CEN/CLC/WS BIOMAT

Date: 2024-06 (v2.0)

prCWA XXXXX:20YY

Secretariat: UNE

Methodology for data-driven management of production processes

Contents

Page

[NOTE to the drafter: To update the Table of Contents please select it and press "F9". To recreate the Table of Contents, select *Custom Table of Contents – Options* and choose the appropriate headings/titles to display. For further instructions, see the *CEN Simple Template Quick Start Guide*.]

European foreword

(To be completed afterwards)

This draft of the CWA is issued for pubic commenting before its publication.

This CEN/CENELEC Workshop Agreement (CWA XXXXX:YYYY) has been developed in accordance with the CEN and CENELEC Guide 29 "CEN/CENELEC Workshop Agreements – A rapid prototyping to standardization" and with the relevant provisions of CEN/CENELEC Internal Regulations - Part 2. It was approved by a Workshop of representatives of interested parties on XXXX-XX-XX, the constitution of which was supported by CEN following the public call for participation made on 2024-02-23. However, this CEN/CENELEC Workshop Agreement does not necessarily include all relevant stakeholders.

The final text of this CEN/CENELEC Workshop Agreement was provided to CEN and CENELEC for publication on XXXX-XX-XX.

Results incorporated in this CWA received funding from the European Union's Horizon 2020 research and innovation programme, project BIOMAT under grant agreement No 953270.

The following organizations and individuals developed and approved this CEN/CENELEC Workshop Agreement:

XXXXXXX

Attention is drawn to the possibility that some elements of this document may be subject to patent rights. CEN and CENELEC policy on patent rights is described in CEN and CENELEC Guide 8 "Guidelines for Implementation of the Common IPR Policy on Patent". CEN and/or CENELEC shall not be held responsible for identifying any or all such patent rights.

Although the Workshop parties have made every effort to ensure the reliability and accuracy of technical and non-technical descriptions, the Workshop is not able to guarantee, explicitly or implicitly, the correctness of this document. Anyone who applies this CEN/CENELEC Workshop Agreement shall be aware that neither the Workshop, nor CEN and/or CENELEC can be held liable for damages or losses of any kind whatsoever. The use of this CEN/CENELEC Workshop Agreement does not relieve users of their responsibility for their own actions, and they apply this document at their own risk. The CEN/CENELEC Workshop Agreement should not be construed as legal advice authoritatively endorsed by CEN and CENELEC

Introduction

The purpose of this document is to facilitate the implementation and optimization of production processes through data and to increase knowledge about them.

The need to cover the start-up arises in the context of the BIOMAT Project, funded by the Horizon 2020 programme, with contract number 953270. This project aims to develop innovative bio-polyurethane foams, modified with nanoparticles, to replace the current ones (derived from petroleum) and to demonstrate it in five sectors: furniture, bedding, construction, building and automotive.

BIOMAT is a representative project of other public and private initiatives in the domain of the Circular Economy characterized by the following needs: a) reducing dependence on petroleum by replacing it with natural and sustainable products, b) managing the heterogeneity of natural raw materials and, c) overcoming the complexity of providing products with the functional characteristics demanded by the market.

In this context, the needs this methodology responds to are aimed at increasing knowledge about production processes through data, answering the following questions:

- *What are the stages of transformation from raw materials to the final product and what components are used? That is, what is the physical sequencing?*
- *What technology is used to support these transformations? That is, what is the layout?*
- *Where should the measurements of the relevant variables, raw materials, process and final product be taken?*
- *What is the mathematical model of the behaviour of a set of properties of the final product as a function of raw material and process variables?*
- *Is it possible to provide traceability to both the final products and the mathematical models that explain them (whether they are based on classical mathematical techniques, artificial intelligence or hybrids)?*
- *Knowing a valid final product combination, is it possible to optimize the process through a model-based scenario analysis, which saves the expenditure of resources?*

With this approach, the document answers these questions, proposing a step-by-step methodology that allows the proposed objectives to be met: improving knowledge of processes, especially in the field of the circular economy, facilitating their implementation and optimization through the intensive use of data.

