



A Digital Ecosystem for Improving Product Design

Sylvain Lefebvre¹ (✉) , Michaël Lecointre³, Benoit Lardeux¹, Jean-Marie Guyader², Olivier Aubrun³, Birane Touré³, and Maher Jridi¹

¹ L@bIsen, Isen Nantes, 33 Quater Avenue du Champs de Manoeuvre, 44470 Carquefou, France

{sylvain.lefebvre,benoit.lardeux,maher.jridi}@isen-ouest.yncrea.fr

² L@bIsen, ISEN Brest, 20 Rue Cuirassé Bretagne, 29200 Brest, France

jean-marie.guyader@isen-ouest.yncrea.fr

³ ACOME, 50140 Romagny-Fontenay, France

{michael.lecointre,olivier.aubrun,birane.toure}@acome.fr

Abstract. Digitization is reaching to every corner of the industry. The industry 4.0 (I4.0) movement initiated a move towards a stronger reliance on data in the manufacturing domain in order to improve processes and product quality. Multiple works highlight the potential benefits of deploying artificial intelligence or big data management platforms for industrial companies to improve their processes and provide a better understanding of their production tools. Many I4.0 work often assume the existence of interconnected machinery, sensors, and Manufacturing Execution System (MES) in the company and assume that most data is already available from these interconnected systems on the production line. Unfortunately, this does not reflect the state of many companies whose production systems are not interconnected due to historical reasons or security and normative issues. This report describes a big data architecture for the collection, storage and analysis of industrial prototype data. We provide details on how such an architecture can be structured and how it supports the engineering cycle in a partner company through a case study.

Keywords: Big Data · Industry 4.0 · Analytics

1 Introduction

Digitization is reaching to every corner of the industry. The industry 4.0 (I4.0) [8] movement initiated a move towards a stronger reliance on data in the manufacturing domain in order to improve processes and product quality. Multiple works highlight the potential benefits of deploying artificial intelligence [3,20] or big data management platforms [12,22] for industrial companies to improve their processes and provide a better understanding of their production tools. The extensive use of data collection and data management tools in the I4.0 paradigm allows companies to implement data-based approaches to manage their production. A data-based approach, per opposition to a model-based approach, is

a methodology where decision making is based on observations on the studied phenomenon rather than on theoretical models. For example, the product engineering cycle can be significantly improved by the usage of advanced data analytics [19].

However, several obstacles hinder the adoption of those approaches in smaller companies across Europe. For example, several studies [6, 15] highlight the difficulties to collect data, as well as a lack of clear return on investment from the deployment of data-based approaches.

Indeed, many I4.0 works assume the existence of a high connectivity of the various components of the shop floor in the company and assume that most data is already available from these interconnected systems on the production line. Unfortunately, this does not reflect the state of many companies whose production systems are not interconnected due to historical reasons or security and normative issues. We argue in this paper that the later type of companies can benefit from data-based I4.0 processes through the use of historical data already present in their IT systems. These historical data sources include, for example, design documents, spreadsheets, normative documents, or specification documents. For example, spreadsheets [4] are often stored in a shared file system over the organization. It is an easily accessible source of information. These represent a significant amount of data available in a large variety of low-structured format which poses a data integration challenge. At the same time, collecting these data has a lower impact on IT infrastructure and production lines, compared to the deployment of new sensors and data collection systems on the manufacturing shop floor. Therefore, the collection of such historical data may be useful to bootstrap data-based processes at a lower cost.

In this report, we describe through a case study in the cable making industry how a big data infrastructure based on I4.0 can be deployed to improve the engineering process. To achieve this goal, multiple obstacles must be overcome:

1. Collect historical data in multiple heterogeneous semi structured format
2. Gather collected data in a central Datalake with an appropriate data model
3. Provide useful data visualization for production engineers.

This work lies at the intersection of several approaches to enhance manufacturing companies processes by using data (See Sect. 2). Our approach was deployed in a case study described in Sect. 3. The main operating principle of the data platform relies on the design of a virtuous cycle of automated importation and user-based update of data sources (See Sect. 4). Finally, we discuss the impact of the platform in Sect. 5 and provide conclusions and future research areas are highlighted in Sect. 6.

