6. Maintenance. A digital twin is a complicated software and hardware system. How can the maintenance of the necessary data, software and hardware be cost-effective? As noted above, a twin combines life-cycle information, measurements of the assert state and simulations. All these must be maintained so that they reflect the as-built state of the facility. They must be included in the facility´s management of change.

7. Computational overload, edge and cloud. A comprehensive digital twin will require extensive computational resources. These resources will likely be distributed over a hybrid cloud that combines private clouds with vendor platforms and HPC resources. This system needs good design and operation to work. This design will need to carefully define what should be done at the edge of the system and what is done in the cloud.

8. Uncertainty, Validation and Data Science. Finally, a digital twin is only as good as the data and models used in the system. Data must be cleaned and reconciled. Models must be validated and tuned to ensure that they follow the state of the facility. Even "first principles" models must be tuned. Setting up effective methods for tackling uncertainty and model alignment should be a fruitful area for collaboration between control and process engineers and the data science community.

5 A research program for digital twins

How can digital twins be made sustainable, maintainable and useful despite these challenges? We believe that a solution will collaboration between computer scientists, control and simulation engineers, data scientists and end-user technical specialists. Here we will present the computer science and data science parts of solution. This research program combines the sub-disciplines of knowledge representation, natural language technologies, formal methods, scalable computing and data science. This knowledge of technologies must be informed by the deep domain knowledge that is embedded in the digital twins' simulation models and is owned by the facility's engineers – chemical, petroleum, mechanical, electrical and control – and managers.

Challenges related to scope, usability, integration and maintenance can be addressed if we adopt a *semantic backbone* that supports the integration of several different types of digital twin application around a shared understanding of a facility's design. To be successful, this backbone will need to find a pragmatic balance between comprehensiveness and maintainability. It will also need to build on standards in a way that prevents reinvention of the wheel and allows modular construction of semantic models. Recent advances in the construction of ontologies using templates (Forsell, 2017) promise to allow this.

Simulation providers can use these ontologies to allow exchange of model configurations with engineering databases and the exchange of calculated results with the monitoring and optimization layers of the digital twins. Point-to-point connections through tag cross-reference lists can be replaced with declarative mappings – where data in the simulation results or configuration is mapped to items in the semantic backbone.

Pragmatic standards are needed. Unfortunately, semantic standards such as ISO15926 have not lived up to their promise as tools for integration and data exchange. This is partly due to a lack of tool support and to a choice of technology, Express, that became dated and hampered progress. We also believe that the standard development attempted to be comprehensive at the expense of solving smaller, realistic problems along the way. Semantic models used by our group aim to take the best from ISO15926 where it is possible. We also will take account of the industry initiatives in DEXPI (www.dexpi.org) and CFIHOS (http://uspi-global.org/).

Good semantic models can also address the usability problem. Mapping data to concepts that are used by the end-user allows automatic generation of graphical interfaces that meet a specific user's needs. The OptiqueVQS (Soylu et al., 2018) framework is just such a tool. Our research group is working further with this approach to implement faceted search (Klungre & Giese, 2017).

A successful digital twin also requires the correlation of structured data from designs, measurements and simulations with unstructured data from logs and documents. This requires the structured data to be interpreted and transformed into structured data in the process model. There are now many commercial and academic tools for parsing and processing text. The IBM Watson framework is one such commercial offering (Chen et al. 2016) and can be used to parse and process texts. However, language algorithms that are trained on general data sets do not perform well when confronted with oil and gas terminology. A process of domain adaption is needed to improve performance. This process is made difficult by the smaller corpuses of data that are available for training. The SIRIUS centre is working on this domain adaption challenge, with promising results and good performance in solving standard challenge problems (Nooralahzadeh, Øvrelid and Lønning, 2018).

The challenges of maintenance and computational overload can also be addressed through using formal methods to design and monitor the *deployment* of a digital twin. As we noted above, a digital twin is a collection of interacting computational components, deployed across one or more cloud platforms and including edge devices. The behaviour of this system is

https://doi.org/10.3384/ecp181531

difficult to predict, especially at design time. However, simulation tools for the computer systems themselves can be used to test different deployment plans and resolve challenges. The same model can also be used as a monitoring tool for the deployed system – a digital twin of the digital twin (Johnsen, Lin & Yu, 2016).

Finally, there remains the challenge of uncertainty, validation and data science. The digital twin is built on models. To quote George Box (Box, 1979), all these models are wrong, but some are useful. A digital twin will contain many models. Some will be based on physical principles: structural, geometrical and process simulations. Others will be purely empirical, based on machine learning. These models must be validated against observed facility behaviour and aligned so that they mirror observed normal behaviour. Aligning models to observed data is difficult and remains an art. Finding out whether a discrepancy is due to an error in data, a wrong parameter, poor model structure or an actual malfunction in the facility requires a good understanding of the facility and well-developed judgment. This is true whether the models a rigorous physical model or a machine-learnt empirical representation.

