Tool Chain to Extract and Contextualize Process Data for AI Applications

Lincoln Sherpa*, Ralph Müller-Pfefferkorn, Udo Enste, Gregor Tolksdorf, Michael Kawohl and Michael Wiedau

DOI: 10.1002/cite.202300004

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Summarizing a key use case of a research workstream of the German publicly funded KEEN project, methods and tool chains are demonstrated to extract and to contextualize process data in an automated way based on engineering information. The contextualized process data serves as a high-quality data source for machine learning methods. The article covers the applied basic methodical approaches, design decisions and the results of a successful pilot installation of the developed tool chain.

Keywords: Artificial intelligence, Data contextualization, Data management, Process industry, Tool chain

Received: January 10, 2023; *revised:* March 09, 2023; *accepted:* April 03, 2023

1 Introduction and Problem Statement

The overall goal of the research activities presented in this paper is to develop new system integration approaches to decrease configuration efforts for extraction and context-based data handling of device information and production data needed by artificial intelligence (AI) solutions in process control. One use case of a profitable AI-based application addressed in this paper is the calculation of indicators for real-time improvement of energy usage in process plants using surrogate models as baselines for proper guidance of energy efficiency. The addressed problem to be solved is the engineering effort to use this known approach for all relevant energy managed units (EMUs) (e.g., industrial complexes, process plants, plant units) within a production site. In this paper a methodical approach and tool chain will be presented by a use case to minimize the engineering effort generating and maintaining a huge amount of surrogate models, necessary for a comprehensive view on all relevant EMUs of a production site.

1.1 Surrogate Models for Energy Performance Analysis

Increasing energy and resource efficiency is one of the major challenges in process industry. The normative standards around ISO 5000x [1] describe a plan-do-check-act (PDCA) cycle to establish a comprehensive energy management process within a production site. To identify improvement potentials, suitable energy performance indices and calculation methods considering meshed energy flows must be used. In addition, the user has to verify a continuous performance improvement. Therefore, the energy performance indicator (EnPI) must eliminate non-influenceable aspects like outside temperatures or economic-driven fluctuating plant loads in order to verify improvement measures. New approaches using context models and data-driven surrogate models representing normalized EnPIs baselines give an answer to these tasks [2].

In [3], a new approach was presented to find suitable and consistent EnPIs that fulfill the requirements of correction of indicators by the influences of ambient conditions or process utilization. Two methodologies were introduced and combined: First, using data-driven surrogate models, so called best demonstrated practice (BDP) curves, which can be used as baselines to compare calculated standard EnPIs with values in the past under similar process conditions. This approach was presented in more detail in [4]. Secondly, an aggregation method was proposed to exploit process and energy information within a plant and process hierarchy (bottom-up approach) in order to support root cause analysis in case of decreased efficiencies. Both methodologies can be used in real-time to steer a process in a most efficient way.

[1]Lincoln Sherpa (lincoln.sherpa@tu-dresden.de), [1]Ralph Müller-Pfefferkorn, [2]Udo Enste, [3]Gregor Tolksdorf, [3]Michael Kawohl, [3]Michael Wiedau https://orcid.org/0000-0001-7153-7292

[1]Technische Universität Dresden, Centre for Information Services and High Performance Computing, Chemnitzer Straße 46b, 01187 Dresden, Germany.

[2]LeiKon GmbH, Kaiserstraße 100, 52134 Herzogenrath, Germany.

[3]Evonik Operations GmbH, Paul-Baumann-Straße 1, 45772 Marl, Germany.

 © 2023 The Authors. Chemie Ingenieur Technik published by Wiley-VCH GmbH

ISO 50001 demands the use of EnPIs, which have to be compared with an energy baseline. As stated before, one suitable approach for a baseline is using a BDP curve, fitted as a data-driven surrogate model, which describes the best historically observed performance of an EMU (e.g., industrial complex, process plant or plant unit) as a function of non-influencing factors (see Fig.1) [5]. This curve can be used to provide valuable information on the efficiency of the current operation of a plant and the improvement potential to plant managers and operators. The modeling method of the BDP is based solely on the evaluation of historical process data and employs a novel combination of known surrogate modeling techniques using clustering, model fitting and model simplification by backward elimination. The method can be applied to real process data of large industrial production sites. A surrogate model is a parameterized mathematical structure (black-box model) representing an input-output behavior. Such approximation models only represent the data that has been used to parameterize them, so extrapolation usually is not possible and their accuracy depends on the density, accuracy and reproducibility of the available data.

