From Big Data to Smart Data

Oliver Breuer*

FLUIDON GmbH, Jülicher Str. 338a, 52070 Aachen, Germany*
E-Mail: oliver.breuer@fluidon.com

Industrial Internet of Things (IIoT) and Industry 4.0 are very popular buzz words today. The „me too" factor is pretty high and attracted companies are faced with an overwhelming market of data management solutions. But despite the large amount of data that can be collected from industrial facilities, the real benefit is behind colourful graphics and charts. To get there, the data provided by the connected components of an IIoT capable system has to be analysed and put into context. So, the question is not what can be done with all the collected data but how to generate useful information.

Keywords: Digital Twin, Simulation, IIoT, Model-based Systems Engineering
Target audience: Digitalization, Connectivity and Communication, Design Process

1 Introduction

Monitoring industrial systems is no new invention at all. Equipping a system with a few sensors and logging their data has been best practice for decades. A more sophisticated approach is condition monitoring as an early approach for predictive maintenance, years before it became one of the most popular catch phrases for IIoT.

The real differences between conventional and smart systems are their grade of connectivity and the simple existence of certain sensors due to their integration just into the used components. It means that more components have more to say and are provided with a forum to do just that. That sounds quite simple, but distributing data in an industrial environment is a challenge for various reasons, including security restrictions.

Beyond the challenges that data acquisition entails, data processing is the key to further benefits. IIoT provides contextual analysis of the same data in different domains and levels.

2 Smart Data

Frequently, measured data only becomes useful information in the broader context. This transition requires some kind of intelligent processing.

2.1 A trivial analogy

A car driver is an excellent example of a human data processor. His task is to use all relevant information provided by his senses to drive the car safely. The differences between experienced and unexperienced drivers are

- the definition of "relevant information",
- the definition of "safe operation" and
- the individual grade of control over the system.

Unexperienced drivers are often overwhelmed by the amount of available information. Traffic, road condition, weather, traffic signs, pedestrians, buildings, even the radio programme – all this will be noticed by this driver and taken into account. The data analysis is very time-consuming, and hopefully his skills are sufficient to react adequately.

By contrast the experienced driver is able to filter the incoming information a lot better. He knows the relevant bits in the data stream that affect the behaviour of his vehicle. The reason for that is that he has got a much better idea of how his car reacts and what the requirements for a safe journey are. Objects beneath to the road are, for example, no threat as long as they do not intend to move towards the car's heading. Moisture on the road is more dangerous in leafy corners than on motorway straights, and so on.

The experienced driver has developed a model of his environment and the objects he uses and is able to feed it with current data. He knows how slippery a wet, leafy road is and when his car begins to lose wheel grip. Having analysed the relevant data and applied that to his model, he is able to predict how fast he can drive through the next corner without hitting anything.

2.2 Analysing data

As shown in 2.1, the quality of prediction depends on current data, knowledge about the behaviour of a system (the model) and experience. Screening the relevant data and assessing it manually is an appropriate way for a manageable amount of data and simple systems, but it reaches its limits with increasing data amounts and system complexity as occurring when the full potential of new technologies like IIoT is to be exploited. Referring to the driver example, even experienced drivers get overstrained sometimes if the needed information cannot be recognized at once. So, there is the need for an experienced driver with very quick comprehension who can analyse data based on his earned knowledge.

A promising approach to a quick and complete examination of data is automated data processing. A contemporary data processing system can observe millions of values in fractures of seconds, compare them with thresholds and display warnings to the operator. Despite its performance, this approach alone would be insufficient having the large amount of data in mind that can be provided by IIoT systems, at least if one would like to tap the full potential. If the correct relations are established in the system, the correlation of all this data unveils a deeper information layer. The instance that can make these relations is the model.