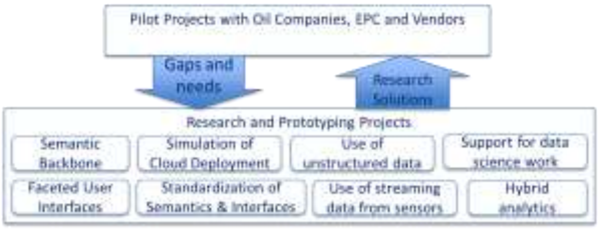

Figure 4. A research agenda for digital twins in the oil and gas industry.

A maintainable digital twin will contain structured tools that allow validation and tuning of all the models in the system. We believe that hybrid analytics – the combination of data science with physical and engineering simulations – is a valuable and fruitful area of research. Machine learning can benefit from being constrained by the laws of physics, while the laws of physics contain parameters that are uncertain or expensive to measure. Good statistical practice is needed in the engineering communities and engineering knowledge is needed among data scientists.

We are working on two problems related to data science for digital twins. The first of these is related to data access. Data science projects in industry are currently not scalable. Each new implementation needs to start from scratch, finding data, checking it and making it available. Our proposed semantic backbone will allow data science solutions to be rapidly transferred to similar sites in an organization (Kharmalov et al., 2017).

The second area of research is related to the use of sensor data in data science. When a data scientist talks about streaming data, they usually mean a sequence of discrete event records – like tweets or sales transactions. The stream of data from analogue sensors is subtly different. The underlying signal from a sensor is continuous. The process of digitization itself introduces uncertainty and error in the calculation. Filtering and data compression provide further sources for error. Common data science frameworks expect data in vectors at common times. Production of this from a time series data base requires interpolation. All these details increase the cost and decrease the usefulness of data science work. A well-defined semantics and query tool for time-series data from sensors could solve many of these challenges.

Companies in the oil and gas sector are installing digital twins now, using commercially available platforms and siloed applications. This provides academics with an opportunity to engage with the observed problems of our colleagues in operations and maintenance. This means that we need a research program that engages with operations and today’s digital twins through pilots.

Each pilot has a narrow enough focus to be tractable. The companies we are collaborating with have linked these installations to a well-defined business case. Current pilots are ambitious: if successful they will bring previously unachieved levels of interdisciplinarity, effectiveness and access to data in design, operation and maintenance. At the same time, the pilots are focused on one specific business problem. By working with existing pilots and proposing new pilots we plan to establish a virtuous cycle, where shortcomings in today’s technology and methods can be filled with research-driven innovations.

6 Conclusion

This paper has attempted to give a synthetic overview of an important, if hyped, element of today’s digitalization landscape. The digital twin is, in many ways a rebranding of several generations of on-line systems for simulation and decision support.

Marketing and vendor communication is raising expectations about digital twins. We welcome this interest and believe that it gives an opportunity to integrate operations technology and information technology. However, we also believe that implementing digital twins is inherently difficult. Solving the challenges that we have identified will require the best efforts of many computer scientists, data scientists, engineers and managers. The role and importance of simulation can only be expected to grow in the next decade. Our research agenda can help ensure that we do not disappoint our managers.

Acknowledgements

The SIRIUS centre is funded by Norwegian Research Council, project number 237898. It is co-funded by partner companies: Statoil, Schlumberger, DNV GL, IBM, SAP, Computas, OSIsoft, Numascale, Dolphin Interconnect Solutions, Kadme, Evry and Fluid Operations AG.

We thank the participants from SIRIUS partner companies and others at the workshop on digital twins held in Oslo on 20th March 2018.

References

Alam, K.M. & A. El Saddik. 2017. "C2PS: A Digital Twin Architecture Reference Model for the Cloud-Based Cyber-Physical Systems." IEEE Access 5:2050-2062. https://doi.org/10.1109/ACCESS.2017.2657006 .

Box, G. E. P. 1979, "Robustness in the strategy of scientific model building", in Launer, R. L.; Wilkinson, G. N., Robustness in Statistics, Academic Press, pp. 201–236.

Cai, Yi, B. Starly, P. Cohen, & Yuan-Shin Lee. 2017. "Sensor Data and Information Fusion to Construct Digital-twins Virtual Machine Tools for Cyber-physical Manufacturing." Procedia Manufacturing https://doi.org/10.1016/j.promfg.2017.07.094.

