zhou, J. Chen, Y. Svetashova, G. Cheng, E. Kharlamov, Scaling usability of ML analytics with knowledge graphs: Exemplified with a Bosch welding case, in: IJCKG, 2021, pp. 54–63.
10. D. Zhou, B. Zhou, J. Chen, G. Cheng, E. Kostylev, E. Kharlamov, Towards ontology reshaping for KG generation with user-in-the-loop: Applied to Bosch welding, in: IJCKG, 2021, pp. 145–150.
11. DOME4.0, Digital open marketplace ecosystem 4.0, https://dome40.eu/, accessed 14 March 2022 (2022).

12. Z. Zheng, B. Zhou, D. Zhou, G. Cheng, E. Jiménez-Ruiz, A. Soylu, E. Kharlamov, Query-based industrial analytics over knowledge graphs with ontology reshaping, in: ESWC (Posters & Demos), Springer, 2022.
13. D. Zhou, B. Zhou, Z. Zheng, E. V. Kostylev, G. Cheng, E. Jimenez-Ruiz, A. Soylu, E. Kharlamov, Enhancing knowledge graph generation with ontology reshaping – Bosch case, in: ESWC (Demos/Industry), Springer, 2022.
14. M. Andresel, D. Stepanova, T.-K. Tran, C. Domokos, P. Minervini, Neuro-symbolic ontology-mediated query answering (2021).
15. Y. Shi, G. Cheng, E. Kharlamov, Keyword search over knowledge graphs via static and dynamic hub labelings, in: WWW, 2020, pp. 235–245.
16. Y. Shi, G. Cheng, T. Tran, J. Tang, E. Kharlamov, Keyword-based knowledge graph exploration based on quadratic group steiner trees, in: IJCAI 2021, 2021, pp. 1555–1562.
17. Y. Shi, G. Cheng, T. Tran, E. Kharlamov, Y. Shen, Efficient computation of semantically cohesive subgraphs for keyword-based knowledge graph exploration, in: WWW, 2021, pp. 1410–1421.
18. X. Wang, J. Chen, S. Li, G. Cheng, J. Z. Pan, E. Kharlamov, Y. Qu, A framework for evaluating snippet generation for dataset search, in: ISWC, 2019, pp. 680–697.
19. X. Wang, G. Cheng, J. Z. Pan, E. Kharlamov, Y. Qu, BANDAR: Benchmarking snippet generation algorithms for (rdf) dataset search, IEEE Transactions on Knowledge and Data Engineering (2021).
20. X. Wang, G. Cheng, E. Kharlamov, Towards multi-facet snippets for dataset search, in: PROFILES/SEMEX@ISWC 2019, 2019, pp. 1–6.
21. X. Wang, G. Cheng, T. Lin, J. Xu, J. Z. Pan, E. Kharlamov, Y. Qu, PCSG: pattern-coverage snippet generation for RDF datasets, in: ISWC, 2021, pp. 3–20.
22. T.-K. Tran, A. Le-Tuan, M. Nguyen-Duc, J. Yuan, D. Le-Phuoc, Fantastic data and how to query them, arXiv preprint arXiv:2201.05026 (2022).
23. B. Zhou, T. Pychynski, M. Reischl, E. Kharlamov, R. Mikut, Machine learning with domain knowledge for predictive quality monitoring in resistance spot welding, Journal of Intelligent Manufacturing (2022) 1–25.
24. B. Zhou, Y. Svetashova, S. Byeon, T. Pychynski, R. Mikut, E. Kharlamov, Predicting quality of automated welding with machine learning and semantics: A Bosch case study, in: CIKM, ACM, 2020, pp. 2933–2940.
25. B. Zhou, Y. Svetashova, T. Pychynski, I. Baimuratov, A. Soylu, E. Kharlamov, SemFE: Facilitating ML pipeline development with semantics, in: CIKM, ACM, 2020, pp. 3489–3492.
26. DataCloud, Enabling the big data pipeline lifecycle on the computing continuum, https://datacloudproject.eu/, accessed 14 March 2022 (2022).
27. D. Roman, N. Nikolov, A. Soylu, B. Elvesæter, H. Song, R. Prodan, D. Kimovski, A. Marrella, F. Leotta, M. Matskin, G. Ledakis, K. Theodosiou, A. Simonet-Boulogne, F. Perales, E. Kharlamov, A. Ulisses, A. Solberg, R. Ceccarelli, Big data pipelines on the computing continuum: Ecosystem and use cases overview, in: ISCC, IEEE, 2021, pp. 1–4.
28. OntoCommons, Ontology-driven data documentation for industry commons, https://ontocommons.eu/, accessed 14 March 2022 (2022).
29. M. Yahya, B. Zhou, Z. Zheng, D. Zhou, J. G. Breslin, M. I. Ali, E. Kharlamov, Towards generalized welding ontology in line with ISO and knowledge graph construction, in: ESWC (Posters & Demos), Springer, 2022.