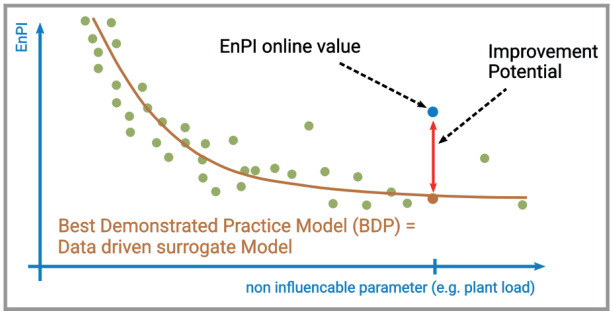


Figure 1. Best demonstrated practice build as a surrogate model.

The methodology to get these kind of normalized efficiency indicators is described in [6]. First of all, a suitable indicator has to be chosen. Usually, the EnPI is used, which is defined as

$$EnPI(j) = \frac{E(j)}{m(p)} \tag{1}$$

where $E(j)$ denotes the energy demand and $m(p)$ the amount of product during a specified time interval. j denotes the type of the energy source.

The generation of a BDP model is a multistep approach:

1) data acquisition,
2) raw data pretreatment,
3) data clustering,
4) model identification using surrogate modeling techniques, and
5) the BDP usage by providing dashboards for plant manager and operators showing the actual improvement potentials.

The method can be used for real-time process monitoring. By visualizing the improvement potential of the current operating point, performance deviations are identified and can be reduced by the plant personnel.

The concept is also suitable for reporting according to ISO 5000x [1]; especially the requirements of ISO 50003:2021 [7], which demands the verification of a continuous improvement of the energy performance. The concept was applied amongst others to process data of a large production plant of INEOS in Cologne. It has been used and enhanced in practical use from 2018 up to now.

1.2 Engineering Effort to Build Site-Wide KPI Applications

The analysis of material and energy efficiency is commonly based on evaluating balance volumes. Therefore, the knowledge of the energy flow network at a site or plant, including the energy that is bound in material flows, is a matter of special importance and relevance. Besides the amount and direction of flows including internal recycles, information is required relating to flow categories, related measurements and possibly additional characteristics like the enthalpy of products. In addition, generic calculation algorithms and material or energy type specific master data constants, like enthalpy of material and utilities have to be linked to the data elements that represent the energy flows. The goal is to give the available measurement data a process-oriented context in order to obtain additional information based on the structural information. Context awareness is a key element in today's data handling and processing [8].

Based on a structural process-oriented data model, a company can describe the energy and material flows of a site in an intuitive but also formal manner so that all information needed for real-time energy or material performance calculations is merged and structured in one holistic knowledge base. Within such a structural model the nodes of the resulting graph represent a so-called EMU, which typically could be an industrial complex, plant or plant unit and the directed edges of the graph represent the material and energy flows (see Fig. 2).

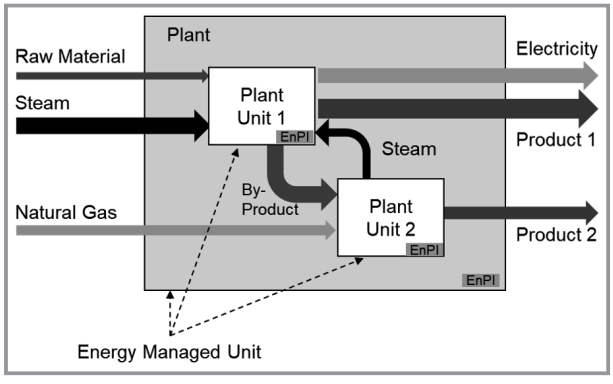


Figure 2. Energy managed units.

