Systems & Control Transactions

Original Research Article
Peer Reviewed Conference Proceeding
ESCAPE 36 - European Symposium on Computer Aided Process Engineering
Sheffield, UK. 21-24 June 2026

PSE PRESS

Digital Twin Supported FAIR Electronic Lab Notebooks for Simulated Experiments

Amy Koch[a*], Isabell Viedt[a], Leon Urbas[a]

[a] Technische Universität Dresden, Process-to-Order Group
* Corresponding Author: amy.koch@tu-dresden.de

ABSTRACT

The use of equipment digital twins of standardized, multi-purpose units can accelerate process development and reduce experimental effort. Experimental data are essential not only for identifying critical process parameters and enabling model-based methods within a Quality by Design framework, but also for constructing and validating the simulation models that describe digital twin behavior. To achieve high-fidelity and robust predictive models, structured concepts are required to manage metadata and process-, product-, and resource-specific information exchanged between physical and digital twins. Electronic lab notebooks (ELNs), which contextualize experimental data, must therefore be structured and standardized to ensure interoperability and seamless data exchange. For integration into digital twin workflows and process transfer between equipment instances of the same category, ELNs must comply with FAIR (Findable, Accessible, Interoperable, Reusable) data principles. This work proposes a Semantic Integration Pattern for FAIR-compliant ELNs to document simulated experiments using domain-specific, vendor-independent standards such as DEXPI 2.0 and ISA 88/IEC 61512-1. DEXPI 2.0 supports standardized descriptions of equipment and process steps, while ISA 88/IEC 61512-1 structures experimental recipes. The resulting FAIR ELN Semantic Integration Pattern enables standardized data exchange between ELNs and digital twin tools. Its applicability is demonstrated through an illustrative example documenting simulated saponification experiments in gPROMS® using simple batch reactor model. The results will support subsequent real experiments conducted in a 2L stirred tank unit.

Keywords: digital twins, electronic lab notebooks, gProms, Simulation, Process Design

MOTIVATION

Digital twins can be utilized to support process development, optimization, and scale-up. In modular and multi-purpose process environments, application of digital twins of the equipment enables the reuse of simulation models across different process applications and equipment instances, thereby reducing experimental effort and time-to-process. However, the predictive quality of digital twins critically depends on the availability of high-quality experimental and simulation data together with rich contextual metadata.

Electronic lab notebooks (ELNs) play a central role in capturing experimental data and associated metadata. In current practice, ELNs are often used as flexible documentation tools with limited structure, leading to heterogeneous data representations, incomplete metadata, and poor interoperability. As a result, experimental data rapidly lose their value for reuse, model validation, and transfer between projects, tools, and organizations.

To address these challenges, the FAIR data principles [1] provide a widely accepted framework for sustainable scientific data management. Applying FAIR principles to ELNs is particularly important in digital twin workflows [2], where experimental data from both the laboratory and simulation models must be exchanged between modeling tools, automation systems, and laboratory or pilot-scale equipment. This paper focuses on the documentation of experimental data from simulation models (simulated experiments*)*, which are a key component of model-based process development and Quality by Design (QbD) methodologies.

In this paper, first the relevant background is presented and key domains specific ontology, metadata models, and data models presents. Next, a concept for a structured Semantic Integration Pattern for FAIR-compliant ELNs for simulated experiments based on established, vendor-independent domain specific metadata guidelines are proposed. This Semantic Integration Pattern is explained for a saponification reaction in a 2L batch process using a gPROMS® Process simulation.

METHODOLOGY

Digital Twins in Modular Plants

Digital twins for modular plants have been discussed in terms of a single PEA [3] and for process transfer [2]. Here, the common approach is to use a Product-Process-Resource-Model to structure the digital twin [4] and leverage the knowledge distribution between the equipment vendor and the owner/operator [3] as well as for process transfer between modules [2]. A similar approach is used in the domain of discrete manufacturing with a focus on standardized description of tasks and skills of machinery, e.g. [6], [7], [8]. For modular plants, these equipment-based digital twins can be used to facilitate faster time-to-process. These equipment -based digital twins are integrated into digital twin workflows for process development which typically span multiple asset lifecycle phases, including process concept design, model development, parameter identification, optimization, and scale transfer. In modular plant concepts, digital twins are often associated with standardized equipment classes rather than single physical assets. This enables the transfer of process knowledge between different equipment instances of the same category.