The main expected benefits of this methodology are:

1) *Facilitating the replicability, stability and scalability of innovative processes*
2) *Facilitating the understanding of production processes (e.g. bioprocesses)*
3) *Saving start-up and optimization costs*
4) *Providing traceability to both final products and mathematical models (based on Artificial Intelligence, classical or hybrid techniques)*

1 Scope

This CWA presents a methodology for the data-driven management of production processes from inception to operation, which allows to document their lifecycle and gain knowledge through its application. It provides a description of the methodological approach along with the introduction to its primary objectives.

The objectives of this methodology encompass traceability, involving:

- Structured documentation of data related to observed parameters, materials, and the physical infrastructure equipment used in processes.
- Data acquisition, which facilitates collection of measured data.
- Data analysis, with interfaces designed for extracting insights through charts or dashboards.
- Simulation capabilities, enabling the execution of models to anticipate outcomes before production initiation.

The methodology is preferably to be integrated into an information system that offers the aforementioned capabilities. It also serves as a communication tool, fostering collaboration between stakeholders, such as producers and customers. However, this CWA does not impose any specific technological framework for this system.

NOTE: This methodology does not constitute a use case for digital twin adoption, although their abilities may align with the capabilities of a digital twin.

This CWA is intended to be used by industries characterized by the presence of production processes or engineering consultancies working with them. For instance: manufacturing of materials or products, water and waste treatment and valorisation of subproducts, etc.

2 Normative references

There are no normative references in this document.

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

ISO and IEC maintain terminology databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp/

— IEC Electropedia: available at https://www.electropedia.org/

3.1
Alert
Inform subscribed system clients about the occurrence of a specified event by a predefined message or indication

[SOURCE: IEC Electropedia, International Electrotechnical Vocabulary-IEV 351-43-06]

3.2
Batch
Set of all products in one family and/or group and identified as such and put forward at one time for checking compliance

[SOURCE: IEC Electropedia, International Electrotechnical Vocabulary-IEV 845-27-134]

3.3
Data store
Persistent repository for digital information

Note 1 to entry: A data store can be accessed by a single entity or shared by multiple entities via a network or other connection.

[SOURCE: IEC Electropedia, International Electrotechnical Vocabulary-IEV 741-01-13]

3.4
Design of experiments
Design of experiments (DOE) is a statistical and mathematical tool to perform the experiments in a systematic way and analyze the data efficiently.

[SOURCE: Sethuramiah, A., & Kumar, R. [4]]

3.5
Equipment
Single apparatus or set of devices or apparatuses, or the set of main devices of an installation, or all devices necessary to perform a specific task

[SOURCE: IEC Electropedia, International Electrotechnical Vocabulary-IEV 151-11-25]

3.6
Functional component
Functional building block needed to engage in an activity, backed by an implementation

[SOURCE: IEC Electropedia, International Electrotechnical Vocabulary-IEV 741-01-19]

3.7
Process
Production process
Set of interrelated or interacting activities which transforms inputs into outputs

[SOURCE: ISO/IEC/IEEE 12207:2017, 3.1.33]

3.8
Sensor
IoT device that measures one or more properties of one or more physical entities and outputs digital data that can be transmitted over a network

[SOURCE: IEC Electropedia, International Electrotechnical Vocabulary-IEV 741-02-09]

3.9
Simulation
Use of a similar or equivalent system to imitate a real system, so that it behaves like or appears to be the real system

Note 1 to entry: Simulation serves the purpose of analysing the future behaviour of a system, i.e., making predictions, or the purpose of reasoning on the past behaviour in order to analyse failures. For performing simulations, a model is needed together with actualized sets of data and a platform able to execute the simulation.

[SOURCE: ISO/IEC 20924:2024]

3.10
Traceability
Ability to trace the history, application or location of an object

Note 1 to entry: When considering a product or a service, traceability can relate to: — the origin of materials and parts; — the processing history; — the distribution and location of the product or service after delivery.

Note 2 to entry: In the field of metrology, the definition in ISO/IEC Guide 99 is the accepted definition.