2 Background and Related Works

2.1 Big Data Analytics

The usage of big data analytics approaches [2, 11, 12, 22] enable manufacturing organizations to overcome the 5V challenges (Volume, Velocity, Variety, Veracity,

Value) posed by large amounts of data. For example Sun et al. in [19] designed a platform combining Product Life-cycle Management (PLM) software and data analytics processes to generate optimized planning and task assignment to engineers in the Engineer To Order industry. They rely on the Cross Industry Standard for Data Mining (CRISP-DM) [21] to extract relevant information from collected data.

Datalakes are frequently applied in industry 4.0 related platforms. Commercial offerings from major cloud computing providers such as Amazon¹ or Microsoft² include a Datalake for storage and data analytics. In [11] Kebisek et al. describe a platform for quality level prediction of paint on a production chain. The proposed system aggregates data from different production batches in a data lake and allows for historical data analysis. The work focuses on a single production phase and not on the product design stage. The data only include structured data from shop floor equipment.

Multiple approaches coexist to integrate heterogeneous data sources in such platforms. In [2], Bonnard et al. designed a proxy Application Programming Interface to ensure a common data format for ingestion from various shop-floor level sources and provide a generic industry 4.0 Big Data platform. The platform provides services for data analytics and a set of standard dashboard for helping shop floor workers to assess the state of the production system. The BiDRAC model [16] developed by Sanz et al. provides several use cases related to fault detection and analysis in car paint coating process. Their platform integrates both unstructured data sources and structured data coming from industrial equipment. However, the data extraction process from the unstructured sources is not detailed in the paper. In [9] Kahveci et al. detail the building of a big data infrastructure to collect data in a manufacturing plant. As their approach focuses on real-time process monitoring, they require the deployment of multiple additional equipment in the factory. The proposed architecture relies on standard layers for data integration visualization and dashboards. The collected data is structured in nature as it comes from various industrial equipments. They do not collect or integrate this data with document-based sources.

2.2 Semantic Approaches

Semantic approaches use the tools provided by knowledge modeling and ontologies to facilitate data integration and retrieval. In [14], Patel et al. propose a semantic web-based platform for industrial data sources integration and data based application development. More recently, Semantic Graphs [18] emerged as another approach to solve data integration challenges posed by the multiplicity of data sources in industrial companies. Prominent manufacturers such as Bosch [10], or Siemens [5] implement multiple data integration and analytics

¹ <https://aws.amazon.com/fr/solutions/case-studies/innovators/volkswagen-group/>, last seen December 5, 2023.

² <https://www.microsoft.com/fr-fr/industry/manufacturing/microsoft-cloud-for-manufacturing>, last seen December 5, 2023.

on top of Knowledge graph based platforms. Graph-based approaches provide a flexible schema which facilitates data integration across multiple sources. However, these systems often rely on specific databases with less standard querying languages, which hinders the empowerment of users to extract their data for analysis. Moreover these reports tend to focus on semi-structured input data (JSON, XML), coming from shop floor machines and sensors. Analytics performed on these data focus on the detection of quality issues during fabrication, and not to improve the product engineering phase.

2.3 Machine Learning in Manufacturing Industry

Data-based approaches to decision making often involve the use of predictive models. Therefore the use of machine learning ecosystems is also rising in the manufacturing industry. These platforms share some common infrastructure with big data platforms. In their reviews of machine learning usage in industry 4.0 [20], Tercan et al. demonstrate the usefulness of predictive approaches for quality prediction in manufacturing. However, they also denote the lack of systematic integration and deployment of these technologies in real production settings. They also note that a large fraction of the complexity of such projects lies in the data acquisition process as well as the connectivity of shop-floor machines. They also do not consider the impact of machine learning techniques on the product design cycle. In industrial process monitoring the development of Soft Sensors extensively use big data and machine learning approaches. Soft Sensors are software components which allow to monitor difficult to estimate process outputs or metrics through predictive models. For example, in [7], Kabugo et al. use a big data analytics and machine learning platform based on cloud technologies to model and develop soft sensors for the monitoring of gas emissions in a Waste To Energy Plant. Their approach collects data from various sensors on the machines and evaluate different machine learning models for the studied phenomenon. In this approach, data used for modeling comes from various industrial sensors which are costly to interconnect to data processing facilities in the cloud. They do not integrate unstructured data and documents in their process.