 © 2023 The Authors. Chemie Ingenieur Technik published by Wiley-VCH GmbH

15222640, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cite.202300004 by Cochrane Germany, Wiley Online Library on [11/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The generated graph can be used for EnPI calculations but also to calculate missing measurement data or to provide online plausibility checks using material or energy balance equations. This approach was developed and successfully evaluated within the research project MORE [9]. Also, industrial practice showed that this is a suitable approach to satisfy the increasing requirements for energy performance analysis and energy audits based on ISO 5000x. In case of using normalized EnPI based on surrogate models as baselines, first of all a structural process-oriented model of a plant must be generated and secondly for each combination of EnPI and EMU a surrogate model must be calculated. The application of this approach in industrial use cases showed that this unavoidable effort is a high barrier to use this approach for each plant of a big industrial site. Therefore, a more automated engineering workflow and tool chain is needed to decrease the barriers to implement this innovative method enhancing the energy efficiency in process industry.

1.3 P&IDs and Plant Topology

Computer aided engineering (CAE) software plays an important role in the design phase of chemical plants. As described in [10], the asset lifecycle involves of a variety of software tools from the early phases where the *functional requirements* [11] are defined, over the more concrete *functional design* phase and the detailed *asset specification* up to the *assets in operation*. The collection of data during these design phases is important and good data quality does have a huge impact, e.g., for modern applications using AI technologies [12]. One of the most important diagrams in planning and operation of a chemical plant is the piping and instrumentation diagram (P&ID). This diagram combines the key elements of the major disciplines in chemical engineering: The world of apparatuses and machines, the world of piping and piping networks and the world of electrical and instrumentation. This leads to the fact that the P&ID defines the topology of a plant and can be seen as the most important single document (i.e., fixed and structured amount of information that can be managed and interchanged as a unit between users and systems [13]) for semantic relationships between engineering objects in plants.

While digitization addresses the transformation process from printouts (unintelligent documents) to data, P&IDs still play an important role. This is especially true for all the existing (brownfield) plants relying on P&IDs in the paper-sense of documents. In order to make the conversion to digital objects happen and to have a future-proof way of transmitting and storing P&ID information, we see the need of *semantic P&IDs*. The international initiative Data Exchange in the Process Industry (DEXPI) addresses this need by defining a respective standard (see also [14]). In a semantic P&ID, as we understand it, all contained engineering objects and other elements (like graphical symbols including the connecting lines) are instances embedded in an ontology. This comprises in practice at least the following aspects that need to be machine-readable:

- definition of a reference data library (RDL) as an underlying vocabulary for the meaning of classes, attributes, and relationships
- classification of all objects according to the RDL.

This usually leads to implementations that are database-based with all elements being underlying database elements, but other implementations like triple stores or file-based approaches (using, e.g., XML or JSON) are possible as well.

Despite DEXPI defining a standard for semantic P&ID perfectly fitting the requirements derived from the tool-chain presented here, actually making use of DEXPI comes with challenges and limitations. The main limitation is the actual availability of DEXPI P&IDs due to the maturity status of the DEXPI interfaces in default configurations of P&ID software. While several software vendors proved to be able to export P&IDs according to the standard (some even confirmed by certificates [15]), creating DEXPI P&IDs is still not common practice. One main identified challenge related to this is the accuracy or quality to adhere to when constructing the P&ID. It is for example mandatory to connect pipelines to equipment (i.e., lines to symbols) – it is not sufficient when it only looks like they are connected on the visualizing drawing (leading to a wrong plant topology with open pipelines and unconnected apparatuses and machines). This issue can be addressed by raising further awareness and by tool improvements checking for possibly unintended disconnections. In the long run we expect P&ID creators to get used to these P&ID construction quality requirements and therefore see more DEXPI-compliant P&ID exports used in practice.

1.4 Contextualized Data

Contextualized data can be described as data that is put into a context describing the situation the data is related to or was created in. Typical examples of data can be process measurements, as well as master data of technical equipment. The term *situation* comprises facts like the time or location context that can be linked to even more pieces of information (e.g., batch phase, involved resources etc.).

Another aspect of consideration is maintaining the understanding of contextualized data for a prolonged time. But the knowledge about the data gets compromised with respect to time with loss of individual/group memory or movement of staff, which can create problems during its use or while sharing among others [16].

One approach to resolve this problem is to provide a proper structure of the information by using a standard metadata schema [17]. If the available standards do not fit the requirements of needed information models, a specific schema has to be created with inclusion of metadata fields

© 2023 The Authors. Chemie Ingenieur Technik published by Wiley-VCH GmbH

15222640, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cite.202300004 by Cochrane Germany, Wiley Online Library on [11/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

that describe all the required aspects of the data which is relevant for the specific use case. However, in this case, it is always advisable to reference metadata fields from already available standards to minimize interoperability issues [16]. Furthermore, the implementation of such a schema in a data management platform will enhance the FAIRness of the data (according to the FAIR principles [18]).