Within such workflows, simulated experiments are used extensively to explore design spaces, analyze process dynamics, and support model-based design of experiments. The results of these simulations must be traceable, reproducible, and interpretable to be reused for decision-making and subsequent validation with physical experiments.

FAIR DATA

A critical consideration for the effective utilization and reuse of experimental data is adherence to the FAIR data principles [1]. Alignment with these principles ensures that metadata is sufficiently enriched with descriptive information that captures the context in which the data were collected, including data quality, collection conditions, and other relevant characteristics. In accordance with FAIR requirements metadata must “be richly described with a plurality of accurate and relevant attributes.” Within the context of ELNs and simulated experiments, these principles imply:

- **Findable:** Experiments and datasets can be located based on product, process, resource, purpose, or author.
- **Accessible:** Data and metadata can be retrieved via well-defined access mechanisms, including access restrictions where required.
- **Interoperable:** Data and metadata use shared vocabularies, formats, and semantics to enable interpretation by humans and machines.
- **Reusable:** Data are sufficiently well-described to allow repetition of experiments and reuse in new contexts.

Applying FAIR principles requires explicit information models and controlled vocabularies rather than ad-hoc documentation.

Electronic Lab Notebooks

The increasing shift towards digitalization in the laboratory extends beyond the use of electronic devices to encompass the networking of experimental workflows and the management of research data in accordance with the FAIR principles. Electronic laboratory notebooks (ELNs) support structured, machine- and human-readable metadata capture, enable direct integration of measurement data, and facilitate data traceability, reuse, and long-term archiving.

NFDI4Chem [9] provides guidelines for managing research data in chemistry and chemical domain specific guidelines for the following aspects: type of experiment, planning of experiments, documentation of experiments, data producing methods, data analysis, and publication.

Simulation-Supported DoE

Model-based Design of Experiments (MBDoE) [10] integrates simulation models, typically formulated as differential–algebraic equations, with classical Design of Experiments to systematically plan informative experiments. Simulation models incorporating prior knowledge of process structure and parameter estimates are used to identify experimental variables with high sensitivity to model parameters, thereby maximizing information gain. This model-driven approach enables efficient identification of optimal process conditions while reducing experimental effort, making it particularly valuable in time- and resource-constrained process development.

Information Models and Standards

In addition to the FAIR Data requirements [1], other ontology and data model standards exist to describe the high-level metadata.

Dublin Core Metadata Terms

The **Dublin Core Metadata Element Set** [11] is a widely adopted standard for describing digital and physical resources to support discovery, interoperability, and long-term access. It consists of 15 core metadata elements, such as *title, creator, subject, description, publisher,* and *date*, which are designed to be simple, generic, and domain independent. Dublin Core is formally standardized as ISO 15836. The standard supports extensibility through qualified elements and encoding schemes, allowing more precise descriptions when required by specific communities.

ProMetaS

ProMetaS [12] is a comprehensive metadata schema designed to document and describe data generated in chemical process engineering and industry to support interoperability and reuse. The schema was developed within the **KEEN research project** to address the diversity and complexity of industrial and research data, improving findability and understanding. In line with FAIR principles ProMetaS builds on international metadata standards (e.g., Dublin Core) and was iteratively refined with domain experts to include only relevant metadata elements. The applicability of this schema has been demonstrated for real use cases in process data collection and management in the KEEN project.

The information model divides the data in the process industry into categories: plant info, process info, chemical, sensor, image control valve, time series, and machine learning.

DEXPI 2.0 for Process and Equipment Descriptions

DEXPI (Data Exchange in the Process Industry) is an initiative of the process industry to create a universal data model for the information described in flow diagrams along the asset lifecycle. DEXPI Plant [13] provides a standardized, vendor-independent representation of process equipment and piping and instrumentation diagrams. DEXPI Plant can be used to describe equipment classes and topology. DEXPI Process serves to describe the block and process flow diagram type information which can be mapped to simulated data, particularly for flowsheet simulation. The DEXPI 2.0 und data model serves to provide a consistent reference between simulation models and physical assets.