[SOURCE: ISO 9000:2015]

4 Basic functional component: Data components

4.1 General

The main goal of having a data strategy is to highlight the significance of employing structured data collections, particularly in scenarios involving heterogeneous data sources, both automatic and manual (paper-based or manually typed) media, including sensors, laboratory experiments, and literature. This represents the initial phase in the execution of the methodology, as it lays the groundwork for configuring processes and gathering relevant data.

To develop an effective data strategy to monitor, control and model the production process, it is necessary to record certain information from/of the process that serves as a starting point for its use afterwards (see Figure 1). The key elements of this strategy are:

1. **Production processes**: representation of the stages in which the materials are processed, indicating the equipment required, the materials that enter the process and the products created at it, specifying the parameters measured for stage and material involved.
2. **Materials**: catalogue of raw materials needed to produce the final product and all the semi-elaborated and intermediate products, as well as waste and other type of materials used in the process.
3. **Equipment and blocks**: as production processes are physically executed in equipment (e.g., a reactor) which can be grouped into blocks, the process variables or process inputs and outputs should be linked with them and also with sensors measuring the observed parameters.
4. **Sensors:** integrated within the production process environment, installed as part of one equipment or as an equipment themselves, they are designed to capture and collect data pertaining to various physical parameters (e.g. temperature, flowrate, stirring or any other relevant metric).

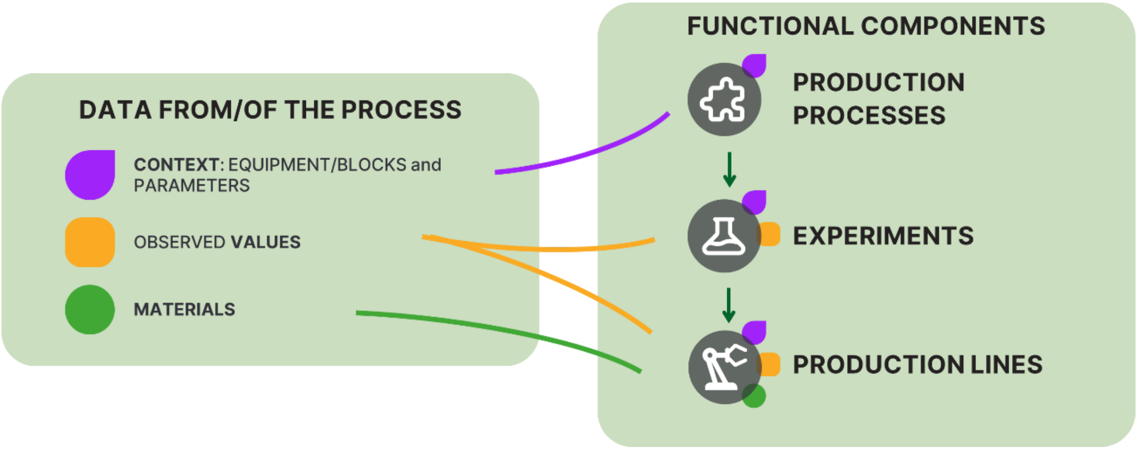

Figure 1 — Data from process and its use in the functional components of the methodology

4.2 Production processes

The descriptive parameters of interest related to the process are identified through a conceptual flowchart consisting of inputs, outputs, stages, and unions.

As a first step, the production process or processes (in case there is more than one production line) shall be identified. The registration of a production process includes three phases:

1. Start of creation of the process: description of the process (location, denomination, description or status).
2. Configuration of materials, stages and sub-processes: the elements of the process and their relationships are established.
3. Confirmation of the process: once the configuration has been completed, the registered process is set as active/inactive as the experiments and production phase, described below, rely on the process definition.

4.3 Materials

The creation of a material catalogue (broad material classification with parameters) ensures consistence, as all products that can be manufactured with their properties (e.g., a PUR foam with "x" density), the raw materials and additives to be used for their manufacture are registered. When entering the production phase, this enables the recording of measured specific characteristic values for each of them.

A double classification into families and subfamilies is used. Materials can be classified into 6 families:

- raw material,
- semi-elaborated product,
- intermediate product,
- final product,
- waste,
- other type.

Each one should be differentiated in several subfamilies as necessary.