Carvajal, G., Maucec, M., & Cullick, S. (2018). Intelligent digital oil and gas fields: concepts, collaboration, and right-time decisions. Cambridge, MA: Gulf Professional Publishing.

Chen, Y., Elenee Argentinis, J., & Weber, G. 2016. "IBM Watson: How Cognitive Computing Can Be Applied to Big Data Challenges in Life Sciences Research." Clinical Therapeutics, 38(4), 688–701. https://doi.org/10.1016/j.clinthera.2015.12.001

DAU. 2018a "DAU Glossary: Digital Twin," https://www.dau.mil/glossary/pages/3386.aspx , accessed 18th March 2018.

DAU. 2018b "DAU Glossary: Digital Thread," https://www.dau.mil/glossary/pages/3385.aspx, accessed 18th March 2018.

DAU. 2018c "DAU Glossary: Digital System Model," https://www.dau.mil/glossary/pages/3384.aspx.

Deaton, D. F., & Kloosterman, J. T., 2007. "Success Stories in Onshore Production Surveillance and Optimization." SPE Annual Technical Conference and Exhibition, 9-11 October, San Antonio, Texas, USA. https://doi.org/10.2118/110574-MS

Grieves, M., & J. Vickers. 2017. "Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems (Excerpt)." In Trans-Disciplinary Perspectives on System Complexity, edited by F.-J. Kahlen, S. Flumerfelt and A. Alves, 85-114. Springer.

Grieves, M.W. 2014. Digital Twin: Manufacturing Excellence through Virtual Factory Replication. LLC.

Hyllseth, M., & Cameron, D. (2003). Operator training and operator support using multiphase pipeline models and dynamic process simulation: Sub-sea production and on-shore processing. In A. Kraslawski & I. Turunen (Eds.), *Computer Aided Chemical Engineering* (Vol. 14, pp. 425–430). Elsevier. https://doi.org/10.1016/S1570-7946(03)80152-9

Johnsen, E. B., Lin, J.-C., & Yu, I. C. 2016. "Comparing AWS Deployments Using Model-Based Predictions". In *Leveraging Applications of Formal Methods, Verification and Validation: Discussion, Dissemination, Applications* (pp. 482–496). Springer. https://doi.org/10.1007/978-3-319-47169-3_39

Kharlamov, E., Mailis, T., Mehdi, G., Neuenstadt, C., Özçep, Ö., Roshchin, M., … Waaler, A., 2017, "Semantic access to streaming and static data at Siemens.", Web Semantics: Science, Services and Agents on the World Wide Web, 44, 54–74. https://doi.org/10.1016/j.websem.2017.02.001

Klostermeier, R., Haag, S. & Benlian, A.

Digitale Zwillinge – Eine explorative Fallstudie zur Untersuchung von Geschäftsmodellen

HMD, (2018), 55: 297. https://doi.org/10.1365/s40702-018-0406-x

Klungre, V.N., Giese, M. 2017 *A Faceted Search Index for OptiqueVQS*. CEUR Workshop Proceedings.

Lee, Jay, Hung-An Kao, and Shanhu Yang. 2014. "Service Innovation and Smart Analytics for Industry 4.0 and Big Data Environment." Procedia CIRP 16:3-8. https://doi.org/10.1016/j.procir.2014.02.001.

Lu, Yang. 2017. "Industry 4.0: A survey on technologies, applications and open research issues." Journal of Industrial Information Integration 6:1-10. doi: https://doi.org/10.1016/j.jii.2017.04.005.

Lunde, G. G., Rudrum, G., Angelo, P., Holmas, K., & Setyadi, G. R. 2013. "Ormen Lange Flow Assurance System (FAS) - Online Flow Assurance Monitoring and Advice". Offshore Technology Conference. https://doi.org/10.4043/24297-MS

Mayani, M. G., Svendsen, M., & Oedegaard, S. I. (2018). Drilling Digital Twin Success Stories the Last 10 Years. Presented at the SPE Norway One Day Seminar, SPE: Society of Petroleum Engineers. https://doi.org/10.2118/191336-MS

Nooralahzadeh, F., Øvrelid, L., & Lønning, J. T., 2018. "SIRIUS-LTG-UiO at SemEval-2018 Task 7: Convolutional Neural Networks with Shortest Dependency Paths for Semantic Relation Extraction and Classification in Scientific Papers. *ArXiv:1804.08887 [Cs]*. Retrieved from http://arxiv.org/abs/1804.08887

Panetta, Kasey 2017, Gartner Top 10 Strategic Technology Trends for 2018., https://www.gartner.com/smarterwithgartner/gartner-top-10-strategic-technology-trends-for-2018/, accessed 23rd April 2018.

