https://doi.org/10.24867/GRID-2024-p46 Professional paper

TRANSFORMING OFFSET PRINTING WITH DIGITAL TWINS AND AI: INSIGHTS FROM RESEARCH INITIATIVES

Chris Trochoutsos [1,3,4], Alexandros S. Kalafatelis [2], Panagiotis Trakadas [2]
[1] Pressious Arvanitidis, printing company, Chalandri, Greece
[2] Department of Research & Development, Four Dot Infinity, Athens, Greece
[3] HELGRAMED - The Hellenic Union of Graphic Arts & Media Technology Engineers, Athens, Greece
[4] Hellenic Open University, School of Applied Arts, Patras, Greece

Abstract: *The manufacturing industry has a long tradition in the European Union, being one of the key employers providing jobs for highly skilled and qualified staff. Despite this, the sector is currently facing challenges related to global competition and rising costs, leading to declining profit margins. Additionally, companies are required to meet new environmental initiatives, demanding an ambitious voluntary commitment effort from their side, which has been proven to be difficult to implement. At the same time, manufacturing companies generate vast amounts of unstructured data daily yet struggle to leverage this data for supply chain optimization. This problem is especially noticeable in situations when there are many semi-structured data sources, like the ones found in offset printing operations (images, sensor data, etc.), typically leading to the creation of data silos storage and inefficiencies. Thus, a systematic approach is needed to optimize the process of receiving, combining, and analyzing data, which can, in turn, promote the creation of innovative business models based on manufacturing data. Similarly, the usage of a digital twin in production offers several advantages, primarily enhancing productivity, quality and cost efficiency. This paper presents two use cases illustrating how Digital Twin technologies and the CyclOps platform can be applied to offset printing operations, digitalizing factory processes and enabling AI-driven models, towards achieving Zero-Defect Manufacturing (ZDM), fostering improved efficiency and innovation in the manufacturing process.*

Key words: offset printing, digital twin, artificial intelligence, CyclOps

1. INTRODUCTION

The printing business is among the major manufacturing sectors globally, producing items such as packaging, flyers, books, magazines, and newspapers. The industry is essential to the European economy, producing an annual turnover of over €88 billion in EU GDP and employing over 770,000 individuals. Nonetheless, the industry is also confronting environmental and economic issues that might adversely affect demand and, consequently, the sector's economy. The production of high-quality printing items is intricate and significantly influences the environment because to the substantial consumption of raw resources (such as water, paper, ink, and aluminum) and chemicals that contribute to environmental deterioration (Kalafatelis et al., 2023b).

Contemporary manufacturing environments utilize IoT devices to comprehensively monitor the production chain, generating substantial volumes of data. These data may be employed to enhance the efficiency of printing operations using Artificial Intelligence (AI) and Machine Learning (ML) approaches, hence advancing the industry. Moreover, the use of AI/ML methodologies in manufacturing settings has demonstrated a significant capacity to optimize processes related to automation, waste reduction/prediction, and product quality, advancing towards Zero-Defect Manufacturing (ZDM).

This article presents two separate use cases in Offset Printing designed to improve production processes, utilizing a Digital Twin and the CyclOps platform. This document is organized as follows. Section 2 delineates the use case pertaining to the application of digital twin technology in printing processes. In Section 3, we describe the CyclOps platform and demonstrate its application on the Offset Printing domain. Finally, Section 4 encapsulates and deliberates the use cases.

2. DIGITAL TWIN FOR OFFSET PRINTING

Digital Twin solutions in the manufacturing sector represent a transformative approach to optimizing production, enhancing efficiency, and driving innovation. A digital twin is a virtual replica of physical assets, systems, or processes, enabling real-time monitoring, simulation, and analysis. By leveraging data from sensors and IoT devices, manufacturers can create detailed models that mirror real-world operations. This allows for predictive maintenance, improved product design, and streamlined supply chain management. Digital twins help manufacturers anticipate issues before they occur, reduce downtime, and increase productivity, ultimately leading to more agile and cost-effective operations in an increasingly competitive market.