2.3 Digital Twin

Using a model in automated analysis allows estimating the condition of a subsystem or component, even though it is not directly monitored by dedicated sensors. It acts as a virtual experienced maintenance engineer who can diagnose a system from many malfunctions by just listening to his machine's noises because of the knowledge-based model in his mind. Beyond that, the approaches of both the maintenance engineer's mind model and the digital model can be used for "what-if" scenarios. It means that it is possible to use a model for extrapolating current data based on a behavioural description. An obvious use case is predictive maintenance. This method tries to avoid unexpected system downtimes due to worn-out components by monitoring the trend of certain indicators and providing an estimated remaining runtime until the need for maintenance. The indicators are calculated by a simulation model that is a replica of the real system – a digital twin.

Real-time data processing for a huge amount of data, a digital twin of the whole system – that offers a whole bunch of new possibilities, doesn't it? Being able to predict every future development of any component in the facility sounds tempting but, in fact, that is not the way it goes.

A major critique on simulation is often the lack of completeness. Simulation models are just as accurate as the behavioural description behind and only within its validity range. To achieve a sufficient grade of accuracy, the model has to be very precisely defined. This is directly related to the computational power required for acceptable calculation times. Such an accurate model representing a whole facility with all aspects would therefore be way too slow to provide the current state of the system in time.

At this point, the car driver example in 2.1 could be helpful a last time. The purpose of the driver's model is to ensure a safe journey. That does not include any other coherence like the photosynthetic uptake of the tree beneath the wet corner while the car passes by, emitting carbon dioxide, although this is of course also an important issue.

But the environmental model of our driver has to include the tree's effect on his car when he hits it in case of a miscalculation, not the biological processes of the wooden plant.

This strategy of simplification and concentration on the relevant things can be transferred to the concept of the digital twin. The guiding principle should not be "Process all of the available data". In fact, a definition of what information is useful for an additional benefit should come first. A simulation model for exactly this purpose is most likely computable in reasonable times and much more maintainable because of its size.

2.4 Use case

Suppose a system contains hydraulic cylinder drives. The drives are position controlled and used for material processing where accuracy and speed have top priority and machine downtime has to be absolutely minimized. Power supply is provided by an adjustable pump unit. The layout of the cylinder drive is shown in Figure 1.

The subsystem is equipped with the common sensors that can be found in many similar systems. A PLC processes the relevant sensor data and activates the pump and the servo valve. The operator is quite up to date and has established a pool for all the accumulating data of his facilities, catchy related to as the cloud. Although the components of the described subsystem have been provided by different manufacturers, they are all able to transmit data to the cloud, either via the PLC or via their own interfaces.

