AIP Conference Proceedings

RESEARCH ARTICLE | JUNE 16 2026

Standardized generation of semantic domain models from content management metadata

Selina Sommer ✉; Maria Teichmann; Sophia Kopf; Lejla Bratic; Raphaela Erb; Wolfgang Ziegler

AIP Conf. Proc. 3432, 060005 (2026)
https://doi.org/10.1063/5.0336657

View Online Export Citation

Articles You May Be Interested In

Aligning the provisioning of product information with agile software development by semantics and correlation rules

AIP Conf. Proc. (June 2026)

Implementing cross-domain linking of service information in the aviation security sector

AIP Conf. Proc. (June 2026)

Challenges and opportunities of semantic information management and cross domain linking in the context of manufacturing

AIP Conf. Proc. (June 2026)

18 June 2026 16:39:35

AIP Publishing

Standardized Generation of Semantic Domain Models From Content Management Metadata

Selina Sommer [a)], Maria Teichmann [b)], Sophia Kopf [c)], Lejla Bratic [d)], Raphaela Erb [e)] and Wolfgang Ziegler [f)]

Faculty of Information Management & Media, Karlsruhe University of Applied Sciences, Amalienstraße 81-87, 76133 Karlsruhe, Germany

[a)] Corresponding author: sose1016@h-ka.de
[b)] tema1015@h-ka.de
[c)] koso1015@h-ka.de
[d)] brle1021@h-ka.de
[e)] erra1011@h-ka.de
[f)] wolfgang.ziegler@h-ka.de

Abstract. The digital transformation in industrial environments necessitates innovative methods for efficient processing and integration of technical information. This paper addresses the increasing demand for semantic technologies in Technical Communication (TC) to enhance interoperability and enable cross-domain knowledge utilization through the standardization of metadata formats and transformation methods. This study focuses on the development and implementation of a standardized approach for transforming XML-based metadata from Component Content Management Systems (CCMS) into ontologies and knowledge graphs, aiming for a solution applicable across different products and companies. The proposed methodology utilizes XSLT for the transformation of metadata, primarily from a CCMS used in industrial TC applications into universal ontology patterns. The research also explores the integration of metadata from other domains, such as Product Engineering involving ERP and Spare Parts Catalogs (SPC). Challenges related to heterogeneous data structures. The integration of Semantic Correlation Rules (SCR) introduced earlier in the TC domain for context-sensitive information retrieval is a key aspect of the developed approach. The project demonstrates the generation of semantic domain models, leading to improved metadata processing, enhanced interoperability, and hence the creation of a unified API facilitated by a Semantic Information Model (SIM) Core framework. The resulting merged ontology enables dynamic linking to relevant information. Future work focuses on establishing a central SIM for sustained metadata consistency and scalability, paving the way for seamless interoperability and cross-company collaboration.

18 June 2026 16:39:35

INTRODUCTION AND OBJECTIVES

The digital transformation of industrial environments is driving an increasing demand for innovative methods to efficiently process and integrate technical information. In the field of Technical Communication (TC), semantic technologies are gaining importance as they enhance interoperability between various information systems. This development necessitates the standardization of metadata formats and transformation methods to consistently process and link heterogeneous data sources.

This research examines the standardized transformation of heterogeneous metadata into a unified semantic model, enabling cross-domain knowledge utilization. Achieving this requires the consistent transformation of metadata from various systems into an ontological model that defines semantic relationships between technical entities [1].

This study focuses on the development and implementation of a standardized approach for transforming XML-based metadata from Component Content Management Systems (CCMS) into ontologies and knowledge graphs. Ontologies and knowledge graphs can play a key role in this process, as they provide a structured representation and allow for interconnection of information. The objective is to establish a standardized procedure for the metadata of the CCMS used in the project, making it applicable across different products and companies. A core aspect of this approach is the processing, transformation, and integration of metadata from various systems while accounting for domain-specific requirements of the modelling pattern. These domain models serve as the foundation for standardized metadata modelling and are supplemented by Semantic Correlation Rules (SCR) to logically link and dynamically

ETLTC2025 International Conference Series
AIP Conf. Proc. 3432, 060005-1–060005-8; https://doi.org/10.1063/5.0336657
Published under an exclusive license by AIP Publishing. 978-0-7354-5404-0/$30.00

provide specific corporate content [1]. The proposed approach aims to eliminate data silos and enable precise, context-sensitive information delivery [3].

INITIAL SITUATION

In this research project, all cooperating companies utilized the same Component Content Management System (CCMS), Cosima developed by the Docufy company (https://docufy.de/produkte/technische-dokumentation/cosima-von-docufy-das-ki-gestuetzte -xml-redaktionssystem/), as the environment for content creation and process automation.

