Filling the Industrial Data Lake in Manufacturing Companies Based on KPIs

An efficient implementation of digitalisation projects in manufacturing is highly challenging. Usually a transformation to digitalisation based on a technology-oriented perspective leads to vast investments and implementation costs. In order to exploit the potential of digitalisation entirely, defined digitalisation targets derived from an overall corporate strategy are required. This paper describes a new systemic methodology based on a generalised automation pyramid allowing the optimal placement of required sensors in a production line based on highest impact key performance indicators (KPIs) identified from the corporate strategy broken down to production level. The new approach is demonstrated on the hand of the vial production line of Sola Messwerkzeuge GmbH exploiting synergy effects systemically achieving cost efficiency.

Keywords: Digitalisation, industry 4.0, KPIs, data analysis, systems engineering

Introduction

Production planning and control as well as quality management in a production facility always require up-to-date production data. While information is always available in the same, high-resolution cycle in continuous or sequential systems, object-related processes in logistics or manufacturing industry are characterized by random events. For example, employees arrive at a plant at different times in the morning, or the material arrives at the production facility at exactly the time it is processed - just in time or as an enhancement: just in sequence. This reduces stocks in production and thus avoids unnecessary investments or costs. Especially because the proportion of non-value-added activities is minimized.

With the introduction of these principles, it is no longer possible to exactly calculate the moment an event will take place - like the arrival of material. Such systems are not deterministic and follow stochastic distributions. Non-deterministic systems can only be automated with stochastic models. The quality of these models depends here on the accuracy of the identification of underlying distributions in a production line. In practice, that identification is highly challenging as the broad necessary database is usually not available. And if figures are available, these distributions are highly time variant, e.g. with product variance.

In many companies, the status quo is usually a manual collection of operational data - either in paper or in digitalised form. The required data is entered manually into a form and confirmed or signed. This usually works well until problems in production occur. In the latter case the forms are often handed in collectively, i.e. at the end of a shift. And this means that there is no longer a sufficiently precise data basis for an optimal production planning or control. However, to achieve the paradigm shift to Industry 4.0, where the product determines the path through
production, an order must necessarily be clearly identifiable and precisely traceable at any time.

Companies facing that challenges with automation, digitalisation and big data technology are driven by various external and internal factors such as e-commerce, increased customisation combined with reduced batch sizes or cost reduction and competitiveness.

These drivers have different influences on the company’s motivation for digital transformation. Especially hygiene factors have an essential impact. In this context, hygiene factors, referred to product perspective, represent an economic imperative for implementation, otherwise the product cannot be sold anymore. As a consequence, customers are not willing to pay for automation. In order to remain competitive, manufacturers are forced to implement automation without being able to generate financial added value by the end customer, e.g. by increasing sales prices. Other hygiene factors are, for example legal regulations or increased customer demands.

That paper addresses that challenges by introducing a systemic methodology compromising technology and costs against a company’s strategy using only highest impact production indicators. The described methodology was developed with Sola Messwerkzeuge GmbH in Austria. The paper shows the application of the methodology exemplary on the hand of a production line for vials as part of Sola’s spirit level production.

In a former project, Sola targeted that challenge like many other companies driven by technology. A data lake should be implemented sourced by ident sensors providing track and trace for the product focussing a part-production line. The suppliers offer regarding the required industrial-standard sensors with their related hard- and software equipment plus further required touch screens led to high estimations of investment costs – compared to the benefit. As for tracking parts and tracing processes, a cost-effective, reliable and industry-compatible technology is required. In the beginning, several technologies for part tracking and tracing have been investigated with regard to their usability, applicability and profitability in the considered context. As a result, RFID technology has been identified as being well suited but involving very high investment costs. Further investigations show that the benefits of this technology are limited to little advantages specific for the focussed part of the production line.

That has stopped the project – again like in many other companies. However, there was a demand for a more methodical and application-oriented approach for standardized use in the complete companies’ production lines.

Based on the actual intention of that first project – turning the event driven production line deterministic by generating production line data in real time – the systemic methodology was developed. The paper shows the exemplary application of the methodology on the example of one KPI and starts with a determination of the most impacting parameter targeting the strategic goals driven here by hygienic factors of the market. By exploiting the information technology driven layer principles of the automation pyramid that goal is broken top-down to the digital twin of the vial production line to determine the required locations of sensors and
appropriate sensor types. A following bottom-up approach through the information layers allows a further costs consideration. That process repeated in an iterative manner to find the optimal trade-off between corporate strategy and location of sensors in the production line by identifying its required production data resolution by the described iterative process leading to a cost-effective choice.

By applying that methodology consequently to other related production data indicators, the already determined devices can then be exploited optimally leading to an adequate generation of key performance indicators of production, keeping databases or data lakes lean and that under minimal implementation and maintenance cost.

In the exemplary case of Sola, the shown key performance indicator extracted from the corporate strategy could be implemented even without placing further sensors in the production line. The methodology showed, that the indicator could be implemented by simple correlation of already acquired data from the database. Further, that data can be easily augmented by few ident sensors in the production line to extract other key performance indicators.