3 Case Study: Product Prototyping

Product design and prototyping phase is an essential part of the industrial production cycle. This phase happens before the mass production of the product. In our approach we model this phase as an iterative cycle with four steps:

1. *The design step*: Engineers receive desired specifications from the R&D department or from clients. They proceed, based on their experience of the industry or some accumulated data on past designs and simulation tools to propose a product design that should follow the specifications,

2. *The build step*: Based on the target design from the previous step, the study engineers send a prototype fabrication order (FO) to the manufacturing floor and provide machine parameters for each of the building or processing phases of the target product,
3. *Test step*: Once the prototype product is built, qualification engineers proceed to some tests on the built product to check the conformity to the specification,
4. *Analysis step*: If the prototype product does not conform to the tests, study engineers gather data from the production phase and test phase and proceed to various analysis in order to improve the design of the product.

Engineers iterate over this process to come up with a design that can full-fill all target specifications.

Iterations which do not come up with the proper design result in wasted raw materials and machine time that could otherwise be used for more productive fabrication. It is therefore of great importance to be able to record all attempts at fabrication to avoid repeating previous mistakes and anticipate possible interactions of various design decisions.

This design data analysis capability is of great importance for many manufacturing organizations. However, as shown by [12] this data is often gathered in multiple silos across the organization and it remains difficult to provide engineers with a common entry point for their access and analysis.

4 Solution Architecture

Therefore the building of this data platform focused on the ingestion of various documents and required the development of three layers as presented in Fig. 1.

The first layer is the *data ingestion and extraction layer*. This layer, described further in Sect. 4.1 processes raw input files and data sources to map them on the schema of the Datalake. The *Data Integration Layer* maps and consolidates the various entities resulting from the data extraction layer into a common Datalake schema. Finally, the *Access Layer* enables product designers to retrieve data as well as discover and model relationships between different collected data sources.

4.1 Data Extraction

The Ingestion Layer in Fig. 1 is composed of three sub-components to achieve data collection as well as extraction, transformation and loading (ETL). The first part is the Data Exchange Zone (DEZ). This DEZ is a temporary storage area where the source data coming from the manufacturing company IT system or Shop Floor data sources can be stored before processing. This space provides a set of standardized APIs for data transfers. The data exchange zone stores the data temporarily until its ingestion in the Datalake is confirmed. Once the ingestion is successful the data in that staging area can be deleted.

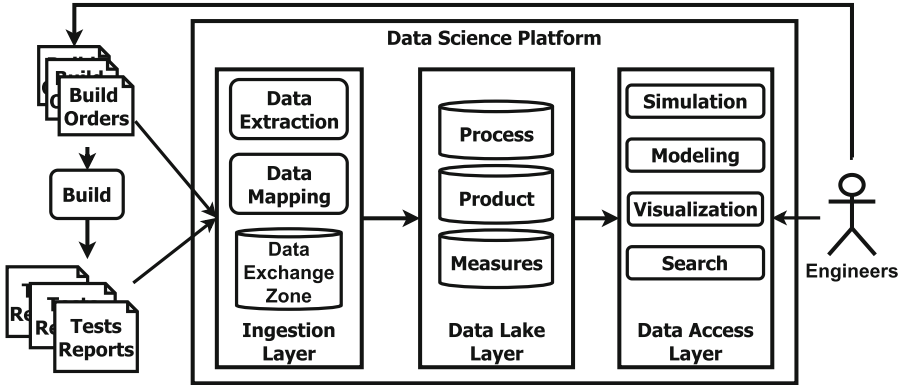


Fig. 1. The collected data are stored temporarily in the Data Exchange Zone, which provides several APIs for data transfers. The Data extraction and mapping processes transform the data to conform to the Datalake schema. The access layer provide Engineers and Analysts with tools to Search Visualize and Model the data in the Datalake

A set of data input processes update the data exchange zone continuously. In the case of file based ingestion of semi-structured documents which constitutes the majority of ingestion for the design cycle use-case described in Sect. 3, data is refreshed every 24h, by taking all modified files in that interval. This delay was chosen as acceptable for the users. Incoming data in the platform are from heterogeneous sources that can be separated in two main classes, based on their structuration level:

Database Extractions: These data sources provide a well-structured data schema that can be easily mapped to the target Datalake schema.