2.1 Application of Digital Twin in Offset Printing Manufacturing Use Case

Living in a digital era, a modern and competitive printing company understands that digitalization is key for moving towards the new age of business and commerce. The digitalization process is of paramount importance for an organization to improve its production line. One of the most successful paradigms of digitalization, to be adopted by the printing company, is the Digital Twin. The Digital Twin scheme offers the possibility of creating a digital copy of the physical resources and production line. The process comes with several advantages like being able to simulate facilities in a protected environment, with minimal risks and costs.

In this use case, the company's target is to transform its printing process, which is the main pillar of operations. The printing process lacks real-time decision-making, prevents standardization, and has a significant environmental impact. Monitoring, process optimization, and proactive maintenance will be possible thanks to the transformation of the forenamed process through the Digital Twin experiment. More specifically, the Digital Twin is intended to assist human operators in the difficult machine (line) selection process, which is to send orders to the best available printing line.

Selecting the most suitable machine line to print an order (also referred to as a job) is one of the most important decisions in the production process. In the use case, there are three (3) different lines of printing: one (1) 5-colour printing line, one (1) 8-colour printing line, and one (1) 4-colour printing line. One of the most important decisions of the production process is to select the most suitable machine line for each unique print job (order).

To better understand the process in the use case, the figure above presents the machine line selection activity in a more detailed way. First, the order arrives. Then the production schedule is checked to see which jobs are in priority and if any jobs are currently being printed. The suitable machine line is selected based on the order features that we will present in the following subsection. Finally, the order features and selected machine lines are stored in the company's database, to be used by other departments for other tasks. Figure 1 presents an abstractive illustration of the machine selection process.

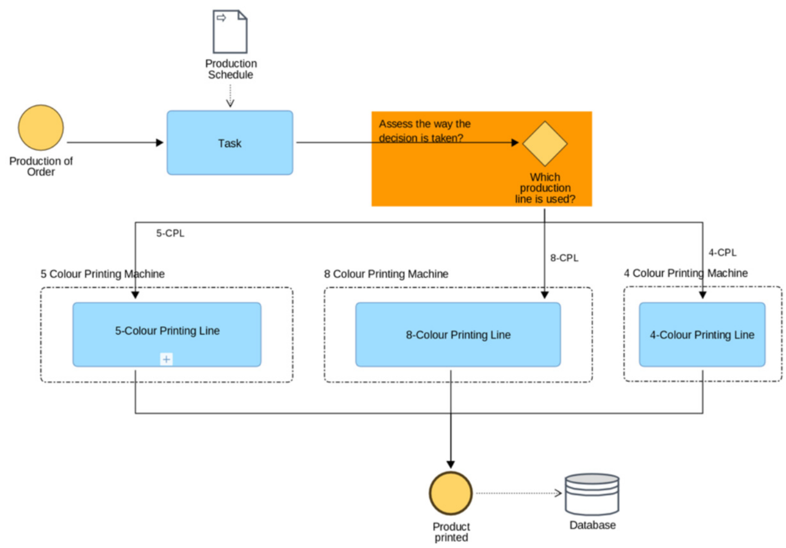

Figure 1: The machine selection process

The following parts have been taking into consideration into design specifications:

1. mapping the digital process landscape, showing all the processes of the physical unit,
2. user stories,
3. design of the experiment,
4. the specific task of machine allocation which will be optimized.

The beginning is the modelling of the Digital Process Landscape. This task is of paramount importance, as it discloses all processes that can be digitized. It should be noted that the Digital Twin experiment is essentially a digitalization step, created to optimize one or more tasks.
Modelling the Digital Process Landscape comes with several steps and advantages:

1. Instantiation and adaptation of process landscape for Customer needs. Each customer may come with different needs, which the enterprise should listen to and follow strictly. As the physical process is adapted to customer needs, so is the digital process.
2. Specification of Digitalization Goals. The enterprise should identify the goals of the digitalization process. In the current experiment, several measurable KPIs have already been specified.
3. Identification of Digitized Processes. In an industrial unit, not all processes can be digitized and further optimized. Furthermore, some processes are much more important for the enterprise, and when optimized, a significantly larger gain is obtained. Finally, there exist processes that are easier to digitize and optimize. All these factors are taken into consideration.
4. Definition of Data and Information Flow of Digitized Processes. The generated data to be used plays a critical role in the optimization task. All the necessary inputs and outputs of all processes in terms of data are required to be defined.
5. Specification of selected Digitalization Processes. Several of the processes will be selected as the digitalization processes to be optimized.