A more comprehensive description of why there is a need for data context management is given in [8]. The importance of contextualized data in process monitoring and optimization results in the need of an enhanced IT/OT infrastructure of a process plant.

Fig. 3 shows the addressed infrastructure using established IT/OT systems like programmable logical controller (PLC) and decentralized control systems (DCSs) as shop-floor systems as well as process information management systems (PIMSs) and analytics & dashboard solutions supplemented by a contextualization platform, which can include different kind of data models representing context relationships of data.

One source to build these kind of information models is the totality of engineering information, like what is contained on a modern P&ID. Looking at an exemplary P&ID, the following types of engineering information can be extracted:

- equipment information (what types of apparatuses and machines, having what kind of technical data, are used in the plant?)
- plant topology (how are apparatuses and machines connected via pipes?)
- location of sensors in the topology (where do measurements take place?)
- instrumentation IDs representing the role an apparatus or a device like a sensor or actuator will have.

Of course, other typical deliverables of the engineering process (e.g. process flow diagrams (PFDs), 3D models, specification sheets, electrical loops etc.) contain other important engineering information, but the P&ID itself already is a powerful means for giving process measurement data a meaningful context.

Well contextualized data, preferably structured according to available international standards, is of added value not only for human interpretation, but most importantly for the application of machine learning [12]. In this article we will describe a tool chain concept that demonstrates how dashboards, visualizing relevant key performance indicators (KPIs) that are calculated with the help of an AI, can build on the correct extraction and contextualization of engineering information and process data, in our use case given by a P&ID and a PIMS system.

2 The Tool Chain

The AI-based application for real-time energy improvement calculations addressed in the previous chapter needs a structural, best possible hierarchically subdivided context model of the meshed network of energy flows within a production site. The main information to build this kind of data model can be deviated by CAE data, especially by P&ID diagrams. Usually PFDs, which will typically be designed before a P&ID will be specified, would also be a suitable source of context data, but in industrial use, these kind of diagrams are often not maintained accurately enough during the life cycle of a plant.

The ultimate goal of the addressed tool chain in this paper is to deliver automatically generated dashboards for energy performance evaluation of the main, most important EMUs provided with minimal effort in system setup by re-

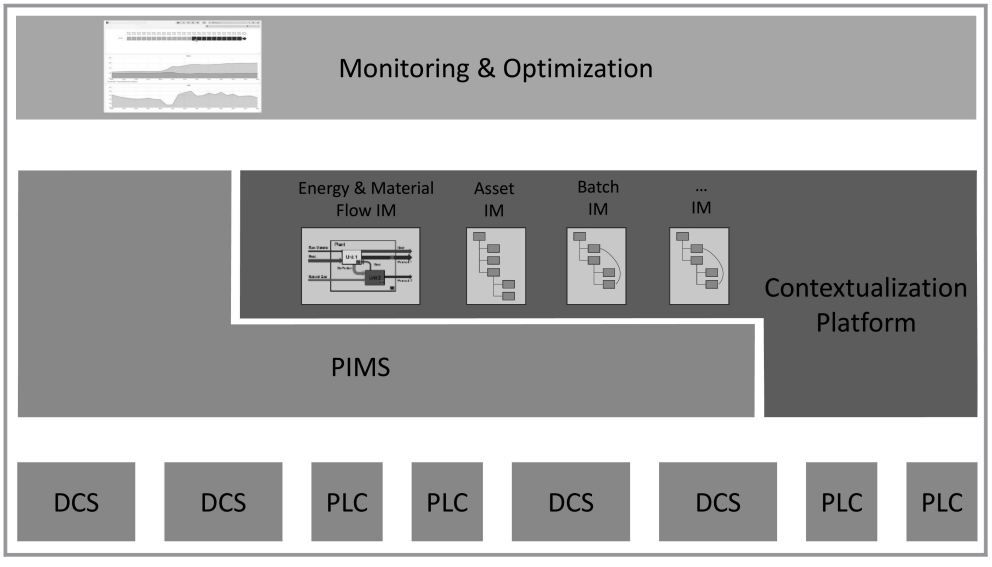


Figure 3. General infrastructure architecture concept of IT/OT systems.