Additionally, a third additional classification may be considered, if necessary, independent of the previous ones, according to specific needs of the user.

Each material from any of the families, shall be defined by categories: parameters to be controlled in the process (e.g. density, grain size, etc.).

4.4 Equipment and blocks

All the existing equipment used in the process shall be identified to establish a connection between the information (parameters) represented in process diagrams and the physical apparatuses employed and installed in the process. Records shall include:

- Equipment: machines and tools used in the process.
- Blocks: logical grouping of equipment.
- Connection of monitored information to equipment and blocks: data and their relationship with the processes in which they are involved shall be linked with the equipment and blocks. Hence, the recorded parameters shall be linked with the equipment in which they are measured.

4.5 Sensors

Sensors track relevant parameters informing the status of the production process and constitute a particular case of equipment with a dual aspect:

- Hardware information: sensors shall be recorded as a special equipment to include their key attributes such as manufacturer, model or specifications.
- Measured information: data acquired by sensors shall be transferred by methods as described in subclause 6.2 'Data integration'.

5 Production start-up

5.1 General

The primary objective of the production start-up is to provide the required support for correlating the data from utilized materials and the values of process properties, both when initiating new processes and when managing established production processes. Serving as the second stage in the execution of the methodology, it is necessary for:

- Enabling the traceability of the products manufactured,
- Collecting data for model generation and production monitoring using the structured framework defined in Data Strategy.

The functional components used in the production start-up to aid in correlating data with production stages are (see Figure 2):

1. **Batches:** catalogue of records that provide traceability by establishing a unique relationship between the series of materials produced, and the measured parameters and the materials used during their manufacturing production process.
2. **Experiments**: predefined configuration for a process which consists of the collection of recorded values for every parameter within a process. It documents the trial stages leading up to a final production line for a production process.
3. **Production line**: integrated collection of data regarding the product specifications, the associated production process, the reference experiment, the used materials and the measured parameter values for the production process.

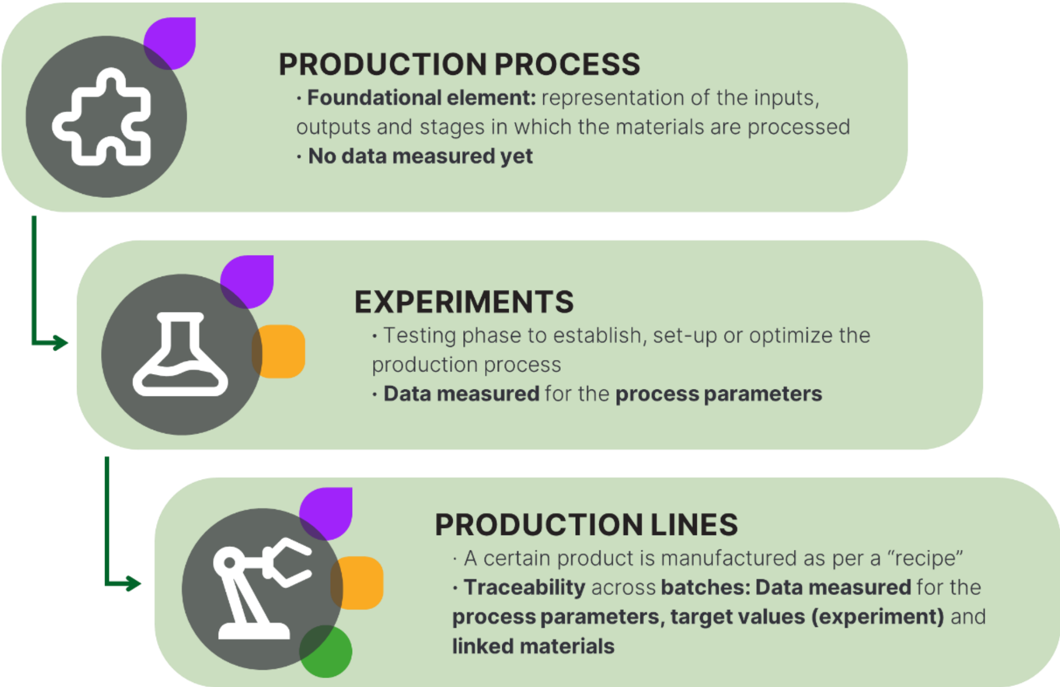

Figure 2 — Functional components used in the production start-up and their correlation with the production stages

5.2 Batches

Traceability is achieved by creating a unique correlation between the series of produced materials, the measured parameters, and the materials used in the manufacturing process.