**Basics of Digitalisation and Automation - Impact by Technology and Costs**

Industry 4.0 in production is driven by several technological developments accompanied by mass application and the related cost reductions of the required automation technology.

Before digitalisation and automation efforts are implemented in a production line it is recommended to analyse their potentials in the specific production. Figure 1 illustrates the implementation of automation in production in a stepwise approach, by

1. **Organising workflows and workstations,**
2. **Mechanising work processes and finishing with**
3. **Automated production.**

Usually, these steps are summarized by the term “OMA-principle”.
Figure 1. General Procedure of Production Automation

1. Derivation of digitalisation targets from corporate strategy.
2. Organisation of material and information flows.
3. Identification of activities and classification into CI's.
4. Digitalisation of work instructions.
5. Determination of current degree of automation.
6. Determination of required sensor technology.
7. Mechanisation of work processes.

Determination and weighting of automation potentials due to target definition.
Specification of material and information flows, routes and layouts for each workstation. Calculation of additional parameters, e.g. by using key figures.
Collection of activities and classification into small units, called configuration items (CI's). CIs are used to manage production processes, e.g. if the capacity utilization of certain workstations is too high, CIs can be shifted quickly to other workstations with less capacity utilization.
Derivation of work instructions from CIs. Transfer, integration and storage of optimized organisation in IT/worker leading system.
Evaluation of automation potential through analysis and optimisation of process flows in order to reduce costs and avoid errors.
Investigation of suitable sensor quantity, type and placement in order to record production data and generate key figures.
Optimisation and improvement of worker's manual production activities. Derivation of required actuator technology for mechanising workstations, e.g. pick-by-light systems, automated material delivery or supportive assembly devices.
Analysis of potentially used automated machines. Specification of measures for further automation taking the defined targets (step 1) under consideration. Assessment of profitability and comparison with alternatives.

Source: Adapted from [1].

As shown in step 6, selection of suitable sensors and the generation of operational and business data is an essential part of production automation procedure. Information from ident systems are the basis for that step. These systems are a decisive development in sensor technology for improving the automatic acquisition of operating data in Industry 4.0 applications. They allow a clear identification of objects such as workpieces, tools or transport containers. For the hardware-technical implementation, a unique identifier or marking for identification must therefore be attached to the object. That identifier can be en masse printed, engraved or similar on objects, e.g. in the form of bar-, UDI- or QR-codes. Or in an intelligent variant in the form of a programmable data carrier with radio technology. Then further information can be transferred and read out contactless and dynamically on the object. This means that further information is also available when there is no contact to the database, e.g. in logistics.

By linking to a database, additional information, such as the corresponding customer order, is also available.

Since this database is filled permanently with data from sensors in production, the amount of information in this database is constantly increasing. Therefore, that database it is also called a data lake. The goal of a data lake in production must be to have a complete digital image of the physical system from which all the required data can be derived. In that case this data lake forms a complete digital twin over the entire lifecycle, from factory planning to product ideas and the disposal of products.

In a data lake, a wide variety of data such as technical information, production and quality data or information from the supply chain is collected by the various IT
systems involved and checked for consistency and correctness. This step is necessary to comply with requirements such as single source of truth. Only in that case, it can be ensured that the stored inventory is generally valid and therefore reliable at a certain point in time.

Data in the data lake can be further used in shop floor analytics to make system dynamic predictions. Such a forecast could be the identification of demand. It could also take into account special events that will generate exceptional demand. For example, a major sporting event that will create exceptional demand for television sets. This requires system dynamic models which can be adapted based on historical data - self-learning and improving and in further developments even combined with artificial intelligence (AI).

With such predictive methods in demand management, delivery reliability can be improved while reducing reserve stocks at the same time. In addition, these methods can also be used as a smart service to manage customer inventories.

However, the data lake is not only crucial for data storing, analysis or future smart services – it also links the IT world with automation technology. The role of production and its automation in digital transformation is undisputed. Körner et al. [2] argue that automation “requires multiple IT-systems to cope with resulting challenges due to complex manufacturing systems” [2]. The merging of IT (virtual world) with classical industrial processes (physical world) by embedded hardware and software systems leads to Cyber-Physical Systems (CPS) that demonstrate the core elements of smart factories. This evolutionary, technological transformation represents the vision of Industry 4.0 [3].

At that point - where physical systems meet information technology - it must be noted that the prioritization of the requirements of classical IT clearly differs importantly from that of automation technology. In classical IT, the highest priority is given to confidentiality. This means that data may only be read, transmitted or modified by authorized users. In second place are the requirements for integrity: data must not be changed unnoticed and must be consistent. In third place is the availability, the prevention of system failures or access to data within a certain period of time. In automation technology, the exact opposite ranking applies, which requires i.e. other safety concepts.