 © 2023 The Authors. Chemie Ingenieur Technik published by Wiley-VCH GmbH

15222640, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cite.202300004 by Cochrane Germany, Wiley Online Library on [11/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

use of digitally available engineering information. The term 'tools' comprises algorithms/techniques as well as common software solutions with user interfaces. The finally provided dashboards can be used by operators, plant operation managers or energy managers of process plants.

As shown in Fig. 4 the starting point of the tool chain is an electronically available P&ID, which is preferably stored using the CAE export standard DEXPI. Assuming that this P&ID is in addition annotated with standardized meta data for plant information, the FAIR principle (findable, accessible, interoperable and reusable) is supported, and the correct P&ID can easily be chosen and automatically interpreted. For the purpose of the tool chain, two main kinds of information are extractable out of the DEXPI files in a system independent format.

First of all, the instrumentation tags of all available measurements can be parsed. If a company uses a consistent classification and naming of measurements in all operation technology (OT) systems, the instrumentation tags within the CAE system will be the same as the names of the variables within the DCS or PLC available by common communication standards like OPC unified architecture (OPC UA). Therefore, the parsed instrumentation tags of the DEXPI file can be used to generate a variable list to build up a PIMS historian of all relevant measurements of a plant or plant unit. Also, necessary master data like the measurement ranges or measurement units can be explored using the DEXPI file or alternatively by exploring the configuration data of the DCS. In case of an inconsistent usage of tag names in CAE and OT systems, a mapping, in a best case a rule-based mapping, must be provided. Based on the automatic configuration of a data transfer between DCS and PIMS, trend data of relevant measurements can be provided from scratch with minimum configuration effort.

As a second result from parsing the DEXPI file, a first structural model of the topology of the material and energy flows as described in Sect. 1.3 can be generated automatically. Therefore, a more intelligent parser and importer for a contextualization platform is needed, which can interpret common P&ID structures and elements in order to provide relevant nodes, which represent potential EMUs as well as directed edges, which represent the energy flows within the context model. Each edge can be linked automatically to one or several data points of the PIMS using also the data parsed in the DEXPI file. As a result, the formerly non-interpretable data points of the PIMS (like flow measurements) can now be used for data analyses incorporating the added knowledge of the sources and sinks of material and energy flows.

In a third step, the process-oriented context model must be extended by additional data. Based on the requirement analysis of ISO 50001 the most important EMUs must be specified. Therefore, the balance envelop (or balance boundary) of each relevant EMU must be defined. In addition, the knowledge of energy and material categories (e.g., electricity, steam, etc.), which are relevant energy carrier coming into or leaving an EMU including their master data like enthalpy, must be configured.

Now, in a fourth step, the surrogate models can be automatically generated by a model builder. For each specified EMU a surrogate model can be built for each relevant EnPI according to the descriptions made in Sect. 1.1. The model builder uses the knowledge of incoming and leaving energy and material flows of an EMU to pick the dedicated historical data from the PIMS. The model builder just needs the additional information and corresponding addressing path within the PIMS of the non-influenceable parameter, which should be considered by building the best demonstrated practice curve as described in