Batches are categorized as follows:

1. **External batches**: used to identify series of the same external materials incorporated into the production process. They should be recorded to ensure full traceability of internal batches but they are not strictly necessary.
2. **Internal batches**: comprise data essential for tracking, monitoring and reporting all the obtained products throughout a production process's lifecycle.

For each internal batch, specific details shall be identified:

- external batch of materials used,
- measured property values registered during production,
- target values set for the process parameters based on a selected experiment.

Each internal batch shall be defined within a production line. This line shall connect a production process with an experiment, enabling the association of recorded data and used materials by generating new internal batches.

5.3 Experiments

During the start-up configuration of a production process, experiments are the test executions needed to adjust the parameters to obtain a certain final product.

To effectively record these tests, experiment registration includes the following steps:

1. Creation of the contextual information (e.g. Name, description details, etc.).
2. Selection of the process where they are carried out, that consists of inputs, outputs, stages and unions comprising different parameters.
3. Link the set of values measured for each parameter.

Every experiment shall be marked as active or inactive, depending on they being successful and available to be used in a production line or deprecated and not available for their use in a production line.

Since experiments are recorded for selected processes, the registered values of several experiments for a same process may be compared to take advantage of their data analysis.

Techniques such as Design of Experiments can be considered to assist the generation of experiments. They are useful to optimize the execution of experiments by reducing the required number of trials while maximizing the mapped knowledge space and facilitating their replicability.

In case they are considered, the experiments suggested by these techniques should be easily recorded as the actual experiments in which the different combinations of set-ups are applied and the results obtained are measured.

5.4 Production line

A production line serves to log the measured values for a production process and the specific materials used considering a predefined configuration, i.e. an experiment, selected during the experimentation phase.

This represents the final phase in building the traceability for the data managed in a production process using this methodology. Thus, available information (materials, measured parameters and established configuration) is linked to the resulting product. The creation of a production line includes these steps:

1. General information:
    a. Definition of the product to be manufactured: it shall be registered in the materials catalogue.
    b. Selection of the reference process for the production line: it shall be assigned to the production line.
2. Association of specific materials: configuration of materials used in the process. Their registered properties in the materials catalogue or their measured values shall be considered.
3. Association of target values for process parameters: this constitutes the recipe to be followed in the production line. A registered experiment can be used as reference. If not, the user shall define the corresponding target values and tolerances for every parameter.
4. Link of measured data: data related to the process shall be linked to the production line associating their measurements with their respective parameters. Every batch shall record these values when it is run.

6 Monitoring

6.1 General

Data acquisition is crucial for monitoring the performance status of production processes during their operation. The data shall be stored in a data store populated automatically via Machine-to-Machine methods or by human intervention, involving file storage or manual recording.

Effective data capture ensures the creation of a comprehensive collection and accurate referencing of information to the specific process and the corresponding event when it is recorded.

Data acquisition management in monitoring shall follow these steps:

1. **Data integration**: first step encompasses the use of various tools for storing information, combining automated methods with protocols for seamless data capture from databases or SCADAs, and manual methods where structured templates are used for data recording.
2. **Data association with a process**: after data is consolidated in a common data store, in the second step the relationships shall be registered between the measured parameters, their specific location in the production process and the event (e.g. experiment) during which they were recorded.

6.2 Data integration

Data integration refers to the set of methods that allow to transfer the collected information in the process to the data store. These methods are:

- Automated: the different systems shall communicate and exchange data in a reliable, compatible, and scalable manner with the data store. Given the variety of data sources that can provide information, software solutions like API (Application Programming Interface) or REST (Representational State Transfer) may be implemented.
- Manual: templates shall be created to ensure consistency in manual data uploading given a certain reference such as a process or an equipment that cannot transfer machine-to-machine its data.