For a CPS, it is therefore decisive whether it is derived from classic IT or automation technology. Notwithstanding the non-hierarchical approach of CPS, the established model of automation pyramid “to structure the different applications in a functional and hierarchical manner […] is still present and very common” [2].

The automation pyramid is based on various information layers commonly found in IT systems combined with the hardware-based approach of automation technology. The number of hierarchy layers of the pyramid varies, as in research some layers are removed or merged together [2]. In the given context, the automation pyramid is considered with three layers for simplification.
With the increasing use of sensor technology in production, the number of data stored will also rise rapidly. A concept or ontology must be elaborated at an early stage to determine not only the type (in what form) and place (where) data will be stored but also a time variant rating or importance of the data. In the case of central databases, for example, a distinction can be made according to the time at which the data is stored and the associated number of server accesses. It must also be taken into account that the data can have different meanings and must be evaluated appropriately. But over time, data can also become meaningless or even harmful. If data is randomly written to a database without a structure, a data grave is quickly created in which no user can find or assign data. The hope in practice is then often that at some point an AI or other big data analysis tools will be invented and developed that structure and analyse the chaos. However, this is currently not foreseeable and does not correspond at all to the systems dynamics-based principles of lean production: avoiding waste [4].

An important role in production planning plays a well-defined, manageable set of KPIs. They are the basis to bring the principles of lean thinking to databases allowing an easier organization and assignment of data by keeping communications and the number of database entries slim. Also, that reduces the maintenance effort of the database - see also the takt principle of lean manufacturing.

The Role of KPI and Key Figures in Digitalisation

The previously mentioned lean approach for filling the data lake with production data must be transferred to the generation of KPIs. Finding the KPI with the highest impact on a target of the corporate strategy is crucial for the company’s success. This importance of KPIs and machine availability data (MAD) is confirmed by several authors. Schuster [5] highlights the importance of continuous measurement, monitoring and reporting of key figures like machine availability (MA) or Overall Equipment Effectiveness (OEE) for efficient production management (planning and control). Moreover, he suggests using supportive IoT-solutions that prepare and provide real-time machine and production information for variance/root-cause analysis and deriving measures [5]. Kang et al. [6] describe KPIs such as efficiency, availability
or throughput as “critical for manufacturing operation management and continuous improvement” [6].

As this paragraph shows, in literature and practice, the terms “key figure” and “KPI” are often used synonymously. Thus, a clear definition of both terms is required.

According to the International Organization for Standardization (ISO), a key performance indicator (KPI) is defined as a “quantifiable level of achieving a critical objective” [7]. Kang et al. [6] take up this definition and add that KPIs are “a set of quantifiable and strategic measurements […] that reflect the critical success factors of an enterprise” [6]. Moreover, they demonstrate that “KPIs are not independent and may have intrinsic mutual relationships” [8].

When differentiating KPIs and key figures, the terms “strategic measurement” and “a company’s success” are crucial, as these reflect the main characteristics of a KPI.

In the context of this paper, KPIs are related to a strategic context by indicating the degree of strategic target achievement. With reference to the definition of indicators, a KPI consists of two components:

- measurable: key figures based on specific, quantifiable values.
- non-measurable: indices, like corporate reputation, employee motivation or satisfaction [9].

That means, KPIs should essentially be based on key figures and rely as little as possible on indices. The key figures are measured whereas the indices are extracted from the data lake.

For example, the above mentioned MA represents a typical key figure which can be measured. Whereas the superordinate OEE as KPI is built out of the product of the figures for availability, effectiveness and quality ratio. Hence, the MA influences as part of availability the OEE.

However, it must be mentioned that different definitions of MA exist in literature – having the same key statements [6, 8]. MA represents production data by expressing the percentage of actual operating time (AOT) relative to planned operating time (POT) [8].

Due to that reliance on the MA e.g. Boschi et al. [10] define the MA as a key business factor (KBF) and a KPI, too, because it considers “the production parameters used to evaluate the production trend and the process behaviour […]” [10].

Those relationships and interdependencies of KPIs with system quantities and between other KPIs are discussed in further detail in Stricker et al. [11]. Their paper establishes relationships between several “KPIs in the field of maintenance and machine monitoring” [11] with MA and OEE under consideration.
Methodology

The newly developed methodology describes a systemic approach for generating a KPI systemically broken down from corporate strategy level and implemented on production level with adequate hardware and software components.

The methodology is characterized by an iteration loop-based approach with the aim to structure KPI's in order to identify the KPI with the greatest effect on achieving corporate strategy's objectives.

Its aim is to determine the clear definition of optimal sensor placements for filling the data lake, i.e. generating the required data at production level such as machine data, by considering implementation costs and other driving factors for data analysis like resolution, update rate, margin of error.

As previously described, KPIs have relationships and interdependencies. Thus, another aim of the framework is to implement the generation of related KPI deviated from the existing sensor data at production level without generating additional effort or costs.

The methodology is developed with respect to the automation principle and its prioritisation of availability, integrity and confidentiality. As a consequence, requirements like single source of truth principle can be fulfilled even in combination with an IT driven data lake. In combination with the IT and automation technology approach of thinking in hierarchical information layers the costs of investment can be controlled.