ISA-88 / IEC 61512-1 for Recipe Descriptions

In addition to the process description, it is necessary to describe the experimental procedure, or “recipe.” ISA-88 which is formally standardized in Europe with the IEC 61512 [14] specifies a hierarchical recipe model consisting of procedures, unit procedures, operations, and phases. This structure is well suited for documenting experimental workflows in both simulation and laboratory contexts. In the FAIR ELNs, ISA-88 is used to: (1) structure the experimental procedure of simulated experiments, (2) define input parameters, control actions, and execution order, and (3) support repeatability and comparison between experiments.

PEA Harmonization for Modular Plants

Additionally, for modularization further requirements with respect to the documentation are needed. Here, it is important to be able to describe both process and equipment in a modular way. To address this challenge, the Process-to-Order Group has introduced a PEA Harmonization [17] which aligns the hierarchical modular terms from process engineering [16] and automation engineering [17] and links it to the process steps in the DEXPI Process specification [18]. This harmonization enables a direct link from the domain specific terms in modularization to domain specific process steps used in the process industry to describe process phenomena in different levels of abstraction.

REQUIREMENTS

Based on experience from experimental and simulation studies, several challenges motivate this work. First, the same process step can be carried out in models as well as different apparatus configurations, leading to ambiguous documentation. This is an issue for both real experiments as well as simulation runs. A second issue is that equipment of the same category may exist in multiple instances with different geometries, sensors, or performance characteristics. This is important because depending on the purpose of the simulation, the same simulation models can be applied to different equipment instances, but with different parameterization. Another issue is sensors. Sensors with different accuracies and variances are often used without explicit documentation of uncertainty models, however, uncertainty models and estimates play a significant role on the predicted outcome of a simulation. Another challenge is the data itself, experimental data may originate from both online (time-series) and offline (sample-based) sources, which are rarely integrated consistently. These challenges highlight the need for structured, semantically rich ELNs.

Another issue to address is the separation of the domain specific data. Here, it is necessary to have a separation of data and metadata: Raw simulation results, derived results, and metadata must be clearly distinguished. Next, it is necessary to structure the domain specific data. As in the ProMetaS [12], it is necessary to have structured description for the different domains. For ELNs specifically, this means separate descriptions of process, product, and resources. Therefore, equipment topology, process steps, and material information must be described using standardized models.

In addition to the above listed challenges, it is further necessary to delineate between the type of simulation which is being carried out. Here, the follow categories are suggested:

- Static simulations for design space identification,

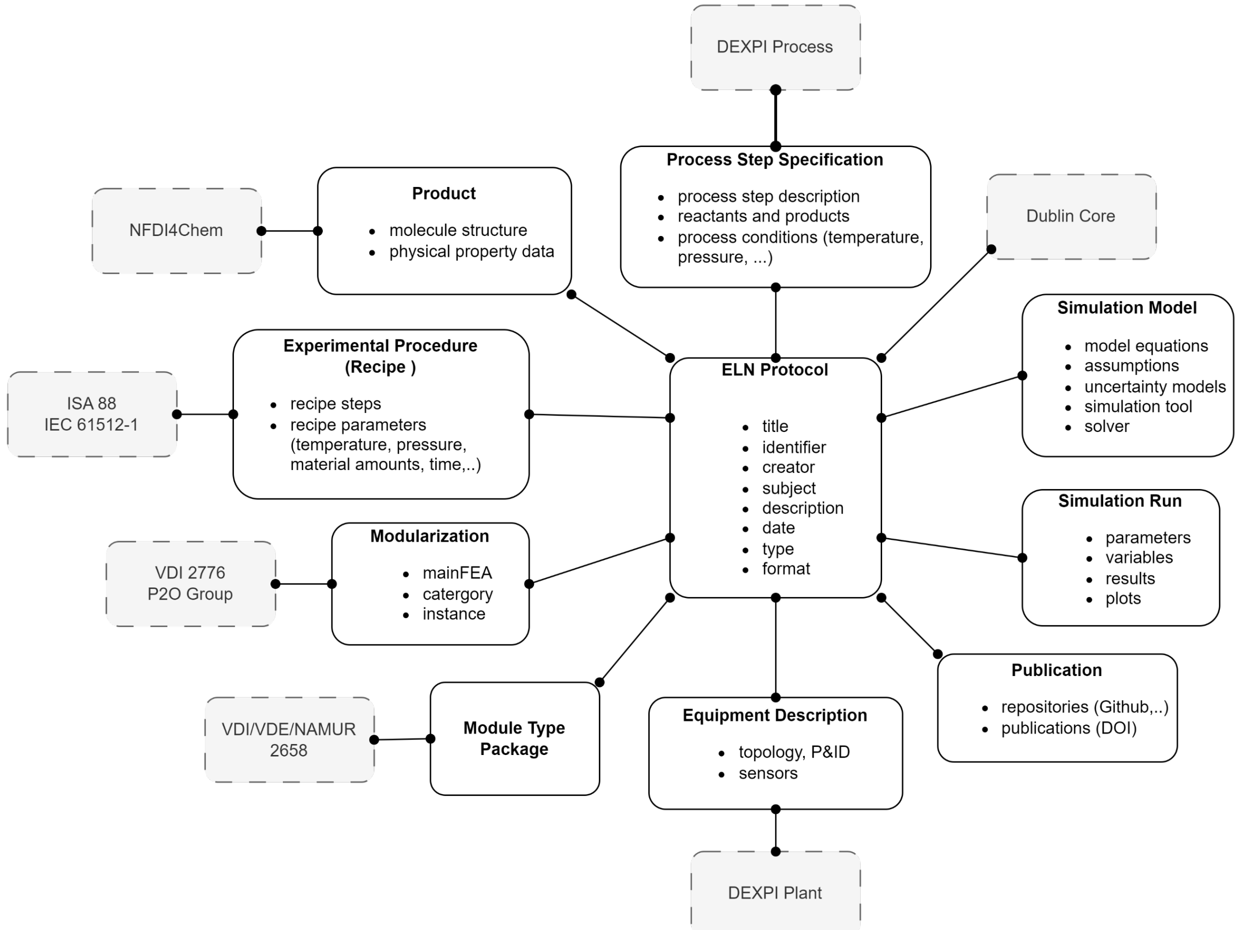

Figure 1. ELN Semantic Integration Pattern for documentation of simulated experiments.

parameter estimation, process design, and optimization

- Dynamic simulations for system behavior analysis, model discrimination, and control loop design
- Stochastic simulations for model-based experimental design and uncertainty analysis/reduction

ELN SEMANTIC INTEGRATION PATTERN FOR SIMULATED EXPERIMENTS

To address the above-mentioned challenges, an ELN semantic integration pattern is proposed for documentation of simulation runs for application in digital twin workflows for modular plants. Here, the Dublin Core [11] was utilized for the ELN Protocol, ISA 88 for the experimental procedure, and ProMetas for domain specific categories. The following domains for ELNs documenting simulated experiments are identified as:

- **ELN Protocol** for the explicit identification of experiments. Each ELN entry must have a unique and persistent identifier and should include metadata according to the Dublin Core [11]. Specifically, experiment title, unique identifier, creator, subject (purpose of experiment), description, data, type of simulation, format of the simulation, and the source.
- The **Product** must include a description of the chemicals. Here, information such as molecule structure, physical property data, physical property models, etc. is included. The NFDI4Chem [9] provides domain specific guidelines depending on the chemical domain.
- The **Process Step Specification** describes the type of process step which is considered and utilized terms from the DEXPI Process specification
- The **Equipment Description** to describe the topology, P&ID and available sensors in alignment with the DEXPI Plant [13] .
- The **Experimental Procedure** must be documented

in a reproducible, machine-readable form. Definition of the experimental procedure using ISA-88/IEC 61512-1 [14] structures to include recipe steps and parameters.