6.3 Data traceability

The connection of data series with the respective experiments and production lines where they have been measured and the models in which they have been used for training purposes is important for providing context. This activity involves referencing the data to the specific production process and the corresponding event (experiment or production batch) when they have been measured and the models in which they have been used.

Effective data association allows for accurate analysis and interpretation, and shall be facilitated in the following contexts:

- Experiments: the system must allow for the linking of data to specific experiments. This link provides a clear association between the data collected and the experimental conditions under which it was gathered.
- Production line: similarly, the system shall enable the association of data with specific production lines, ensuring that data is accurately correlated with the relevant production parameters.
- Models: since data-driven models are created using data measured in experiments that belong to processes, they shall refer to this source data.

For every data linkage, the system shall register the following elements:

1. **Parameter**: Identify and record the specific process parameter in an experiment or batch to which the data is connected.
2. **Data source**: Document the source of the data, referring to the specific record of the data store where this data was integrated.
3. **Timeframe**: Specify the time window during which the data was recorded.

For every model trained, the system shall provide the following supporting information:

a) **Experiments**: Link to the set of experiments that contain the data that has been used in training the model.

b) **Training information**: The minimum and maximum values of the variables for the set of experiments used to train the model shall be provided as contextual information on the boundaries of the model.

7 Process control assistance

Assisting in the control of production lines by generating alerts based on measured or predicted data is key for reducing production costs and enhancing overall efficiency. Chronologically, it shall be implemented after the initiation of the production phase.

Every process control element includes:

1. **Parameter**: it shall be clearly indicated within the process to which it belongs.
2. **Control method**: it can be based on measured values or predictions from models available for the process.
3. **Alert setting**: alerts shall be configurable to respond to specific conditions or deviations that are described by tracked parameters.

8 Management of models

The management of models for specified production processes, particularly through the use of techniques like machine learning, represents an advanced functional component of this methodology. It effectively leverages the data strategy and management principles established in preceding stages.

This encompasses the utilization of two types of models:

1. **Custom-developed**: created by specialists, these models are tailored to specific requirements and complexities of the production processes.
2. **Generalist:** designed to be accessible and usable by non-specialist users, facilitating broader application and understanding.

The elements considered by this methodology in managing models are:

- Model-process mapping: models either custom or generalist shall be mapped to an existing process. This involves aligning the models' inputs and outputs (parameters to predict) with the data structure of a given process. This alignment maintains consistency between the actual and predicted values.
- Data selection for generalist models: in the development of generalist models, data shall be readily accessible from the data store. This is key to safeguard that the model is built based on relevant and accurate information of the process.
- Model metrics: to evaluate the performance of the models (either generalist or custom) the system shall provide some metrics and charts that allow to determine the validity of the model. For instance, these metrics can be R^2 *score*, *Mean Squared Error* or *Mean Absolute Error*, and a chart plotting real values versus predicted values that can be included.

9 Scenario analysis (digital twin)

The purpose of scenario analysis is to facilitate user interaction with process-related models to evaluate how modifications in the parameters impact final outcomes, thereby allowing the evaluation of various "what-if" scenarios before their actual implementation in the process.

Each scenario analysis shall create one "Execution". Every execution shall incorporate this information:

- **Model**: base simulation.

- **Name**: to identify the execution.
- **Date**: when the execution was created.
- **Values** for input parameters: which will be used to obtain the predicted outputs.
- **Output results**: predicted values considering the input values.

To promote the exploitation of execution results as well as its consistency, all the parameters of every Execution will follow the data structure provided by the process for which the base model was configured.

Bibliography

[1] ISO/IEC 20924:2024, Internet of Things (IoT) and digital twin — Vocabulary

[2] ISO/IEC/IEEE 12207:2017, Systems and software engineering — Software life cycle processes

[3] ISO 9000:2015, Quality management systems — Fundamentals and vocabulary

[4] Sethuramiah, A., & Kumar, R. (2016). Statistics and Experimental Design in Perspective. Modeling of Chemical Wear, 129–159. doi:10.1016/b978-0-12-804533-6.00006-8