The following Figure 2 presents the complete Digital Process Landscape. This Landscape reveals the most important processes that play a significant role in the printing production line.

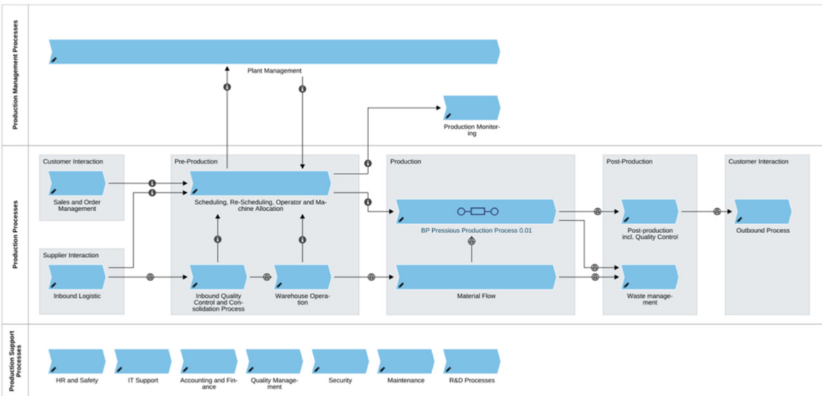

Figure 2: The Digital Process Landscape.

The operators would like to simulate a part of the production process, more specifically the machine line selection.
The economics department would like to know the cost of printing each unique order.
Given a specific order, the operators would like to know which is the optimal printing machine line.
The department that handles scheduling would like to know if an order can be printed on an alternative machine line.
The following Figure 3 is an abstractive high-level illustration of the Digital Twin architecture. On the left side, an abstract illustration of the physical production flow is shown. A customer arrives, giving a specific order. Then the order is passed to the Sales department. Given an order, the Sales department proceeds with the order analysis and outputs the paper type, quantity, size, and format. The total cost is computed by the Finance department and then the order is given to the Production department. Via the sensors, the extraction of important features like accuracy ratio, ink consumption, and colour accuracy is possible. On the right side, the Digital Twin is presented. Older and simulated orders are given for cost computation. Then the ML models can be trained to learn optimal configuration for machine selection. Newly arrived

orders can be given to the trained ML models and via the Digital Twin we have access to the optimal configuration (best machine) for the specific order.

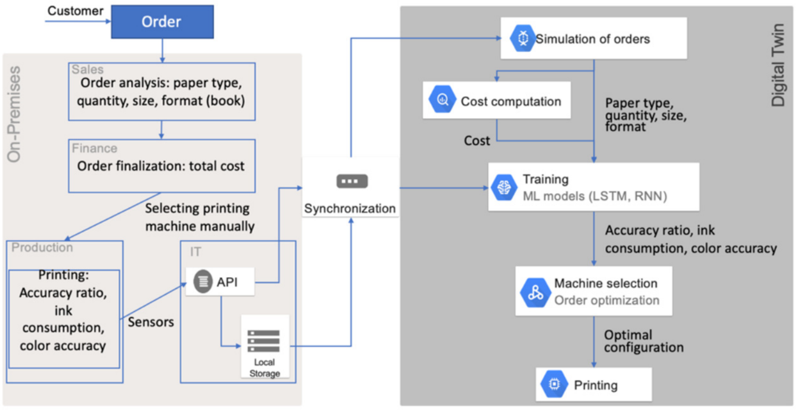

Figure 3: The Digital Twin design

The implementation of business process modelling (BPM) is done with ADONIS, a full-featured business transformation suite created specifically for process management by B.O.C company. The three most important components that we want to model: a) order features that help the operator allocate jobs to machine lines, b) the cost computation of each machine line, and finally, c) the actual decisions that the operators use to select the suitable printing line.