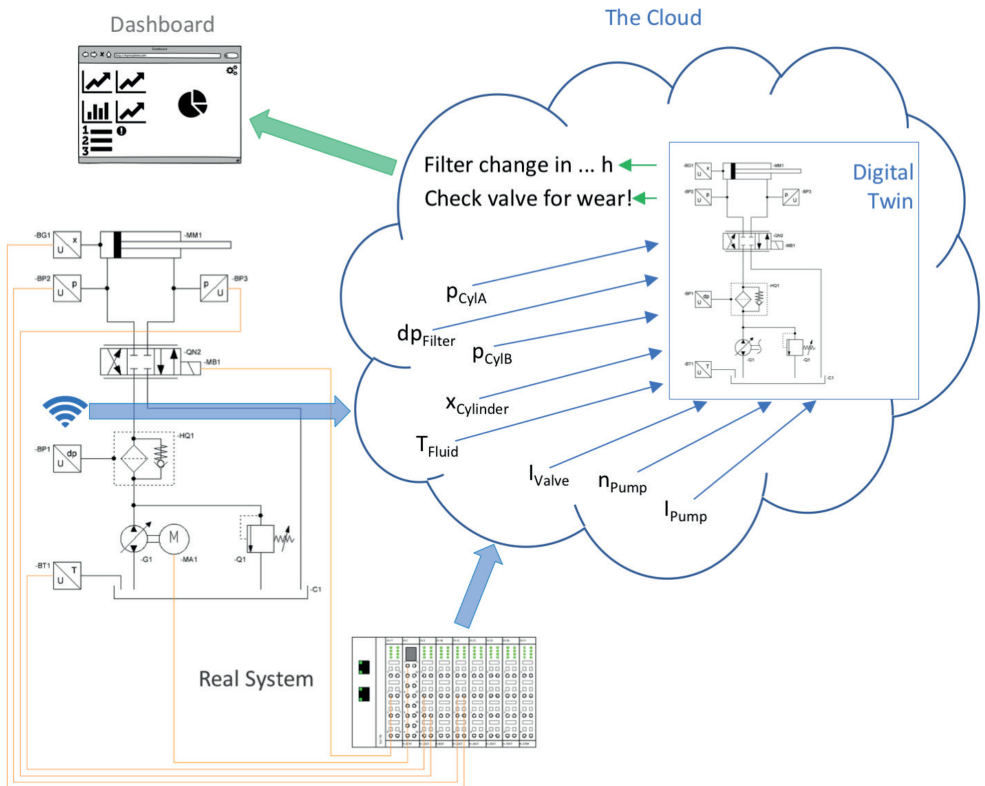

Figure 1: The digital twin as data analyser

As mentioned, the operator is especially interested in keeping the machine running and remaining a high standard of accuracy. Concentrating on the exemplary cylinder drive, there are at least two components that can be related to these requirements: the high-pressure filter and the servo valve.

Some operator's questions could therefore be

- When does the filter have to be changed if we keep the current operating conditions?
- When will the servo valve be too worn-out to keep the tolerances while maintaining the current speed?

or maybe even

- With the current system condition, how fast can we produce without affecting quality?

The filter is a component that has to be maintained frequently, depending on run-time, fluid contamination, or even sudden stress failures, so monitoring this component closely could reduce machine downtime. The servo valve suffers from wear of its metering edges, which has an impact on system performance and accuracy. The similarity between both components is that they cannot simply raise their arm and say "Hey, operator, I'm tired!".

The condition of these components can only be determined by context. The valve control signal value, for example, has to be checked against the current cylinder movement profile and the differential pressure over the filter depends on filter load as well as fluid viscosity. These dependencies are described within the model of the hydraulic part of the cylinder drive, the digital twin, and it is part of the cloud.

3 The Value Chain

The creation of digital twins, as well as the creation of simulation models in general, requires know-how and data. That can be a problem in terms of intellectual property, because the plant manufacturer is not necessarily the manufacturer of the used components and could therefore have limited access to required data. One way to deal with this information gap is to estimate the missing data based on experience or context. Another way is to add new processes to the component manufacturer's value chain.

3.1 Model Exchange

An obvious alternative is the deployment of the physical component along with the corresponding digital twin by the component manufacturer (Figure 2). In this case the twin is technically a boxed simulation model with a defined interface. This is not a new approach to the exchange of models, in fact it is one of the initial drivers for the Functional Mock-up Interface (FMI):

> "The FMI development was initiated […] with the goal to improve the exchange of simulation models between suppliers and OEMs." (Modelica Association, /1/)

So, using the FMI protects intellectual property but allows the usage of a model that has been created by the manufacturer itself and should therefore be as accurate as possible. The model, according to the FMI standard called a FMU (Functional Mock-up Unit), becomes part of the digital twin that is located in the plant operator's cloud.

3.2 Component as a Service

The new flexibility offered by IIoT also allows an outsourced analysis. The component manufacturer gets access to certain values in the cloud of the plant operator and analyses the data using his own component twin. This approach works best when the component does not have too many logical dependencies on the rest of the system.

Combined with a maintenance contract, the plant operator actually pays more for the function provided by the component than for the component itself (Figure 3). A further aspect is the possibility of continuous adaptation and improvement of the component due to the manufacturer's access to real-time operating conditions. He could therefore install an improved version of his component at the next maintenance appointment. So, although the exchange of data with third parties is a major problem for many companies, both sides can benefit.