- **Modularization:** Functional Equipment Assemblies (FEAs), module category, and instance as outlined in the PEA Harmonization.
- **Module Type Package:** Link to the data in the real plant along with the Human Machine Interface (HMI).
- **The Simulation Model** to document versions, solvers, and assumptions must be recorded: Documentation of assumptions, observations, and derived quantities.
- The **Simulation Runs** to document the parameters, variables, results, and plots unique to the simulation run.
- **Publication:** Storage of ELN entries in a repository enabling search, access control, and reuse.

This ELN Semantic Integration Pattern ensures that simulated experiments are documented consistently and can be FAIRly [1] integrated into digital twins for modular plants.

ILLUSTRATIVE EXAMPLE

To demonstrate the utilization of the ELN Semantic Integration Pattern, a small illustrative example is provided. Here, a simple batch model created in gPROMS® Process was used to validate kinetics of a saponification reaction at $T = 313\,K$ according to the following equation:

$$CH_3COOC_2H_5 + NaOH \rightarrow CH_3COONa + C_2H_5OH$$

The simulation model employs three equations: Equation 1 describes the concentration profile of each reaction component i over time with the assumption of constant volume V, Equation 2 describes the Arrhenius rate constant kinetics, and Equation 3 describes the reaction rate.

$$\frac{\mathrm{dCi}}{\mathrm{dt}} = -\mathrm{r} * \mathrm{V} \tag{1}$$

$$k = k_{i,\infty} \exp\left(\frac{-E_{A,i}}{R\,T}\right) \tag{2}$$

$$r = kC(1)C(2) \tag{3}$$

Next, for the simulation run values are assumed for the parameters of the Arrhenius rate expressions:

$$k_\infty = 3.5e^7 \frac{L}{mol\ s}$$

$$E_A = 48000\ \frac{J}{mol}$$

Additionally, it assumed that an equimolar ratio of the reactants, 1 = ethyl acetate ($CH_3COOC_2H_5$) and 2 = sodium hydroxide ($NaOH$) is present in the reactor at the start of the reaction time. The results and plots from the simulation run are stored in a gPROCESS Case File.

Figure 2 shows a delimited ELN Protocol in XML utilizing the ELN Semantic Integration Pattern. The ELN Protocol highlights the following domains from the ELN Semantic Integration Pattern from Figure 1: ELN Protocol, Modularization, Simulation Model, and Simulation Run. Here, an example of the general meta data in accordance with the Dublin Core [11] is provided in the ELN Protocol section. The **Modulization** description of M22 which is an instance of a 2L Glass Stirred Tank Unit whose main Functional Equipment Assembly is a reactor. Next, the **Simulation Model** is described according to equations (1) to (3), the modelling assumption of constant volume, along with a list of parameters and units, as well is the solver and any uncertainty assumptions. Lastly, for the **Simulation Run** domain estimates of the Arrhenius parameters is provided as well as the initial amount of moles of the reaction constitutes in the vessel at the start of the reaction. The name of the Case File where the simulation results are stored is also provided in the XML code.

CONCLUSION AND OUTLOOK

The proposed FAIR ELN Semantic Integration Pattern addresses key challenges in documenting simulated experiments for digital twin workflows in modular plants. By combining domain-specific standards with general metadata vocabularies, both semantic richness and broad interoperability are achieved. The integration of DEXPI Plant [13], DEXPI Process [18], ISA-88/IEC 61512-1 [14], and Dublin Core [11] provides a structured, vendor-independent approach to experimental documentation which can be extended to modularization and simulation. However, it should be noted that this work focuses on the concept, thus the illustrative example in Figure 2 is underspecified. This example is not intended to be exhaustive in scope; rather, it aims to describe the ELN documentation method for a simulation run, not the subsequent utilization of the simulated results within the digital twin.