Document Files A large quantity of data is stored in semi-structured tables inside work spreadsheets, PDF, or Word documents. These data sources may have no structure at all and require a specific extraction process.

While the first kind of data is well structured, data coming from document sources poses a heterogeneity problem. Therefore, the data extraction process has to account for several styles of data sources. The data extraction module supports this role.

Located in the extraction layer (Fig. 1) the data extraction module is a framework that allows the user to extract data from unstructured documents and transform it to various table model database, based on a set of customizable rules. Inspired by the works of Shigarov et al. [17] this module processes each document with the sequence of steps described in Fig. 2. After the document identification step, two types of rules are applied:

Segmentation Rules The first set of rules, called *segmentation* rules, allows users to transform each document’s pages into multiple sections to create a tree-shaped object, which recursively subdivides the document into small

sections of meaningful content. The developer defines what is considered as a meaningful content section. Each segmentation rule is composed of an identifier, a capture condition, a capture range, and a section type. A capture condition is a set of predicates that switches the file content capture on or off based on the current content of the file. Such predicates can be, for example: matching the content of a cell or a line with a regular expression, matching a line number or matching a column number. The capture range specifies the vertical range of capture, for example: full line or the number of columns in the case of spreadsheets. Each rule associates the captured data with the specified section type, which reflects the nature and basic structure of the captured data (tabular data, text data, etc.).

Mapping Rules The second set of rules are the *mapping* rules. These rules associate each section or set of sections to a schema mapping function. The mapping function takes a set of sections as well as a target schema and returns a list of objects conforming to the target schema. The developer can define a schema for each mapping rule. This approach provided enough flexibility to extract the data in our use case.

Following the data extraction quality is essential for the good usage of the platform, especially since the input formats can evolve over time. Therefore, the platform administrators maintain a data extraction quality dashboard. The results of extraction for each document can be tracked and errors are recorded in order to establish a quick diagnostic when data is not available in the system. Interaction with the final user, as well as the repeating ingestion process, provide means to efficiently correct issues by re-typing documents with inaccurate formats and ingest them again post modification. The users monitor two main metrics:

Importation ratio Ratio of successfully recognized documents over the total number of documents in the Datalake. This ratio provides a good estimate of the amount of correct documents in the data set,

Extraction ratio For each table in the model, the ratio of documents that have an entry in it. This provides an overview of the completeness of the extraction process.

Extraction reports are created to enable users to follow and improve data input. With these reports, users can check whether the documents they expect to find were loaded in the Datalake or not. If the data extraction failed, the user can identify the root cause of importation failure. Thanks to the continuous importation process, users can update failed documents to correct it so that it can be re-imported at the next scheduled data collection.

4.2 Data Lake

The Datalake is a big data storage facility, that stores data post extraction. Data are stored in progressively more structured formats, from the raw extraction of the various data sources to the target data model for analysis. The first category

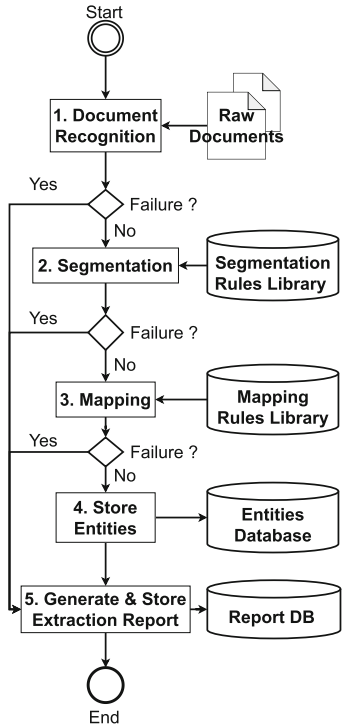


Fig. 2. Rule-based extraction process overview. Each document goes through two phases: a segmentation phase and a mapping phase, then extracted objects are stored in the target database and the system creates and stores an extraction report