For that reason, the core characteristic of the previously described automation pyramid with its multi-dimensionality is exploited. In contrast, the generic literature rarely considers that multidimensionality as shown in Figure 3 (cf. [2, 12]).

Figure 3. General Use of Automation Pyramid as Multidimensional Hierarchical Layer Model, Linking Process Driven Automation Technology and IT

The focus of this multidimensional pyramid illustration is the key subject of the developed methodology – the systemic and iterative generation of KPIs by applying and executing a stepwise reference modell of the pyramid in a top-down and bottom-up approach.
In the context of this paper, layer I represents the lowest level, where the actual manufacturing process with its physical sub-processes “controlled by devices and sensors” [2] is imaged. Simultaneously, layer I is the foundation for the upper generation of KPIs. The multidimensionality of layer I reflects the real-world manufacturing processes. Subsystems represent, according to their respective systemic definition and interpretation, e.g. individual machines of the production line or warehouse, machine groups or complete production lines. Events then are states that trigger the transmission of data and communicate the results of a process step, for example signal of completion or arrival of material. Events are thus initiated by a trigger signal, e.g. timestamps at the beginning or end of processing.

Besides manufacturing operations control, layer II integrates the transmitted data from layer I and external data from other systems (MES, ERP from layer III). This data is gathered over the entire system, or here production landscape or network, for generating the KPIs. In this layer, the sensor values recorded in layer I are compared with the target values. The data measured in layer I is converted to integrated data, which is required for the calculation of the KPIs (layer III). Layer III, the top level, represents the target level and defines the short- and long-term objectives. It is linked to the ERP system and entails the definition, calculation basis and visualisation of the KPI, derived from the long-term corporate strategy and digitalisation project objectives. Körner et al. [2] highlight that in practice, business planning in this layer is executed by applying ERP tools for managing tasks, inventory and resources.

Systemic Procedure for KPI Generation

The newly developed systemic methodology is characterized by its idea to generate KPI systemically by abstracting the general pyramid hierarchy to the specific required information for KPI calculation with a top-down and bottom-up approach meeting the corporate strategy. This methodology retains the strict, hierarchical procedure from IT (see Figure 4).

Figure 4. Automation Pyramid with its Multi-Dimensionality and Derived from Overall Corporate Strategy
Basis for the implementation of the methodology is a clearly defined, abstract project target. The targets of the digitalisation projects are derived again from the objectives of digitalisation. This requires a holistic digitalisation strategy that is aligned with the corporate strategy and vision of the company. In order to verify the target achievement of the digitalisation projects, it is necessary to determine and generate specific KPIs. The stepwise application of the pyramid model enables the methodological generation of such KPIs by defining and specifying the requirements for KPI generation. A systemic approach to the generation of KPIs is essential in order to identify and assess the overall relationships and correlations between them. As a result, before applying the methodology, potential correlations between KPIs must be considered. Thus, similar KPIs can be generated additionally with little effort by using synergies in the collection of data for calculating key figures and KPIs.

Although the methodology is characterized by a systemic manner, the procedure for applying the concept is described stepwise. Another important characteristic of the methodology is its user-driven data collection and generation. User-driven means that data generation is not randomly performed by the subsystems themselves but is selectively and solution-oriented initiated by the end user. This is an important prerequisite for keeping the data lake lean and controllable. This is supported by an automatically generated KPI specification data sheet during application of that methodology.

**Step 1: Derivation of the KPIs from the Strategic Objectives**

After determining the corporate strategy and deriving strategic projects with specific project targets, the first step is to identify the KPIs to be determined, which are derived from the general strategic objectives of the company. An important requirement for KPI definition is that the specific KPI is measured with at least one key figure. Thus, the key figure to be determined has to be defined and abstracted from the KPI. Secondly, the purpose and goal of KPI generation and key figure calculation is described in this step. With regard to this, the linking context and reference between key figure and determined goals as well as additional key figures is created. Therefore, related key figure and KPI are gathered.

**Step 2: Determination of the Basic Structure of the Pyramid**

In this step, each layer of the pyramid is elaborated. Starting with layer III, the KPI is defined and the calculation basis for KPI and key figure is determined. For layer II, the data for calculation are described. The data are differentiated between required data for calculation on layer III and measured data from layer I and external systems, like from MES-/ERP systems such as transaction-related data or master data. This step completes with the determination of data types for data collection on process level (layer I).
Step 3: Identification and Specification of Iteration Criteria

In order to take into account the effort required to implement the KPI generation system and the requirements for the KPIs in terms of time resolution and update rate, it is necessary to define iteration criteria for the respective KPIs. With these criteria, the KPI generation is evaluated in several iteration cycles and a decision on termination or further optimization is made. Therefore, the criteria must be differentiated into soft and hard criteria. Soft iteration criteria are merely guiding and non-quantifiable, whereas hard iteration criteria are quantified, i.e. they are measured with a unit and allow a clear decision about an abortion, continuation or completion of the iteration process. Hard iteration criteria are either of variable character, i.e. all values can occur and are acceptable, or of decisive character, i.e. there is a fixed target value to be reached. During this process step, a unit and a start or target value will be assigned to the hard iteration criteria.