Each order when given by a customer is characterized by several features. Afterwards, the order is enriched with some additional information coming from the customer. The ones that we exploit for the Digital Twin are presented in the Table 1.

Table 1: Dataset for the digital twin

	Description
ID	auto-increment integer identifier of a printing order
Delivery time (days)	the number of days that the order shall be printed and delivered to the customer
Ink Varnish	the ink varnish selection is considered as a boolean attribute with True or False possible values
Colour (4 or >4	This categorical variable denotes the colour requirements of the particular printing assignment. In offset printing, the most requested colour requirement is the 4-colour printing (class 1), followed by 4+1 colour printing (class 2 that involves the use of special / pantone colours, for example gold and/or silver) and grayscale printing (class 3 entailing only black and white colours)
Quantity	the number of paper pieces requested in this specific order. Quantity takes integer values extending to up to large numbers, depending on the type of printing assignment (e.g. newspaper, poster, etc.). Type: Each specific printing order has an associated discrete type of category, also denoting required specifications related to the post-press procedure
Quality	The quality of the paper is associated with the specific printing order. The *Quality* parameter takes string values depending on the properties of the requested paper. The most used ones are 'Velvet' (the most frequently used), 'Uncoated', and 'Illustration/Gloss' paper quality
Job name	each job comes with a small description, mostly including the customer's name, and requested type
Sides	a job may be required to be 1-side or 2-side
Weight (gr)	the weight of the paper to be used in the printing process, measured in grams
Press sheets	it is a larger than-requested sheet that fits multiple smaller printed sheets. For example, a press sheet may include multiple pages of a book, which are then cut to create single pages. Its maximum size may be 70x100 cm
Dimension	the dimensions of a printing order come in the following format: 350x280 mm - 500x280 mm
Cost 4-col	the cost of 4-col line to print the order
Cost 5-col	the cost of 5-col line to print the order
Cost 8-col	the cost of 8-col line to print the order
Machine ID	the machine line to print the order

The next step towards the Digital Twin is to create a digital representation of the decision process, via an interactive diagram. The diagram reflects how the printing company selects a specific printing-machine line for a given order:

1. First, the delivery time is checked. If the delivery date is less than 2 days, the simpler digital printers are selected.
2. The next step is looking at the ink varnish, given by the job order. If the ink varnish selection is True, the order is printed on the 5-colour line. If the ink varnish is false, the algorithm proceeds to the next step.
3. Next, the job colour feature is checked. If the job colour is larger than 4 colours, the 5-colour line is selected. If the job colour is equal to or less than 4 colours, we move to step 4.
4. The next features to be used are Quantity and Printing sheet. If the quantity of printing sheets is less than 500, the 4-colour line is selected to print the order. If the value is equal to or higher than 500, the algorithm proceeds to next step.
5. Next, the paper's weight is checked. Books, journals, newspapers, and magazines fall under the category of weight equal to or less than 170 grams. On the other hand, posters, leaflets, business cards, and folders have a weight higher than 170 grams. If the weight is higher than 170g, the job is to be printed on the 4-color or 5-color lines, depending on the cost. The cost calculation is done by the pre-quotation department. If the weight is less than 170g, we move to step 6.
6. Dimension is the next feature to be taken into consideration. Here, the dimension is transformed to "True 4-col" or "False 4-col". This means that either the job can be printed on the 4-colour line or not. If the value is "True 4-col", a cost comparison between all the lines is performed. In the other case of the dimension being "False 4-col", a cost comparison between the 5-colour and the 8-colour lines is done. The line with the lowest cost gets to print the job.

Before reaching the final step of printing, a comparison with previous jobs on similar lines is made. If no exception arises, the printing process begins. If an exception arises, the job is reallocated to a new machine. In the case of exceptional circumstances, a job can be outsourced to another site.
The following Figure 4 presents the physical process in full detail: features, inputs, outputs, and decision rules. Rectangles represent actions or processes; diamonds represent conditions and lines represent the flow.