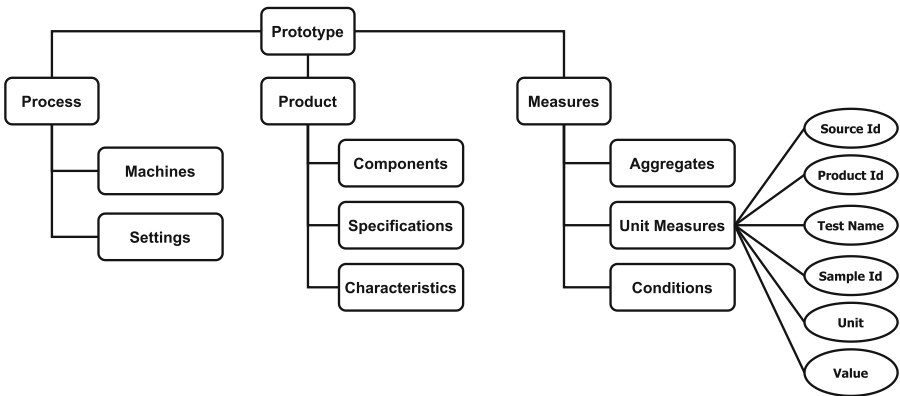


Fig. 3. Data Model: Collection of tables describing Process, Product and Measures data.

of tables consists in sets of entities extracted from input sources. These entities are then mapped and aggregated into a common data model that is more suitable for analysis and retrieval by the users. This data model, visible in Fig. 3 is based on the high-level taxonomy of data for machine learning applications in industrial context describe by Tercan and Meisein in [20]. Data is stored in multiple collections of tables, one per element of the taxonomy:

Measurements: This collection of table is a set of fact tables including all measurements done on all finite or semi-finite products in the chain. These data can come from quality monitoring systems or qualification tests results. Data sources include MES systems, manually filled spreadsheets or manufacturing quality monitoring systems.

Process: This table collection is a set of dimension tables related to the configuration settings used for a given machine and Fabrication Order (FO). Sources can be structured such as Supervisory Control And Data Acquisition (SCADA) or Manufacturing Execution Systems (MES) data or semi-structured data such as spreadsheets or images.

Products: This collection is another set of dimension tables containing the requirements and technical specifications of the products. Sources include Product Lifecycle Management systems, Product Design systems and fabrication order documents.

Each of these collections of tables include tables with similar schema as highlighted for the Unit Measures collection in Fig. 3. All tables in the Unit Measure collection have a similar set of fields, prefixed with the name of the value. This results in a star schema centered on the measurements tables. Advantages of such schema is a more efficient query execution as well as understandability of the tables by the users. On the other hand, the quantity of data is bigger because of the schema denormalization.

4.3 Access Layer

The data access layer is designed to provide production engineers the means to search, visualize, analyze, and model their data.

Data Retrieval. To enable engineers to search and extract data, we chose the open-source component Metabase³. It was selected because of its user interface that enables non expert users to build queries efficiently in a What You See is What You Get fashion. Users can create, record, and share dashboards and queries. The system is Open Source and can be deployed easily. It also allows for data virtualization, which permits modifications to underlying tables and data models to be transparent for the final users. Its administration interface also allows to document the created views extensively.

³ <https://www.metabase.com/>, seen December 5, 2023.

Visualization. The visualization tool is a specific web interface developed in Python to enable exploration and discovery of relations between variables in a given dataset.

The tool retrieves existing queries from the data catalog through the Metabase API, and provides users with the opportunity to build an understanding of the data through automatic data selection and visualization. This provides Data Understanding and Data Preparation tasks based on the CRISP-DM process [21]. As a first step, users can retrieve data from queries stored in the Datalake. Then, through the tool, users provide additional details such as the columns that should be considered as variables, as well as the nature of each of those variables.