The metadata is structured into product and information classes, distinguishing between intrinsic and extrinsic attributes, following the PI-Class method [6, 7]. Beyond the CCMS, this study also considers data from additional systems used in content creation. This includes data from Enterprise Resource Planning (ERP) systems, which label items in a bill of materials (BOM) using a component classification from master data management. The delivery of relevant information to end users is facilitated through systems such as Spare Parts Catalogs (SPC) and Content Delivery Portals (CDP), as outlined in [5].

The systems utilized for content creation enable the export of metadata, serving as the foundation for subsequent processing steps. While standardized metadata structures simplify data handling, company-specific customizations limit reusability—both within individual organizations and across corporate boundaries. To overcome the limitations of existing XML-based metadata models, semantic technologies are gaining increasing importance. The Resource Description Framework (RDF) and Ontology Web Language (OWL) provide structured and flexibly expandable metadata modeling [1]. By linking metadata from various systems and explicitly defining semantic relationships, ontologies enhance information provision through dynamic, context-sensitive retrieval of relevant content [4].

CONCEPT

A structured framework is required for modelling ontologies in order to define semantic relationships between different entities within a system. RDF and OWL syntax are used to represent the ontology. This study examines information from various data domains, i.e the Content Engineering domain and the Product Engineering domain. Both domains use different classifications to categorize product-related data. In Content Engineering, content topics are categorized according to the PI classification based on product classes and information classes [6, 7]. In Product Engineering, the categorization of physical parts follows a component classification [2].

Extracting and analyzing CCMS Metadata for Ontology Creation

Initially, the focus is placed on the Content Engineering domain. The foundation for ontology creation is the analysis of extracted metadata. The metadata in the Content Engineering domain from three companies participating in this study originate from the same source system (CCMS Cosima). Metadata configuration data were exported from the CCMS as native XML files, ensuring a homogeneous structure across all datasets. An analysis of the metadata from the CCMS revealed that all companies follow the PI-Class method. This classification approach enables organizations to efficiently manage their content within a CCMS by providing a standardized classification system. Moreover, it optimizes content creation, enhances content management processes, improves user experience, and facilitates better access to information, as outlined in [1, 6, 7].

18 June 2026 16:39:35

Standardizing Metadata Modelling with Universal Ontology Patterns

XML-based metadata serves as the starting point for transformation. The conversion process is performed using XSLT technology, which enables the translation of standardized XML formats from a Component Content Management System into an ontology.

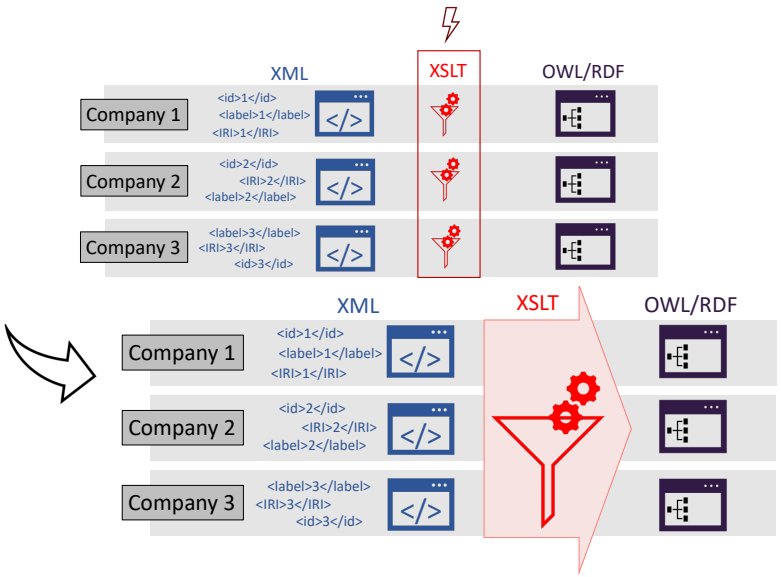

Figure 1. XML Translation into a Basic Ontology using Universal Patterns

To standardize the process of translating an XML file into an ontology, we developed a parametrized XSLT script adapted to the configuration file from the given CCMS. As we defined also a standardized model pattern for the target ontology, the XSLT transformation has to be adapted analogously to other CCMS sources.