Step 4: Top-down Approach – Definition of KPI Requirements for Each Layer

In that step, the requirements for real-time performance and resolution are developed for each layer. Real-time behaviour is usually distinguished between soft, hard or no real-time. The hard real-time is in the safety context, i.e. a failure to meet this time requirement leads to damage, e.g. high costs or useless values. In contradiction, a violation of the soft real-time does not lead to additional costs. Thus, hard real-time is mainly necessary at the lower layers of the automation pyramid (for operational control). It is needed for information evaluation, so that the information can be time stamped. A soft real-time is sufficient for planning activities.

Update rate and resolution are decisive inputs for the sensor level. The resolution refers to the object being viewed and represents, for example, the unit that is measured, e.g. individual part, tray or pallet.

The update rate refers to the time period of the observation: e.g.: on an hourly, shift-related or daily rate.

These factors represent aspects of data processing and data storage by providing information about

(1) how many measuring points are desired (resolution)?
(2) how often should be measured (update rate)?

These criteria are essentially determined by the characteristics of the technologies (e.g. sensors) used. For this reason, the logically shortest detection interval on the one hand and the critical minimum detection interval on the other hand must be determined for the resolution.

Finally, the margin of error and the trigger response have to be defined. Margin of error also affects the selection of the sensors.
Trigger response represents the origin signal for a layer’s processing or detecting activity, i.e. a layer is triggered by an upper (top-down) or a lower layer (bottom-up). Additionally, the trigger response can be initiated actively or passively.

**Step 5: Bottom-up Approach – Description of Layer Functionality in Detail**

In this step the data to be recorded is specified. For this purpose, the data including a description of the criteria is compiled in a list and compared with the data from the systems (e.g. MES or ERP). The data is also re-examined for the use of synergy effects in order to be able to generate new KPIs as required with minimal effort. In addition, the data still to be generated is identified.

For layer I, it is specified how often, where and what kind of data is generated, i.e. with regard to the type of data defined in step 2.

In layer II, it is defined what kind of data are required for KPI calculation, e.g. data from subsystem (layer I), cross-subsystem data (from layer III), external data from other systems (ERP, MES). Furthermore, methods for possible verifications of the data, e.g. plausibility checks or cross-checks are determined. To conclude the specification of layer II, the integration of the internal IT, e.g. other departments or into existing systems is investigated.

This step is finalized by giving details about the functionality of layer III. This includes the definition of filters for modifying the KPI, e.g. by merging several subsystems and for collecting additional information e.g. for application such as interfaces or data formatting, storage and integration. In the end of this step, the list of related KPIs, created in step 1, are investigated in terms of their completeness and correctness and updated if necessary.

**Step 6: Establishment of Quantity Structure from Defined KPI**

The iterative application of the methodology is finalized within that step. The results of this step indicate if an iteration cycle is completed, and the results fulfil the iteration criteria.

In this step, the requirements for the investigated KPI are summarized and the costs and efforts for the implementation are examined in detail. This step represents a feedback loop to review bottom-up the performed steps regarding the initially defined goals and strategic objectives and the defined budget.

For this purpose, a list is established that

(1) summarises the data required for the calculation including a description of the criteria in detail (unit, resolution, update rate, real-time behaviour, origin, etc.),

(2) checks which data are already available in the MES and whether the defined criteria are fulfilled,

(3) designates the remaining data to be generated, including a description of the method of collection, and
(4) estimates the costs and efforts for implementing the KPI system i.e. for purchasing and installing the required sensors, integrating data into existing IT, managing and analysing the data lake and for ongoing operation such as service and maintenance of hardware, software infrastructure and system merging.

In other words, in this step, the generated data is linked to a cost function in order to evaluate it in monetary terms and to determine when this data becomes redundant. This linkage to costs is an essential step to maintain the lean structure of the data lake by deleting or updating non-valuable data.

**Step 7: Evaluation and Assessment of Results**

With a final evaluation of results, it is decided whether the stepwise application of the methodology is completed, aborted or if a further iteration is needed, where e.g. the objects under consideration are revised with respect to the number of subsystems or kind of events, real-time requirements or the margin of error are changed, or the quality of the resolution and update rate is adjusted.

In the last two steps, it is also necessary to consider and comply with the previously described requirements for automation projects, e.g. lean data lake, data consistency and single source of truth.

As the production key figures are in many cases derived from similar data bases, synchronisation with other digitalisation subprojects is possible and recommended. In turn, this means that the implementation of the first subproject is the most cost-intensive regarding the required sensor technology.