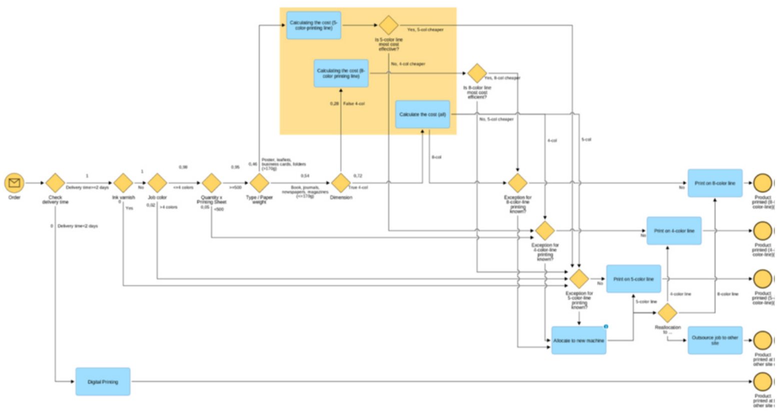

Figure 4: the machine allocation decision diagram, created with ADONIS

3. END-TO-END DATA LIFE CYCLE MANAGEMENT

The ability to cross-reference and integrate heterogeneous data sources is becoming increasingly essential for companies, helping them to gain a competitive advantage. Utilizing these sources, AI-driven applications are transforming vital sectors such as healthcare, manufacturing, and transportation, among others. Within this shift, Data Spaces have evolved as a fundamental framework for facilitating data sharing and exchange, particularly within the context of the European Strategy for Data (European Commission, 2020). This EST seeks to establish a unified data market, augmenting Europe's global competitiveness and data sovereignty (European Commission, 2023).
Despite this progress, the governance of the data life cycle, has not kept pace with technological advances. This gap is especially evident when dealing with semi-structured data from numerous evolving sources,

where manual governance leads to fragmented data silos, making clear the need for a systematic and standardized approach to data ingestion, integration, and analysis to support new data-driven business models and data spaces. Additionally, while current research and development (R&D) efforts focus on specific aspects of the data life cycle (e.g., scalable data management or AI explainability), governance remains largely overlooked. Most organizations still develop their own data architectures, a resource-intensive process that only large tech firms can efficiently manage. Examples include Amundsen (Amundsen, 2019) and Databook (Uber, 2018). Although cloud providers offer layered solutions, data flow design remains ad-hoc, leaving smaller organizations at a significant disadvantage.
To address this challenge, a comprehensive framework is needed to govern the full data life cycle, enabling secure, interoperable, and scalable AI-driven applications for organizations of all sizes. This framework must support large-scale data from diverse, distributed sources to foster efficient data sharing and exchange.

3.1 CyclOps Concept

CyclOps (CyclOps, 2024) aims to redefine the end-to-end data life cycle, enabling organizations to adopt a data-driven culture and utilize data spaces. Achieving this requires automating complex data transformations across large-scale heterogeneous datasets. Currently, this task is handled manually through specialized technologies, making It impractical for large-scale settings. In contrast, CyclOps automates the generation of data processing pipelines, covering the entire spectrum from data consumption to the serving of data, models, and services. In detail, through automated data governance, CyclOps orchestrates the entire data life cycle, ensuring data sharing and exchange in data spaces. A key element of this, is the use of semantically rich metadata, beyond current standards like DCAT, enabling the automation of the components in the data pipeline and content publication. Specifically, utilizing Knowledge Graphs (KGs), CyclOps enables formal models to represent data and metadata, ensuring context and interoperability in line with the FAIR Guiding Principles (Findability, Accessibility, Interoperability, and Reusability) (Wilkinson et al., 2019).
This process is guided by user input and requirements through a Human-in-the-Loop approach, prioritizing a user-centric approach, towards allowing organizations to conduct contextualized analyses, without requiring data management expertise. CyclOps operationalizes the data life cycle through a layered, bottom-up framework composed of five key components, as showcased in Figure 5:

1. DataOps: Automates data management tasks, such as discovery, curation, quality, and integration;
2. AIOps: Offers a decentralized repository of AI algorithms and makes the resulting AI models accessible;
3. Data and Execution Abstraction Layer: Facilitates the development of distributed data preprocessing and analysis;
4. Intent-based Human Interface: Provides a human-centric interface to automatically provide services to exchange, explore, visualize and integrate with data spaces, driving the automatic execution of analytical pipelines;
5. Smart Data Governance and Trust: Acts as a transversal semantic layer that regulates the system, enforces data-related rights, obligations and responsibilities (e.g., GDPR and AI Act compliance), while ensuring interoperability, security and privacy;

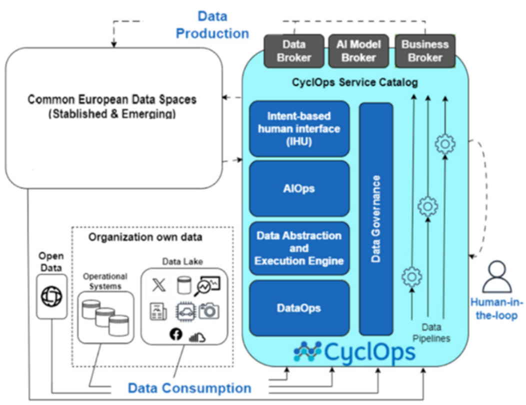

Figure 5: CyclOps conceptual view

3.2 Manufacturing Data Space Use Case

The European printing manufacturing sector plays a pivotal role in employment and skill development, currently facing increasing pressure from global competition, rising costs, and stringent environmental regulations. A significant challenge within the industry is the vast amount of unstructured data generated daily, much of which remains underutilized. Effectively harnessing this data presents an opportunity to optimize production processes, enhance supply chain management, and support the EU's vision for a competitive and sovereign data economy (Kalafatelis et al., 2023a; Kalafatelis et al., 2023b). CyclOps is uniquely designed to address these challenges by transforming unstructured data into actionable insights, enabling manufacturers to meet complex market demands and regulatory requirements more efficiently.

This case study considers a scenario in which two offset printing companies have integrated CyclOps into their operations to tackle persistent issues such as the minimization of their: a) environmental impact, b) frequent product defects, c) machine downtimes, and d) rising production costs. Despite the continuous generation of large volumes of operational structured and un-structured data, these companies have struggled to leverage this information effectively. By adopting CyclOps, they aim to convert their data into valuable insights, optimize production processes, and align with Industry 5.0 principles.

The implementation begins with the DataOps component, which ensures the integration and curation of high-quality data from various factory sources, maintaining data integrity throughout operations. The Data Governance and Trust layer enables secure data sharing, fostering collaboration with partners, suppliers, and research organizations, creating a Manufacturing Data Space environment. This secure environment promotes innovation while ensuring compliance with data privacy regulations. By integrating their data within CyclOps, the companies can utilize the AIOps component to train or utilize existing machine learning (ML) models available on the platform. Meanwhile, the Data and Execution Abstraction Layer manages large-scale data storage and distributed deployment. In scenarios requiring model training, companies can employ Federated Learning (FL) to collaboratively train models on distributed data, such as data collected directly from printing machines, without compromising data privacy. This approach enhances model generalizability, by transmitting only model parameters instead of raw data, thus preserving confidentiality (Sarlas et al., 2023; Angelopoulos et al., 2024). For example, lightweight predictive maintenance models can be trained to monitor printing equipment in real time, proactively identifying potential failures to reduce machine downtime. This predictive capability can minimize costs and enhance productivity by addressing equipment malfunctions proactively (Kalafatelis et al., 2024). Figure 6 showcases the FL training scenario deployed across two printing factories. Additionally, companies can leverage Transfer Learning to adapt pre-trained models to their production line needs, as showcased in Figure 7. This quick adaptation ensures that the benefits of data processing are realized across different machine setups, providing the scalability needed to manage increasing data volumes as the companies expand or modify their printing operations. To enable factory managers and supply chain operators to explore production data in real-time, the Intent-based Human Interface further enhances this scenario setup. By transforming complex datasets into actionable insights, this interface helps identify inefficiencies and guides adjustments to machine settings and workflows. As a result, companies can reduce product defects and material waste, thereby enhancing both operational efficiency and environmental sustainability.