Ontology Structure and Modelling Approaches

A suitable pattern for ontology modelling was then selected. To this end, the basic elements of the ontology were identified as a function of the metadata records and their relationships defined. This project step was supported by tools such as Protégé, which facilitate the provision and visualization of content. To ensure a consistent modelling approach, which will be explained in more detail later, the ontology is built around the following core components:

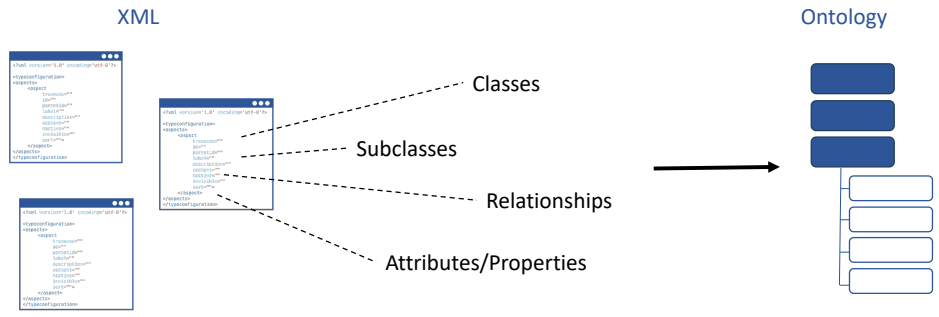

Figure 2. Basic Elements of Ontologies

- •Object Classes: Categories of instances (Figure 3: treename)
- •Instances: Specific examples within an object class (Figure 3: id)
- •Attributes/Properties: Definition and assignment of identifiers (Figure 3: label, description)
- •Relationships: Representation of links between instances (Figure 3: parentid) [1]

18 June 2026 16:39:35

```
<typeconfiguration kernelversion="10.0.14" projectversion="10.0.14">
  <aspects>
    <aspect
      treename="Diff-Komponentenkaskade"
      id="6c7b88f8cdb3e951c0a8008a3411e0d7"
      parentid="9ee265b6cdb3e94fc0a8008a3411e0d7"
      label="Fibre hopper with Discharge head"
      description="Fibre hopper with Discharge head"
      comment="Fibre hopper with Discharge head"
      inactive="false"
      invisible="false"
      sort="0">
    </aspect>
  </aspects>
</typeconfiguration>
```

Figure 3. XML Code used to identify the Basic Elements of the Ontology

Outlook: Enhancing Ontologies with Semantic Correlation Rules (SCR)

These domain models in the form of generic ontologies serve as a basis for metadata modeling. By integrating previously introduced Semantic Correlation Rules (SCR), they enable dynamic, context-sensitive information retrieval. SCR facilitates a dynamic, user-oriented linking of company-specific content by defining logical relationships between metadata tailored to the specific use cases of the company, rather than relying on direct links. A more detailed explanation of the application of SCR in the project framework can be found in [5, 8].

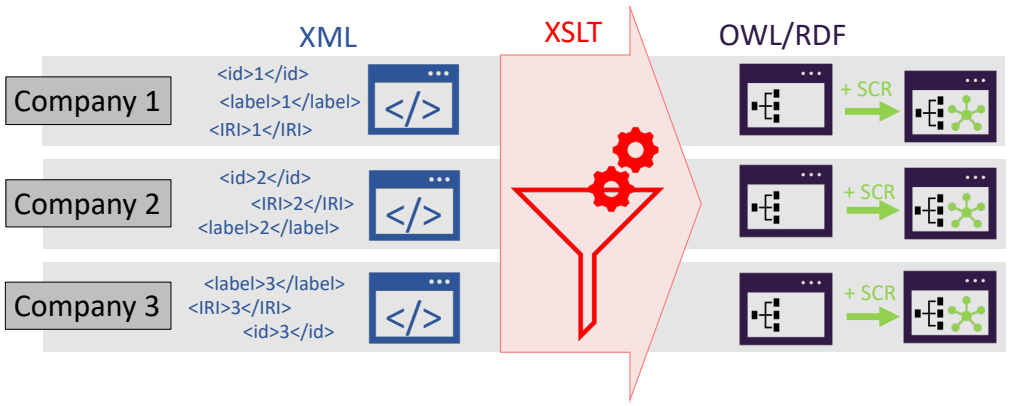

Figure 4. Enhancing Ontologies with Semantic Correlation Rules (SCR)

IMPLEMENTATION

For the transformation from XML to RDF/OWL, an XSLT script is used, consisting of the following components: attributes, classes, relationship types, and custom namespaces. The relationship properties “has_parent” and “has_child” can be used universally and are explicitly defined in the data structure of the CCMS. They specify, for example, that an instance can have one or more child instances, which corresponds to a parent instance and can also be the child of another instance.

A notable aspect of this study is the introduction of a custom namespace, which is defined individually for each company within the script. The rationale behind this approach is to maintain metadata separation when using a uniform, standardized script. This ensures clear assignment and identification of metadata when examining the constructed ontologies.