The first visualization the tool provides is a correlogram. Shown in Fig. 4, the correlogram is an interactive heat-map where variables are listed from top to bottom while response variables are listed from left to right. The color level indicates the correlation direction as well as the intensity of the correlation, which is proportional to the R-coefficient of the computed correlation. We chose to use the Spearman Rank Correlation [1] coefficient to cover a larger number of relationships while keeping the results easy to interpret for users. Some cells have no color and correspond to variable pairs where the correlation coefficient is not statistically significant. The statistical significance of the correlation is verified by checking the p-value as well as confidence intervals on the correlation coefficient. The user can select a cell of the heat-map to visualize the corresponding scatter plot. This tool enables users to rapidly validate their intuition against data with an easy to interpret visualization.

5 Discussion

In this section we discuss the impacts and outcomes of usage of the data platform. The first impact is a significant gain of time for reporting and information gathering about product prototypes. Thanks to the single point of entry for accessing data provided by the platform, some users were able to find relevant data for reporting on current designs in a few minutes or hours. The same process used to take several days before the deployment of the platform. This gain of time for information gathering affects the iterative engineering cycle described in Sect. 3 by accelerating the feedback loop between prototype fabrication and data analysis phases.

The second impact is the quality improvement of data after deployment of the platform. Users observed that the amount of data collected and available in the platform depends on how input files and documents are filled. As developing rules for the many possibilities of filling a line or cell in a documents is infeasible, and the time and resources required for building large annotated databases is not always available, platform users started an internal process for more strict document filling and checking rules. We hope these efforts lead to better data importation success rate in the future.