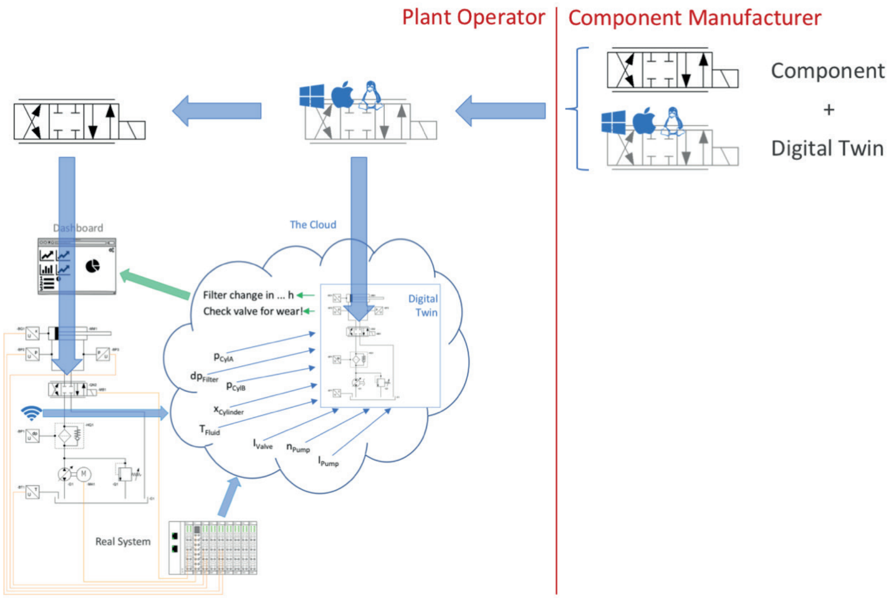

Figure 2: Model Exchange

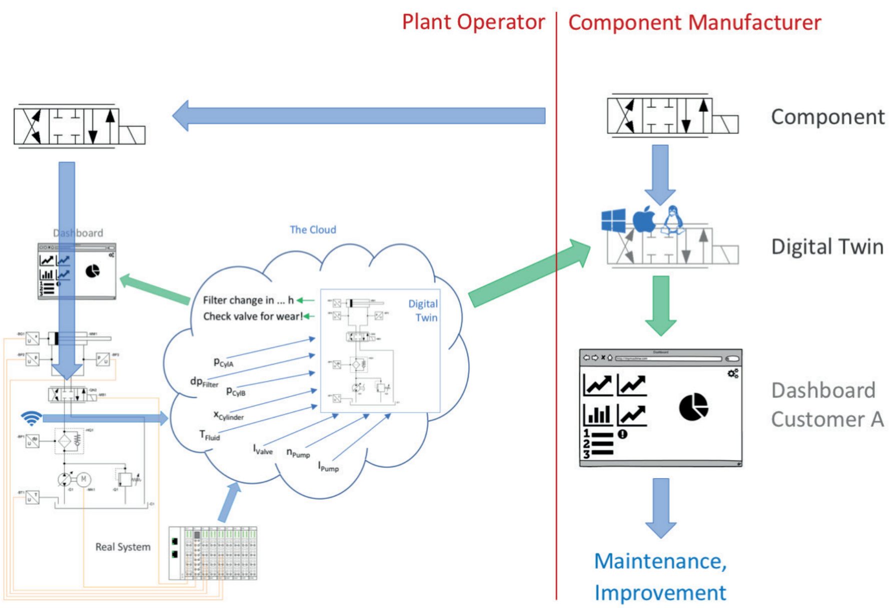

Figure 3: Component as a Service

4 Digital Thread

Working with digital models is an enabler for a connected data flow through design, development and operation of a system. In the past, simulation models where mostly used during the product development phase. Digital twins remove this limitation. Not only that these particular models are used in parallel to the real system's operation phase, they can also serve as prototype for a new clone to be used for future developments, being constantly fed with real-time process data. This opens up the possibility of being able to carry out performance and wear behaviour studies during development on the basis of current data.

5 Conclusion

Simulation models are no longer just development tools. They can accompany their real counterpart over its entire lifetime. The digital twins serve as reference and test bench, they transform raw data into useful information. The particular challenge with digital twins is not so much the creation of accurate models as the derivation of current states based on measured data. The biggest challenge, however, is to even get close to data, at least for third party cloud services. The distribution of data to the other side of the company site wall causes an uneasy feeling for many plant operators.

References

/1/ Modelica Association, *Functional Mock-up Interface*, <http://www.fmi-standard.org> (11.01.18).