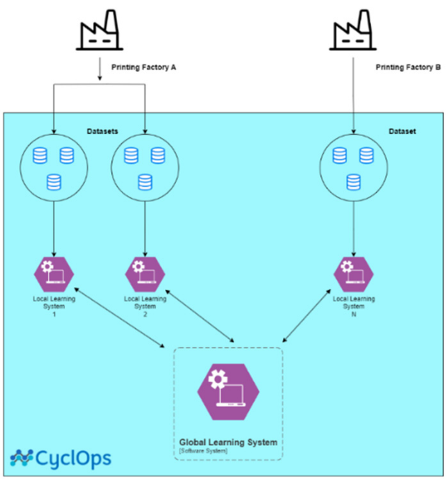

Figure 6: A high-level view of the Federated Learning scheme deployment across two printing companies

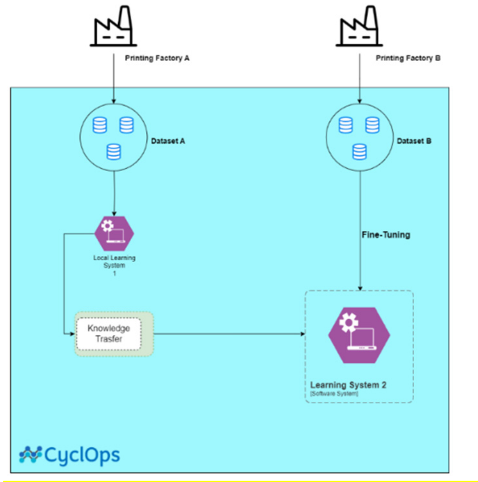

Figure 7: A high-level view of the Transfer Learning scheme deployment across two printing companies

4. CONCLUSIONS

Printing environments exhibit significant unpredictability, making errors an inevitable and prevalent occurrence. These faults can significantly affect several dimensions, from direct economic loss to the environmental consequences of squandered resources. Towards this direction, this study presents two distinct use cases in Offset Printing aimed at enhancing production processes.

The analysis and digitalization process of the printing production line brings many advantages. First, modelling the machine selection process, has enabled the documentation and visualization of real-world decision-making. This process can now be easily simulated in a safe and cost-efficient digital environment, facilitated by a Digital Twin. For new print orders, he simulation accurately predicts the most suitable machine line, allowing operators to input multiple orders and receive optimized recommendations for each. Additionally, past orders can also be checked and validated. The digital model also revealed that some jobs could be processed by multiple machine lines, allowing greater flexibility. Furthermore, the simulation environment provides a platform for parameterizing datasets and fine-tuning configurations related to machine selection. Using ADONIS, different setups can be efficiently tested, enabling informed decision-making with minimal cost. As a result, the integration of new machines into the production line becomes faster and more efficient.

Another important benefit of the Digital Twin is the potential to reduce printing costs. Analyzing a dataset of thousands of orders revealed that, for many, more cost-effective machine lines could have been selected. A comparison between the simulated and actual machine lines demonstrated a cost reduction of approximately 5%, highlighting opportunities for further optimization. Although scheduling conflicts and existing machine utilization may account for some of the inefficiencies, this finding underscores the need for improvements in resource scheduling. In light of this, printing companies are encouraged to explore new ways to optimize efficiency and reduce costs, including partnerships and innovative projects that could address broader challenges within the production line. For instance, digital tools could be employed to automate scheduling and dynamically allocate resources, optimizing both operational costs and human resource management.