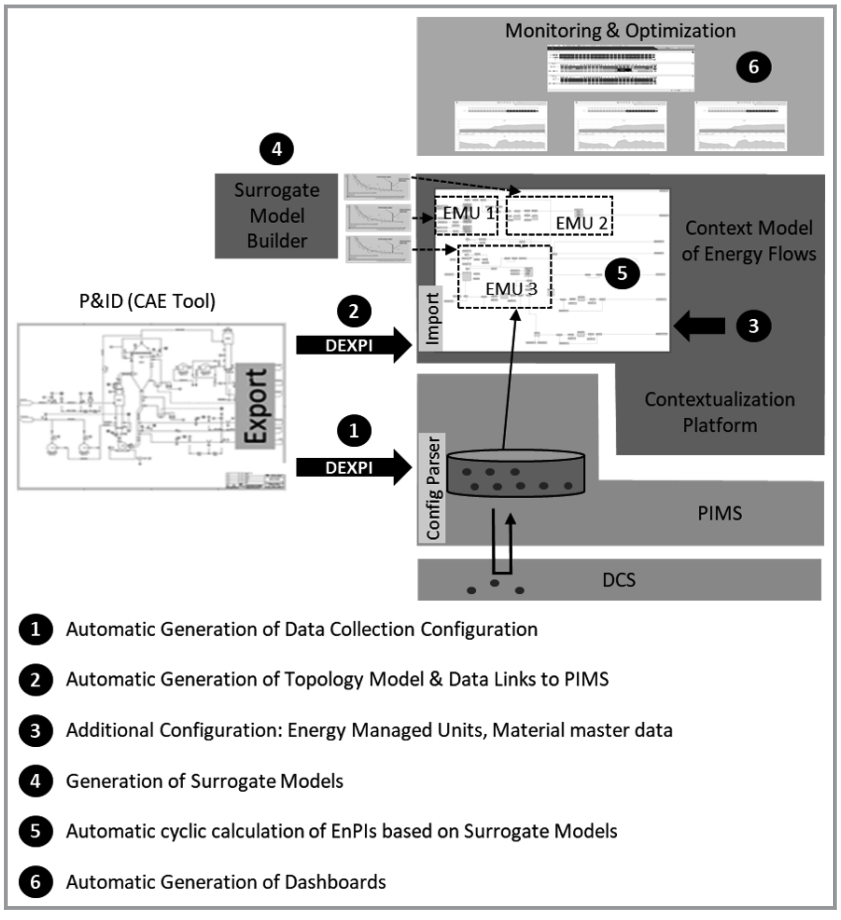


Figure 4. Tool chain to build up AI-based energy performance monitoring dashboards.

© 2023 The Authors. Chemie Ingenieur Technik published by Wiley-VCH GmbH

15222640, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cite.202300004 by Cochrane Germany, Wiley Online Library on [11/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Sect. 1.1. Usually, the surrogate models must be fitted by adjusting the model parameters. The finally built surrogate models will be linked to the corresponding EMUs within the context model.

In order to monitor the energy performance, the contextualization platform can be used to calculate the EnPIs and also the performance improvement potentials using the surrogate models in real-time. All relevant information including the knowledge of the meshed network of energy flows will be available.

Finally, web-based dashboards can be generated automatically. Therefore, dashboard design templates for EMUs as well as for aggregated views of several composed EMUs can be provided. Based on open interfaces of the contextualization platform, the relevant EMUs can be explored and the actual and historic energy performance data can be presented online to the user.

Summarizing the described tool chain, the relevant components are:

- software to build/create intelligent P&IDs,
- export mechanism into a file compliant to DEXPI P&ID specification,
- DCS providing real-time data of process measurements,
- PIMS collecting the historian of tagged measurements,
- parser to extract instrumentation TAGs of DEXPI xml files for PIMS data transfer configuration,
- interface in a contextualization platform to interpret DEXPI xml files building up a representation of an energy flow-oriented plant topology model,
- mechanism to link PIMS data points with energy flow representations in the contextualization tool and to add additional master data needed for energy performance indicator calculations,
- user interface to define boundaries for material or energy balancing of subsets of the plant in the contextualization tool,
- application of an ML method to create (surrogate) models based on contextualized data to calculate KPIs – here especially EnPIs,
- engine for cyclic or event driven calculation of KPIs based on the context model and the corresponding surrogate models,
- mechanism for automatic creation of visualizations of the KPIs in dashboards.

3 Application Results

The approach above describes a tool chain to provide energy performance indicators for operator and plant manager with minimal configuration effort required by exploiting existing engineering data to build up the performance monitoring application from scratch. The general approach and developed tool chain were evaluated at a pilot plant of Evonik [19] in Marl, Germany. The pilot plant represents a distillation column including corresponding plant units upstream and downstream. The P&ID of the pilot plant and the corresponding exported DEXPI file was available by a CAE tool of X-Visual [20]. The KEEN data management platform Dataverse [21] and the metadata schema ProMetaS (process engineering/industry metadata schema), which includes among others plant information meta data, were developed in parallel to the tool chain in order to address the FAIR meta data aspects and data provisioning mentioned in the starting point of the tool chain. Due to the early beginning of the tool chain implementation, the management platform and the metadata schema ProMetaS were not already available but conceptually envisaged in the overall tool chain approach. ProMetaS is described in detail in a KEEN-related paper that is in being published in parallel to this article [22].