Remaining challenges comprise the effort required for initial model alignment, the establishment of controlled semantic vocabularies, and the development of integration tools. However, the long-term benefits in terms of data reuse, reproducibility, and model fidelity outweigh these initial limitations. Future work will therefore focus on the implementation of tool-supported approaches, the automation of metadata capture, and the extension of the framework to facilitate digital twin–supported information exchange based on documented simulation runs.

```xml
<ELN Protocol>
        <Title = "SimpleBatchRunSaponification001" Identifier = "BatchSaponifcation001" Creator = "Amy Koch" Subject
        "Saponification Kinetic Validation"     Description = "Validation of Reaction Kinetics for Ethyl Acetate Saponification
        Reaction" Date = "25.11.2025" Type = "Simulated Experiment/>
<Modularization>
        <mainFEA = "Reactor" Category = "Stirred Tank Unit" Instance = "2L Glass Stirred Tank Unit"/>
</Modularization>
<Simulation Model>
        <simulation tool = "gPROMS Process Academic Research Version 2025.2.0/>
        <Model ="Simple Batch Model"/>
                <Equation>
                        <!-- Species balances -->
                        <Equation variable="C(1)">
                        d(Ci * V)/dt = -r * V
                 </Equation>
                <Kinetics>
                        <!-- Arrhenius rate constant -->
                        <RateConstant name="k">
                                 k0 * exp(-Ea / (R * T))
                        </RateConstant>
                         <!-- Reaction rate: CH3COOC2H5 (1) + NaOH (2) → CH3COONa (3) + C2H5OH (4) -->
                        <ReactionRate name="r" units="mol/L/s">
                         k * C(1) * C(2)
                         </ReactionRate>
                </Kinetics>
                <Parameters>
                        <Parameter name="R" units="J/mol/K"/>
                        <Parameter name="kinf" units="L/mol/s"/>
                        <Parameter name="Ea" units="J/mol" >
                        <Parameter name="T" units="K"/>
                </Parameters>
                <Variables>
                        <Variable name="M22.N(1)" units="mol/L"/>
                        <Variable name="M22.N(2)" units="mol/L"/>
                        <Variable name="M22.N(3)" units="mol/L"/>
                        <Variable name="M22.N(4)" units="mol/L"/>
                        <Variable name="Volume"  units="L"/>
                </Variables>
        <Assumptions = "constant volume"/>
                        <!-- Volume balance (constant volume) -->
                        <Equation variable="V">
                                dV/dt = 0
        </Assumptions>
        <Uncertainty = "none"/>
        <Solver = "DAEBDF"/>
  <Simulation Run>
        <Parameters>
                <Parameter name="R" units="J/mol/K" value="8.314"/>
                <Parameter name="kinf" units="L/mol/s" value="3.57e7"/>
                <Parameter name="Ea" units="J/mol" value="48000"/>
                <Parameter name="T" units="K" value="313.15"/>
        </Parameters>
        <Variables>
                <Variable name="M22.N(1)" value="0.020043"/>
                <Variable name="M22.N(2)" value="0.020043"/>ee/
                <Variable name="M22.N(3)" value="0"/>
                <Variable name="M22.N(4)" value="0"/>
                <Variable name="Volume" value="0.22"/>
        </Variables>
        <Results = "Simulation_Batch_Recipe_20251125_132208.gSCP-PB"/>
</Simulation Model>
</ELN Protocol>
```

Figure 2. Example XML documentation of the ELN Protocol for a simulated exeriment.

ACKNOWLEDGEMENTS

The authors would like to thank our partners Merck Electronics KGaA and Siemens AG in the industrial research project Lab2Plant for the fruitful discussions as well as P2O-Lab Learning Factory at TU Dresden for information about their modules.

AUTHOR IDENTIFIERS

Author ORCIDs:
Koch A: 0009-0009-4740-0590
Viedt I: 0000-0002-5814-5128
Urbas L: 0000-0001-5165-4459