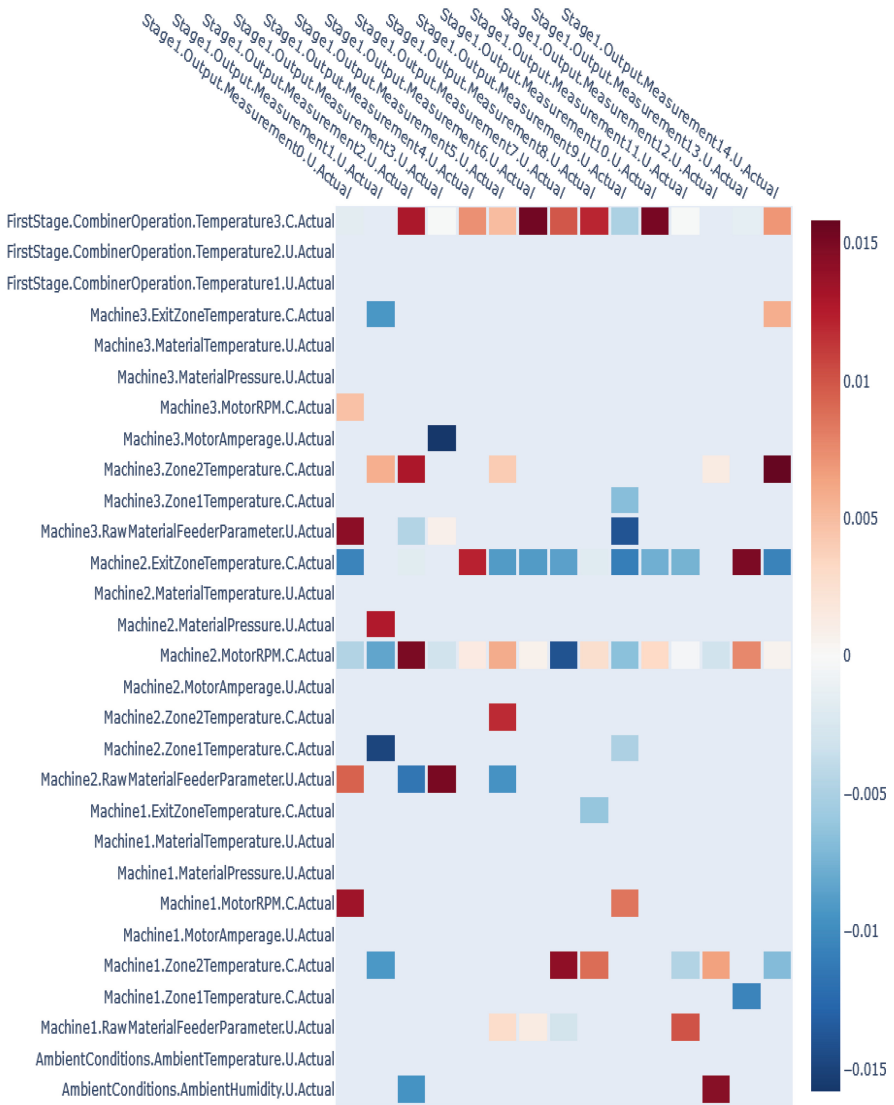


Fig. 4. Interactive Correlation heat-map example, data is from [13]

Another concern was the ease of access to data, which raises concerns against data leakage risks. As the data contains and correlate design data as well as measurements, the risk is the leakage of company knowledge, know-how and intellectual property. Therefore for the acceptability of the approach, it is important to enforce best practices such as end-to-end encryption, provide strict access control, and demonstrate these aspects to stake-holders.

6 Conclusion and Future Works

In this work, we addressed several challenges related to Industry 4.0 implementation for the improvement of product design phase. First, by deploying a continuous data collection system based on historical data and sources already available in the company IT systems, we are able to demonstrate the feasibility of a data-based approach for modeling product features with a lower material cost. We also developed a document extraction framework for tabular data extraction to address the diversity of data sources in the various documents stores of the company. The platform use resulted in a significant time gain according to users, providing the ability to extract analyze product tests results in a matter of hours instead of days previously. This may also result in qualitative improvement of the process by initiating better data input and control. Through continuous data collection and quality feedback, users can improve data quality daily.

In future works, we plan to develop the number of supported data sources, and use Natural Language Processing techniques in order to improve the robustness of the data extraction process and collect more various sources samples. In another part of the work, developing predictive models based on extracted data will enable users to anticipate the impact of design changes on future product. More specifically the study of graph-based techniques for these approaches seems promising to address the variety of data in input sources.

Acknowledgements. Funding for this work was provided by ACOME (<https://www.acome.com/en>) as part of the *Data Architecture for ACOME Factory 4.0* chair established between ACOME and the L@bISEN research laboratory (<https://isen-brest.fr/labisen/>).

References

1. Dodge, Y.: Spearman rank correlation coefficient. In: Dodge, Y. (ed.) *The Concise Encyclopedia of Statistics*, pp. 502–505. Springer, New York (2008). https://doi.org/10.1007/978-0-387-32833-1_379
2. Bonnard, R., Arantes, M.D.S., Lorbieski, R., Vieira, K.M.M., Nunes, M.C.: Big data/analytics platform for industry 4.0 implementation in advanced manufacturing context. *Int. J. Adv. Manuf. Technol.* **117**(5–6), 1959–1973 (2021)
3. Dogan, A., Birant, D.: Machine learning and data mining in manufacturing. *Expert Syst. Appl.* **166**, 114060 (2021). <https://doi.org/10.1016/j.eswa.2020.114060>
4. Fricke, A., Schöneberger, J.: Industrie 4.0 with MS-excel? *Chem. Eng. Trans.* **43**, 1303–1308 (2015). <https://doi.org/10.3303/CET1543218>
5. Hubauer, T., Lamparter, S., Haase, P., Herzig, D.M.: Use cases of the industrial knowledge graph at siemens (2018)
6. Janík, S., Szabó, P., Míkva, M., Mareček-Kolibický, M.: Effective data utilization in the context of industry 4.0 technology integration. *Appl. Sci.* **12**(20), 10517 (2022). <https://doi.org/10.3390/app122010517>
7. Kabugo, J.C., Jämsä-Jounela, S.L., Schiemann, R., Binder, C.: Industry 4.0 based process data analytics platform: a waste-to-energy plant case study. *Int. J. Electr. Power Energy Syst.* **115**, 105508 (2020). <https://doi.org/10.1016/j.ijepes.2019.105508>