A parser and a data coupler developed by LeiKon was used to build up a PIMS historian based on a SQL database by fully automatically extracting measurement data from the DCS SIMANTIC PCS Neo of Siemens [23] using an OPC UA [24] interface. The same was evaluated independently from the pilot plant by using a commercial PIMS (here OSI PI of Aveva [25]) and the DeltaV DCS of Emerson [26]. Hence, the system independency of the automatic built data historian was proven. As contextualization platform the framework IntexcSuite of LeiKon [27] was used. It is an industrial approved solution to build up process-oriented context models usable for online monitoring and optimization solutions. The surrogate model builder was developed also by LeiKon as an Add-on of the IntexcSuite. To present the dashboards, the openly available observability platform Grafana [28] was used. Thus, the whole tool chain building up an energy performance dashboard starting from CAE up to a web-based visualization for operators and plant manager could be successfully demonstrated and the system independent approach could be evaluated.

4 Summary and Outlook

The approach and developed tool chain discussed in this paper provides a solution to deploy AI-based solutions with less engineering and configuration effort. Especially for generic AI applications cost-intensive engineering efforts can be decreased significantly. This was be demonstrated by a use case in the area of energy management. In general, the approach can be used also to build up other kind of context models based on already available engineering data. Thus, the approach provides an enabler technology to propagate AI-based solutions for real-time applications in process industry.

In future, further applications can be build up in a similar way. For these upcoming use cases we recommend using a data management platform with respective meta data schemes from the beginning on, earning the benefits of structured data handover and FAIR data in general. One example of a new application of a tool chain as describe

 © 2023 The Authors. Chemie Ingenieur Technik published by Wiley-VCH GmbH

15222640, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cite.202300004 by Cochrane Germany, Wiley Online Library on [11/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

here could be predictive maintenance solutions based on the context knowledge of the installation constellations of devices within a plant, including their life cycle data extractable by different IT and OT systems.

Acknowledgment

We are grateful for the financial support of BMWK (German Federal Ministry of Economic Affairs and Climate Action, Support code 01MK20014T) for their financial support. Open access funding enabled and organized by Projekt DEAL.

List of Symbols

$E(j)$	[kWh]	energy demand
$m(p)$	[kg]	amount of product during a specified time interval
j	[–]	type of the energy source

Abbreviations

AI	Artificial Intelligence
BDP	Best Demonstrated Practice
CAE	Computer Aided Engineering
DCS	Decentralized Control System
DEXPI	Data Exchange in the Process Industry
EMU	Energy Managed Unit
EnPI	Energy Performance Indicator
FAIR	Findable, Accessible, Interoperable, Reusable
JSON	JavaScript Object Notation
KPI	Key Performance Indicator
OPC UA	OPC Unified Architecture
OT	Operation Technology
P&ID	Piping and Instrumentation Diagram
PDCA	plan-do-check-act
PFD	Process Flow Diagram
PIMS	Process Information Management System
PLC	Programmable Logical Controller
RDL	Reference Data Library
XML	Extensible Markup Language