In conclusion, the methodology aims to determine

(1) what data are required to extract the desired information from the data lake?
(2) whether order tracking or detailed tracking of individual components of an order is being pursued?
(3) in which time periods/intervals signals are to be recorded and data generated by the sensors (per second, minute or hour, per cycle, shift-specific, once a day, etc.)?
(4) how many sensors are needed to generate the required data?
(5) what characteristics the sensors must have (cost, quality, industrial suitability)?
(6) how the acquired signals are transmitted (via Bluetooth, via bus interface, according to which standards and with which protocols)

Therefore, the results of the iterative application of the methodology show which data are needed from the production, in which order, how they are composed and in which temporal resolution they have to be generated.

The ranking then shows which digitalisation subproject should be carried out first as well as the number of sensors required to achieve the target. With consistent use
of incremental data processing the e.g. the number of sensors can be reduced significantly once again in a further iteration cycle.

Exemplary Implementation of the Methodology

The methodology has been developed in collaboration with the company Sola Messwerkzeuge GmbH, headquartered in Götzis (Austria). Sola is an internationally operating SME specialized in the production of high-precision measuring tools. With its 200 employees and subsidies in Germany, Hungary, Switzerland and the US, the company develops, produces and distributes a comprehensive product portfolio with approx. 1500 products in more than 70 countries all over the world.

With regard to software that is required for data collection, analysis and evaluation, the company already uses comprehensive MES- and ERP-systems. The integration of the systems is planned for the next two years. Thus, there is already a solid database for production data implemented, e.g. machine availability data (MAD) or feedback data transmitted to the system for production data acquisition (PDA).

Description of Sola’s Production Processes

In order to explain the methodological generation of the key figure MA in the subsequent chapters by using a real-world scenario from the company Sola, firstly the sample production process under consideration is described.

The total production line is built up by two main sections. The vial production (VP) as the first section, takes place independently of the spirit level production (SLP). SLP is the second section and is physically separated from the VP.

The two sections are connected by a Kanban system. The Kanban system is filled manually by an employee using a Kanban card when a requirement is reported.

The plastic injection-moulded parts for the VP are initially produced in a separate plant of Sola in the same town. In addition to caps, lids and the end caps for the spirit levels, the transparent acrylic glass vials are also produced there. Special demands are made on the high accuracy and the high quality of the acrylic glass (very clear surface) as well as the precise dimensions.

Figure 5. Sample Production Line for Vial Processing at Sola
In the VP, different machines are required for the processing steps, visualized in Figure 4.

The first processing step is the vial shaping for processing the inner radius and markings (step S1). Next, the vials are filled with a liquid (step S2). Then the first test procedure (step S3) takes place at two different workstations in order to ensure that the high requirements mentioned above are met. Using a special manufacturing process, the vials are closed (step S4). This is followed by another high-precision quality inspection (step S5). Finally, the assembly of the modules takes place (step S6).

The finalized modules are stored and placed in the Kanban system for the subsequent SLP or for delivery.

As an innovative and competitive SME, Sola recognized the need to promote digitalisation to maintain productivity and efficiency, and to react flexibly in order to meet the customer’s diverse demands and develop high-precision and optimised products [13]. Sola’s drivers for digitalisation are summarized in Table 1.

<table>
<thead>
<tr>
<th>Drivers and objectives</th>
<th>Challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hygiene factor</td>
<td>Competitiveness</td>
</tr>
<tr>
<td>Strategic objectives</td>
<td>Improved process stability and transparency; increased process availability; reduced quality losses.</td>
</tr>
<tr>
<td>Strategic project</td>
<td>Process optimisation through reduced quality losses at a high degree of capacity utilization; data analysis and reporting through generation of defined key figures/KPIs.</td>
</tr>
<tr>
<td>Project targets</td>
<td>Traceability of the production process through standardization, documentation and evaluation of KPIs; tracking of individual parts (regarding inventories) in production.</td>
</tr>
</tbody>
</table>

Table 1 reflects the initial situation for the development of the described methodology for systemically generating and implementing a corporate key performance indicator (KPI) system.

The previously mentioned concept with a focus on application-orientation in comparison to a technology-oriented approach is described in Table 2.

<table>
<thead>
<tr>
<th>Technology-orientation</th>
<th>Application-orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td>Selection of adequate technology.</td>
</tr>
<tr>
<td>Step 2</td>
<td>Identification of appropriate application for selected technology at production level.</td>
</tr>
<tr>
<td>Findings</td>
<td>High costs for investment and implementation; RFID-technology not beneficial.</td>
</tr>
</tbody>
</table>
Sample Application of the Methodology

Based on the systematic approach described in Chapter 3, this chapter shows the systemic application of the developed methodology on the hand of the vial production at Sola. Congruently with that comes a proof of concept with regard to its applicability and problem-solving contribution.

The specific application of the pyramid model in a top-down and bottom-up approach according to the previously described steps, aims to increase process stability and avoid high quality losses in the long term by introducing digitalisation.

The associated track and trace implementation will improve the extracted values and their reliability. Particular attention must be paid to the relationships and interdependencies between the key figures. At the same time, the aim is to save costs during implementation (through the consideration of key figures’ interrelationships).