REFERENCES

1. M. D. Wilkinson *et al.*, “The FAIR Guiding Principles for scientific data management and stewardship, ” *Scientific Data*, vol. 3, no. 1, p. 160018, Mar. 2016, doi: 10.1038/sdata.2016.18.
2. Koch A, Viedt I, Bamberg A, Urbas L. Applying quality by design to digital twin supported scale?up in modular plants. Chemie Ingenieur Technik 97:1127-1137 (2025). https://doi.org/10.1002/cite.70016
3. J. Mädler, I. Viedt, J. Lorenz, and L. Urbas, “Requirements to a digital twin-centered concept for smart manufacturing in modular plants considering distributed knowledge, ” in *Computer Aided Chemical Engineering*, vol. 49, Y. Yamashita and M. Kano, Eds., in 14 International Symposium on Process Systems Engineering, vol. 49., Elsevier, 2022, pp. 1507–1512. doi: 10.1016/B978-0-323-85159-6.50251-7.
4. Bamberg A, Urbas L, Bröcker S, Bortz M, Kockmann N. The digital twin – your ingenious companion for process engineering and smart production. Chem Eng & Technol 44:954-961 (2021). https://doi.org/10.1002/ceat.202000562
5. A. Koch, J. Mädler, A. Bamberg, and L. Urbas, “Towards PEA Matching from Simulation as Part of a Digital Twin Concept for Scale-Up in Modular Plants, ” in *Computer Aided Chemical Engineering*, vol. 53, Elsevier, 2024, pp. 595–600. doi: 10.1016/B978-0-443-28824-1.50100-9.
6. Schleipen M, Lüder A, Sauer O, Flatt H, Jasperneite J. Requirements and concept for plug-and-work. at - Automatisierungstechnik 63:801-820 (2015). https://doi.org/10.1515/auto-2015-0015
7. Pfrommer J, Stogl D, Aleksandrov K, Escaida Navarro S, Hein B, Beyerer J. Plug & produce by modelling skills and service-oriented orchestration of reconfigurable manufacturing systems. at - Automatisierungstechnik 63:790-800 (2015). https://doi.org/10.1515/auto-2014-1157
8. Pfrommer J, Klein JF, Wurster M, Rapp S, Grauberger P, Lanza G, Albers A, Matthiesen S, Beyerer J. An ontology for remanufacturing systems. at - Automatisierungstechnik 70:534-541 (2022). https://doi.org/10.1515/auto-2021-0156
9. Steinbeck C, Koepler O, Herres-Pawlis S, Bach F, Jung N, Razum M, Liermann JC, Neumann S. Nfdi4chem—a research data network for international chemistry. Chemistry International 45:8-13 (2023). https://doi.org/10.1515/ci-2023-0103
10. Franceschini G, Macchietto S. Model-based design of experiments for parameter precision: state of the art. Chemical Engineering Science 63:4846-4872 (2008). https://doi.org/10.1016/j.ces.2007.11.034
11. ISO 15836-1:2017: Information and documentation — The Dublin Core metadata element set, May 2017.
12. Sherpa L, Müller-Pfefferkorn R, Tolksdorf G, Khaydarov V, Wiedau M, Urbas L. Prometas – a metadata schema for process engineering and industry. Chemie Ingenieur Technik 95:1041-1048 (2023). https://doi.org/10.1002/cite.202200225
13. Wiedau M, von Wedel L, Temmen H, Welke R, Papakonstantinou N. ENPRO data integration: extending DEXPI towards the asset lifecycle. Chemie Ingenieur Technik 91:240-255 (2019). https://doi.org/10.1002/cite.201800112
14. Hughes RC. Preparation by IEC and CIGRE for IEC publication: “IEC 60-2: 1994. high voltage test techniques part 2. measuring systems”. IEE Colloquium on NAMAS Accreditation for High Voltage Measurements and Tests 1997:1-1 (1997). https://doi.org/10.1049/ic:19970280
15. Urbas L. 2nd p2o-symposium 2025 : process-to-order. Tagungsband 2nd P2O- Symposium 2025 : Process-to-Order (P2O) Group : (2025). https://doi.org/10.25368/2025.292
16. VDI 2776-2:2024, Verfahrenstechnische Anlagen - Modulare Anlagen - Design modularer Anlagen, Richtlinie 2, Jan. 2024.
17. Maurmaier ASM. Dezentralisierte kommunikationsarchitektur für direkte modul-modul-interaktion auf basis von standardisierten konstrukten der vdi/vde/namur 2658. Automation 2019 :343-360 (2019). https://doi.org/10.51202/9783181023518-343
18. Cameron DB, Otten W, Temmen H, Hole M, Tolksdorf G. DEXPI process: standardizing interoperable information for process design and analysis. Computers & Chemical Engineering 182:108564 (2024). https://doi.org/10.1016/j.compchemeng.2023.108564