Furthermore, the implementation of CyclOps brings transformative benefits to the printing industry by automating tasks such as quality control using ML models, which reduces machine downtime and enhances production efficiency. By leveraging CyclOps, companies can automating routine tasks, enabling workers to focus on higher-value activities, thus reinforcing the human-centric ethos of Industry 5.0. Another notable advantage of CyclOps is its ability to break down data silos through the development of specialized Data Spaces. This fosters greater collaboration with external organizations and suppliers, contributing to a more integrated and competitive data economy, particularly within the EU manufacturing sector.

Concerning future research avenues, this could be focuses on expanding the capabilities of Digital Twins and CyclOps to incorporate real-time ML algorithms that dynamically adjust production strategies based on evolving data patterns. Additionally, investigating the potential for cross-industry collaborations to share data and best practices through enhanced data ecosystems may also yield valuable insights for the broader manufacturing sector. In conclusion, this study demonstrates that digitalization, simulation, and data-driven approaches can significantly improve the efficiency, sustainability, and flexibility of printing processes. By leveraging technologies like Digital Twins and platforms such as CyclOps, printing companies can not only enhance their operational performance but also contribute to a more sustainable and interconnected manufacturing ecosystem.

5. ACKNOWLEDGMENTS

This research initiative is supported by the European Union's Horizon Europe Framework Programme for Research and Innovation, under the CyclOps project (Grant Agreement Project 101135513). Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the CyclOps consortium and the European Union, which cannot be held responsible for them.

6. REFERENCES

Amundsen. (2019) *Amundsen Data Discovery and Metadata Platform.* Available from: https://www.amundsen.io/amundsen/ [Accessed 13th Sep 2024].

Angelopoulos, A., Giannopoulos, A., Nomikos, N., Kalafatelis, A., Hatziefremidis, A. & Trakadas, P. (2024) Federated Learning-Aided Prognostics in the Shipping 4.0: Principles, Workflow, and Use Cases. *IEEE Access.* 12, 6437-6454. Available from: doi: .1109/ACCESS.2024.3350777

CyclOps. (2024) *Automated end-to-end data life cycle management for FAIR data integration processing and re-use*. Available from: https://www.cyclopsproject.eu/ [Accessed 13th Sep 2024].

European Commission. (2020) *A European strategy for data. Brussels, European Commission.* Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0066 [Accessed 13th Sep 2024].

European Commission. (2023) *European data spaces – Scientific insights into data sharing and utilisation at scale*. Luxembourg, Publications Office of the European Union.

Kalafatelis, A.S., Nomikos, N., Angelopoulos, A., Trochoutsos, C. & Trakadas, P. (2023A) An Effective Methodology for Imbalanced Data Handling in Predictive Maintenance for Offset Printing. In: *International Conference on Mechatronics and Control Engineering, Singapore, Springer Nature Singapore.* pp. 89-98.

Kalafatelis, A.S., Stamou, N., Dailani, A., Theodoridis, T., Nomikos, N., Giannopoulos, A., Tsoulakos, N., Alexandridis, G. & Trakadas, P. (2024) A Lightweight Predictive Maintenance Strategy for Marine HFO Purification Systems. Available from: doi: 10.13140/RG.2.2.25002.22724

Kalafatelis, A.S., Trochoutsos, C., Giannopoulos, A.E., Angelopoulos, A. & Trakadas, P. (2023B) A Stacking Ensemble Learning Model for Waste Prediction in Offset Printing. In: *Proceedings of the 2023 10th International Conference on Industrial Engineering and Applications.* pp. 267-272.

Sarlas, A., Kalafatelis, A., Alexandridis, G., Kourtis, M.A. & Trakadas, P. (2023) Exploring federated learning for speech-based Parkinson's disease detection. In: *Proceedings of the 18th International Conference on Availability, Reliability and Security*. pp. 1-6.

Uber. (2018) *Uber Data Book*. Available from: https://www.uber.com/en-GR/blog/databook/ [Accessed 13th Sep 2024].

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J. & Mons, B. (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*. 3, 160018. Available from: doi: 10.1038/sdata.2016.18