Step 1: Derivation of the KPI from the Strategic Objectives

In accordance with the stepwise procedure described in chapter 3, the corporate strategy at Sola aims at increasing competitiveness in the long term as well as continuous quality improvement. Both objectives are reflected by the companies specific KPI OEE – to be more precise, in its term “plant productivity” (see Figure 6).

In the case of Sola, a prior performed bottleneck analysis already clearly identified that in VP section, process stability is determined by the machines. Therefore, the machine availability (MA) has the dominating impact compared to the total production availability on the KPI and is therefore chosen as measured key figure and is calculated as followed:

\[ MA = \frac{\text{actual operating time (AOT)}}{\text{planned operating time (POT)}} \]

Hence, the purpose of this subproject is to increase data quality and interpretability of PDA and maximize MA. Related KPI is OEE, related key figures are performance, quality and downtime.

On the one hand, POT represents the difference between calendar time and planned time for non-production, e.g. time for maintenance and service. On the other hand it is calculated as the sum of operating time and availability losses such as setup times or downtimes.

The key figure MA further is applied for validating and verifying the feedback data and for synchronising it with the production cycle. This leads to an increased interpretability and transparency for production management. Thus, this key figure has a high practical relevance. In addition, with little effort, further key figures can be generated from the MA and KPIs such as the OEE itself can be specified in figures.
Step 2: Determination of the Basic Structure of the Pyramid

The application-specific structure of the pyramid is illustrated in Figure 6. The key figure machine availability (MA) is distinguished in linked and individual MA. Layer III calculates the linked MA in vial production according to the individual MA. When considering the availability of the overall production line (linked MA), the calculation of the key figure MA results from the individual MA.

Figure 6. Structure of the Pyramid in the Context of Sola's Vial Production for Calculating Linked Machine Availability (MA)

The individual and linked MA are a function of the specific structure. The structure reflects the order and layout of the machines, i.e. in serial sequence (multiplication of individual MA values) or in parallel order (summing up of individual MA values). The structure also refers to the processing steps. For each processing step the structure of the machines have to be defined. Therefore, each individual machine here represents a subsystem and the merging across the structure (parallel or sequential) takes place in layer III. In order to generate useful results, a clearly defined meaning of the term “structure” is required. The calculation of individual MA takes place in layer II, whereas the planned operating time (POT) is delivered from external systems (PPS) and the actual operating time (AOT) is delivered from each machine autonomously. That means, layer I delivers each individual machine status to layer II for calculation. Machine states are differentiated in “machine is running”, “machine is stopped”, “machine is in malfunction”. Malfunction means, there is no OPC signal available. The real AOT is derived from machine status “machine is running” and defined by the time, the machine is actually producing parts actively. Ramping up or waiting for new material to arrive are not included in AOT. Layer II transfers the calculated
individual MAs in a standardized form to Layer III, as a basis for the calculation of
the linked MA on layer III.

Figure 5 summarizes all steps of the methodology applied to the Sola sample.
Step 1 and step 2, described previously in detail, are illustrated on the right side of
the figure, followed by step 3.

**Figure 7. Overview of Performed Steps for Exemplary KPI Generation for
Machine Availability at Sola**

Step 3: Identification and Specification of Iteration Criteria

Several iteration criteria for top-down and bottom-up approach for Sola’s
example have been identified.

The type of criteria is either soft or hard, as explained in chapter 3. A unit is
assigned to each hard iteration criteria (see term in brackets) as well as a
differentiation is made between variable (v) and decisive character (d). A target
value has to be added to the decision criteria.

For the sample set of Sola, the two most important iteration criteria,
amortisation of the implementation and resolution have been selected.

Amortisation of implementation refers to costs in relation to benefit. The costs
for implementation include the installation costs for sensor technology (hardware)
as well as the expenses for database integration and maintenance as well as
assigned, proportionate costs for already existing software architecture and systems
(e.g. PDA in MES system)

The benefit refers to the improved transparency in production, which makes it
easier to optimize production processes. The benefit is more difficult to quantify
than the costs. Therefore, an estimation of benefit is essentially necessary here. For
calculation the amortization it is important to have the same unit for both, costs and
benefits.
The amortization is a function depending on resolution and quality. This signifies that the resolution has a fundamental influence on the amortization. For example, with increased resolution, i.e. resulting in more installed sensors, the costs increase proportional. This implies that it must be determined what benefits can be derived from an improved resolution. As a consequence, resolution is identified as a variable criterion (v), being adjustable in relation to each subsystem. In contrast, amortisation is a decision criterion (d), measured in years and used for countercheck. The amortisation’s target value for indicating a successful iteration cycle has to be less than three years.