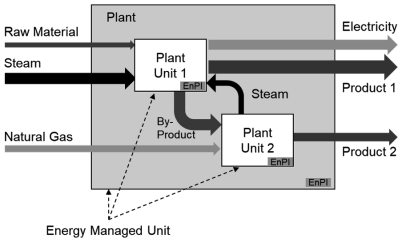
References

[1] ISO 50001, *Energy management*, International Organization for Standardization, Geneva **2021**. www.iso.org/iso-50001-energy-management.html (accessed on December 20, 2021)
[2] S. Krämer, S. Engell, *Resource efficiency of processing plants: monitoring and improvement*, Wiley-VCH, Weinheim **2018**.
[3] B. Beisheim, S. Krämer, S. Engell, Hierarchical aggregation of energy performance indicators in continuous production processes, *Appl. Energy* **2020**, *264*, 114709.
[4] B. Beisheim et al., Energy performance analysis of continuous processes using surrogate models, *Energy* **2019**, *183*, 776–787.
[5] NAMUR NE 162:2018, *Resource Efficiency Indicators for monitoring and improving resource efficiency in processing plants*, Beuth Verlag, Berlin **2018**.
[6] M. Kujanpää et al., *Successful Resource Efficiency Indicators for process industries*, VTT Technology 209, **2017**. https://publications.vtt.fi/pdf/technology/2017/T290.pdf (accessed on January 03, 2023)
[7] *ISO 50003:2021, Energy management systems Requirements for bodies providing audit and certification of energy management systems*, International Organization for Standardization, Geneva **2021**. www.iso.org/standard/77575.html (accessed on December 20, 2021).
[8] U. Enste et al., Gib den Daten einen Kontext: Echtzeit-Kontextmanagement ein Beitrag zur NOA, *atp magazin* **2020**, *62 (9)*, 44–52.
[9] M. Kalliski et al., Real-time resource efficiency indicators, *atp – Automatisierungstech. Prax.* **2016**, *58*, 64–71.
[10] M. Wiedau et al., Enpro data integration: extending DEXPI towards the asset lifecycle, *Chem. Ing. Tech.* **2019**, *91 (3)*, 240–255.
[11] D. B. Cameron et al., A semantic systems engineering framework for zero-defect engineering and operations in the continuous process industries, *Front. Manuf. Technol.* **2022**, *2*. DOI: https://doi.org/10.3389/fmtec.2022.945717
[12] M. Wiedau et al., Towards a Systematic Data Harmonization to Enable AI Application in the Process Industry, *Chem. Ing. Tech.* **2021**, *93 (12)*, 2105–2115. DOI: https://doi.org/10.1002/cite.202100203
[13] IEC 82045:2001, *Document management – Principles and methods*, International Organization for Standardization, Geneva **2001**. www.iso.org/standard/34159.html (accessed on February 27, 2023)
[14] H. Temmen et al., Intelligenter R&I-Austausch, *atp magazin* **2016**, *58*, 54–65.
[15] https://dexpi.org/certificates (accessed on February 27, 2023)
[16] *ISO 23081-2:2021, Information and documentation-Metadata for managing records-Conceptual and implementation issues*, International Organization for Standardization, Geneva **2021**. www.iso.org/standard/81600.html (accessed on January 05, 2023)
[17] ISO 23081-1:2017, *Information and documentation-Records Management Process-Metadata for records-Principles*, International Organization for Standardization, Geneva **2017**. www.iso.org/standard/73172.html (accessed on January 05, 2023)
[18] M. D. Wilkinson et al., The FAIR Guiding Principles for scientific data management and stewardship, *Sci. Data* **2016**, *3*, 1–9.
[19] www.evonik.com (accessed on January 03, 2023).
[20] www.x-visual.com/ (accessed on January 03, 2023)
[21] https://dataverse.org/ (accessed on December 11, 2022)
[22] L. Sherpa et al., *Chem. Ing. Tech.* **2023**, *95 (7)*, in press. DOI: https://doi.org/10.1002/cite.202200225
[23] https://new.siemens.com/de/de/produkte/automatisierung/prozessleittechnik/simatic-pcs-neo.html (accessed on January 03, 2023)
[24] https://opcfoundation.org/about/opc-technologies/opc-ua/ (accessed on January 03, 2023)
[25] www.aveva.com/en/products/aveva-pi-system/ (accessed on January 03, 2023)
[26] www.emerson.com/de- de/automation/deltav (accessed on January 03, 2023)
[27] www.leikon.de/ (accessed on January 03, 2023)
[28] https://grafana.com/ (accessed on January 03, 2023)

 © 2023 The Authors. Chemie Ingenieur Technik published by Wiley-VCH GmbH

15222640, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cite.202300004 by Cochrane Germany, Wiley Online Library on [11/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

DOI: 10.1002/cite.202300004



Tool Chain to Extract and Contextualize Process Data for AI Applications

Lincoln Sherpa, Ralph Müller-Pfefferkorn, Udo Enste, Gregor Tolksdorf, Michael Kawohl, Michael Wiedau*

Research Article: For the application of AI in the process industry, contextualized data is of added value. By building-up a tool chain from source systems providing structural and raw data up to dashboards visualizing key performance indicators of production sites, many steps of the contextualization and AI application for relevant use-cases can be automatized.

15222640, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cite.202300004 by Cochrane Germany, Wiley Online Library on [11/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

 © 2023 The Authors. Chemie Ingenieur Technik published by Wiley-VCH GmbH