Panetta, K., 2016, Gartner Top 10 Strategic Technology Trends for 2017., https://www.gartner.com/smarterwithgartner/gartners-top-10-technology-trends-2017/ , accessed 23rd April 2018.

Parrott, A., & L. Warshaw. 2017. “Industry 4.0 and the digital twin - Manufacturing meets its match.” Deloitte University Press.

Poddar, T. 2018. “Digital Twin Bridging Intelligence Among Man, Machine and Environment”. Presented at the Offshore Technology Conference Asia, OTC: Offshore Technology Conference. https://doi.org/10.4043/28480-MS

Renzi, D., Maniar, D., McNeill, S., & Del Vecchio, C. (2017). Developing a Digital Twin for Floating Production Systems Integrity Management. Presented at the OTC Brasil, OTC: Offshore Technology Conference. https://doi.org/10.4043/28012-MS

Rosen, R., G. von Wichert, G. Lo, & K.D. Bettenhausen. 2015. "About the Importance of Autonomy and Digital Twins for the Future of Manufacturing." IFAC-PapersOnLine 48 (3):567-572. doi: https://doi.org/10.1016/j.ifacol.2015.06.141 .

Sankaran, S., Wright, D., Gamblin, H., & Kumar, D. (2017). “Creating Value by Implementing an Integrated Production Surveillance and Optimization System – An Operator’s Perspective.” Presented at the SPE Annual Technical Conference and Exhibition, Society of Petroleum Engineers. https://doi.org/10.2118/187222-MS

Schroeder, G.N., C. Steinmetz, C.E. Pereira, and D.B. Espindola. 2016. "Digital Twin Data Modeling with AutomationML and a Communication Methodology for Data Exchange." IFAC-PapersOnLine 49 (30):12-17. https://doi.org/10.1016/j.ifacol.2016.11.115.

Sharma, P., Hamedifar, H., Brown, A., & Green, R., 2017. The Dawn of the New Age of the Industrial Internet and How it can Radically Transform the Offshore Oil and Gas Industry. Presented at the Offshore Technology Conference, OTC: Offshore Technology Conference. https://doi.org/10.4043/27638-MS

Soylu, A., Kharlamov, E., Zheleznyakov, D., Jimenez-Ruiz, E., Giese, M., Skjæveland, M. G., … Horrocks, I., 2018. “OptiqueVQS: A visual query system over ontologies for industry”. Semantic Web, Preprint, 1–34. https://doi.org/10.3233/SW-180293

Tao, Fei, Jiangfeng Cheng, Qinglin Qi, Meng Zhang, He Zhang, & Fangyuan Sui. 2017. "Digital twin-driven product desing, manufacturing and service with big data." International Journal of Advanced Manufacturing Technology 94 (9-12):3563–3576. https://doi.org/10.1007/s00170-017-0233-1

Tao, F., and M. Zhang. 2017. "Digital Twin Shop-Floor: A New Shop-Floor Paradigm Towards Smart Manufacturing." IEEE Access 5:20418-20427. https://doi.org/10.1109/ACCESS.2017.2756069 .

Trappey, Amy J. C., Charles V. Trappey, Usharani Hareesh Govindarajan, Allen C. Chuang, and John J. Sun. 2017. "A review of essential standards and patent landscapes for the Internet of Things: A key enabler for Industry4.0." Advanced Engineering Informatics 33:208-229. doi: https://doi.org/10.1016/j.aei.2016.11.007.

Uhlemann, Thomas H. J., Christian Lehmann, and Rolf Steinhilper. 2017. "The Digital Twin: Realizing the Cyber-Physical Production System for Industry 4.0." Procedia CIRP 61:335-340. doi: https://doi.org/10.1016/j.procir.2016.11.152 .

Uhlmann, Eckart, Eckhard Hohwieler, and Claudio Geisert. 2017. "Intelligent production systems in the era of industrie 4.0 - changing mindsets and business models." Journal of Machine Engineering 17 (2):5-24.

West, T.D., and M. Blackburn. 2017. "Is Digital Thread/Digital Twin Affordable? A Systemic Assessment of the Cost of DoD’s Latest Manhattan Project." Procedia Computer Science 114:47-56. doi: https://doi.org/10.1016/j.procs.2017.09.003 .

Yun, S., J. H. Park, and W. T. Kim. 2017. "Data-centric middleware based digital twin platform for dependable cyber-physical systems." 2017 Ninth International Conference on Ubiquitous and Future Networks (ICUFN), 4-7 July 2017.