Step 4: Top-Down Approach – Definition of KPI Requirements for Each Layer

Starting in Layer III, the requirements for the KPI are defined. The real-time performance of the key figure "linked MA" in layer III is analysed as soft real-time. This means that real-time is required for the calculation of the key figure, but the time of value delivery is not that sensitive. The decisive factor is that the value is delivered. For example, for linked MA it does not have a huge negative effect whether the value is available at 6:00 pm or 6:30 pm, but a value delivery at 00:00 am is not acceptable. Interval is differentiated in logically shortest detection interval and critical minimum detection interval. For layer III and linked MA, the shortest detection interval is per second, as it is determined by the lower layers. The critical minimum detection interval depends on various influencing factors for the benefit. For example, if the shift influences the benefit of the implementation and thus the amortisation of implementation, it is necessary to evaluate the key figure once per shift. In the case of Sola, it is necessary to generate the key figure at least once per production batch. This also influences the update rate. Therefore, update rate is set to hourly, which means with every change in production batch. Margin of error refers to error tolerance. If error tolerance, as an iteration criterion, is set to 99%, the acceptable margin of error is 1%. For layer III, trigger response "top-down, active" means, that this layer actively requests new values for calculation from subordinate layer II.

This procedure is repeated for layer II and layer I. In layer II, the individual MA in standardised form is calculated. Its real-time performance is determined by the real-time performance of superordinate layer (layer III). Trigger response of layer II is "top-down, passive" because layer II does not send the information to layer III without a request.

Here can be highlighted that the newly applied methodology shows for layer I that no further sensor devices are required here for fulfilling the iteration requirements as the machine state signal is generated by the machine autonomously and actively forwarded to layer II for calculation. Hence, trigger response is “bottom-up, active”. 
Step 5: Bottom-Up Approach - Description of Layer Functionality in Detail

Starting with layer I, for calculation of the individual MA, the machine state (e.g. “machine is running”) has to be delivered to layer II once per subsystem. In layer II, data from several sources is gathered, i.e. machine state from layer I and information about planned operating time (POT) from production planning and steering (PPS) system. This layer also required a definition of possible plausibility checks for data verification. For the example of individual MA, the machine state “machine is running” can be counterchecked on the one hand, by validating the machine state in accordance with the output counter signal. If machine is in state “machine is running” and parts are produced, the output counter signal has to increase. On the other hand, by summing up the three different machine states at the end of a day, the total time must meet exactly 24 hours.

For layer III, potential filters summarise information for further derivation, e.g. by gathering same types of machines or machines of the same processing step, an overall linked MA can be calculated to extend the expressiveness of the key figure.

With this step, the exemplary explanation of the developed methodology is completed.

If the definition of the criteria and the top-down and bottom-up approach is successful, implementation can be started immediately. If iteration criteria are not yet met, new iteration is necessary by adjusting one iteration criterion and applying the iterative method (top-down and bottom-up) once again in order to complete the KPI definition.

Conclusions

In many manufacturing companies, the users of data proclaim, that they do not have sufficient and transparent data for interpretation, analysis and evaluation. As a consequence, the users often start to place sensors by gut feelings at production level that generate more and more data and fill the data lake unstructured. Several sensors then deliver same or redundant information or, if they have a different resolution, they deliver different data at different time stamps. This leads to an exponential growth of databases as well as increasing costs and effort, and reduced controllability and transparency of the data lake. And single source of truth principle can’t be warranted.

In the last years, Sola implemented, besides the existing MES and ERP systems, in a structured manner a system for machine status feedback using the OPC-UA standard. On this base a highly cost-effective solution for KPI generation by application of the new systemic methodology was found in collaboration.

The data lake mentioned in the title of this paper connects two worlds – the world of the process and its availability, and the world of IT with confidentiality as the highest priority. As previously described, for keeping the data lake small, lean and controllable, it is mandatory to have a systemic and stepwise procedure, based on production automation principles and applicable to each company-specific use.
case by bringing the hierarchical levels to IT and leading to an optimised choice of sensors investment in order to keep the payback time.

Derived from the overall corporate strategy, the novel methodology and its requirements of data consistency and single source of truth principle, simplifies the integration of the KPI into existing systems like MES or ERP. Thus, filling the industrial data lake takes place in a methodological approach. This procedure delivers a specification for KPI generation. Therefore, the developed methodology provides a useful tool to generate KPI in compliance with the requirements for data analysis and storage.

The application of the methodology achieves among other things the following findings:

(1) how many sensors are needed where at the lowest level?
(2) which processing programs are needed for the integrating layer (is MES or ERP needed and if so, which one?)
(3) what does this implementation cost, such as sensor installation or software integration?
(4) what specification does that KPI need to fulfil?

At Sola, the initiation phase for calculating the linked MA as a basis for generating the KPI OEE is performed. Currently, the installation of sensors at production level is completed and the integration of the generated data into the existing systems is in progress.

It was found out, that the application of the methodology leads to a target-oriented, solution-driven and cost-effective implementation.

Due to that, a Sola internal guideline describing the developed methodology was established [14]. In the next step, the methodology is applied to other sections of the production line as well as to the other departments. Finally, the KPI specification data sheet can be used as blueprint for further derivation or modifications of related KPIs.

References