8. Kagermann, H., Wahlster, W., Helbig, J.: Recommendations for implementing the strategic initiative industrie 4.0 - securing the future of German manufacturing industry. Final report of the industrie 4.0 working group, acatech - National Academy of Science and Engineering, Munchen (2013). http://forschungsunion.de/pdf/industrie_4.0_final_report.pdf
9. Kahveci, S., Alkan, B., Ahmad, M.H., Ahmad, B., Harrison, R.: An end-to-end big data analytics platform for IoT-enabled smart factories: a case study of battery module assembly system for electric vehicles. *J. Manuf. Syst.* **63**, 214–223 (2022). <https://doi.org/10.1016/j.jmsy.2022.03.010>
10. Kalaycı, E.G., et al.: Semantic integration of Bosch manufacturing data using virtual knowledge graphs. In: Pan, J.Z., et al. (eds.) *ISWC 2020. LNCS*, vol. 12507, pp. 464–481. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-62466-8_29
11. Kebisek, M., Tanuska, P., Spendla, L., Kotianova, J., Strelec, P.: Artificial intelligence platform proposal for paint structure quality prediction within the industrie 4.0 concept. *IFAC-PapersOnLine* **53**(2), 11168–11174 (2020). <https://doi.org/10.1016/j.ifacol.2020.12.299>. 21st IFAC World Congress
12. Lavasani, M.S., Ardali, N.R., Sotudeh-Gharebagh, R., Zarghami, R., Abonyi, J., Mostoufi, N.: Big data analytics opportunities for applications in process engineering. *Rev. Chem. Eng.* (2021). <https://doi.org/10.1515/revce-2020-0054>
13. Liveline Technologies: Multi-stage continuous-flow manufacturing process (2020). <https://www.kaggle.com/supergus/multistage-continuousflow-manufacturing-process/metadata>. Accessed 01 Oct 2022
14. Patel, P., Ali, M.I., Sheth, A.: From raw data to smart manufacturing: AI and semantic web of things for industry 4.0. *IEEE Intell. Syst.* **33**, 79–86 (2018). <https://doi.org/10.1109/MIS.2018.043741325>
15. Rädler, S., Rigger, E.: A survey on the challenges hindering the application of data science, digital twins and design automation in engineering practice. *Proc. Des. Soc.* **2**, 1699–1708 (2022). <https://doi.org/10.1017/pds.2022.172>
16. Sanz, E., Blesa, J., Puig, V.: Bidrac industry 4.0 framework: application to an automotive paint shop process. *Control Eng. Pract.* **109**, 104757 (2021). <https://doi.org/10.1016/j.conengprac.2021.104757>
17. Shigarov, A., Khristyuk, V., Mikhailov, A.: Tabbyxl: software platform for rule-based spreadsheet data extraction and transformation. *SoftwareX* **10**, 100270 (2019). <https://doi.org/10.1016/j.softx.2019.100270>
18. Song, G., Fu, D., Zhang, D.: From knowledge graph development to serving industrial knowledge automation: a review. In: 2022 41st Chinese Control Conference (CCC), pp. 4219–4226. IEEE (2022)
19. Sun, K., Li, Y., Roy, U.: A PLM-based data analytics approach for improving product development lead time in an engineer-to-order manufacturing firm. *Math. Model. Eng. Probl.* **4**, 69–74 (2017). <https://doi.org/10.18280/mmep.040201>
20. Tercan, H., Meisen, T.: Machine learning and deep learning based predictive quality in manufacturing: a systematic review. *J. Intell. Manuf.* **33**(7), 1879–1905 (2022). <https://doi.org/10.1007/s10845-022-01963-8>
21. Wirth, R., Hipp, J.: CRISP-DM: towards a standard process model for data mining (2000)
22. Xu, X., Hua, Q.: Industrial big data analysis in smart factory: current status and research strategies. *IEEE Access* **5**, 17543–17551 (2017). <https://doi.org/10.1109/ACCESS.2017.2741105>