18 June 2026 16:39:35

Instance-based Modeling Pattern as a Result

When considering an ontological modelling approach, it is crucial to determine whether concepts should be represented as instances or classes. In the context of this research, an instance-based approach was chosen. Instead of defining direct relationships between classes, this modelling pattern focuses on linking instances and enables a more targeted representation of specific resource information. Attributes and relationships are assigned directly to instances. Generic instances are used to represent relationships between classes, as described in [1]. The following figures below show the pattern, its components and the XSLT script on which it is based:

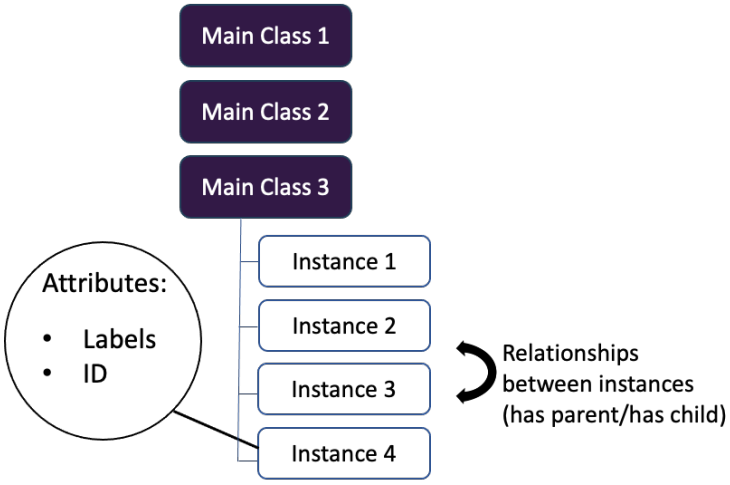

Figure 5. Instance based model pattern

```
<xsl:template match="aspect">
    <owl:NamedIndividual rdf:about="{$custom_ns}#{@id}">
        <rdf:type rdf:resource="{$custom_ns}#Class_{replace(@treename,' ','')}"/>
         <rdfs:label><xsl:value-of select="@label"/></rdfs:label>
         <xsl:if test="@parentid !='NULL'">
         <SIM:hasParent rdf:resource="{$custom_ns}#{@parentid}"/>
         </xsl:if>
        <xsl:for-each select="/typeconfiguration/aspects/aspect[@parentid = current()/@id]">
        <SIM:hasChild rdf:resource="{$custom_ns}#{@id}" />
        </xsl:for-each>
    </owl:NamedIndividual>
</xsl:template>
```

Figure 6. Extract of the XSLT Script used for the Transformation for the Instance based model pattern

Integration of Additional Domains

A key extension of the modeling approach is the integration of an additional domain - the Product Engineering domain - into semantic information management, following the same modeling pattern. This domain also includes the spare parts catalog (SPC), whose ontology is based on logical relationships between hierarchically structured products and their components.

While CCMS metadata can be exported directly to the native XML format, SPC metadata, in the case of the project, is available as an XLSX export and requires a subsequent conversion to XML format. Although the Excel file represents logical relationships between products and their components, in this case, it does not define explicit hierarchies, as illustrated in an excerpt of the XML structures.

18 June 2026 16:39:35

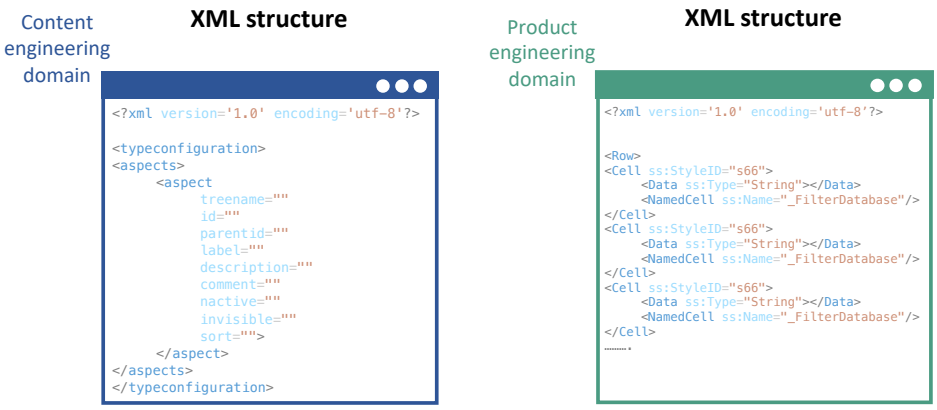

Figure 7. Comparison of the XML structure of the two domains

A particular aspect of the transformation process was handling implicit hierarchies within the Product Engineering domain. Since the export utilized an implicit hierarchy definition of parent-child relations, we adapted the transformation file accordingly. The adaptations were characterized by the fact that domain-specific requirements were added to the standardized XSLT script for CCMS metadata. The former implicit hierarchical relationships were expressed as explicit parent-child relations in the resulting ontology. To achieve this, the following script part was used:

```
<xsl:if test="$parent !='NULL'">
    <SIM:hasParent rdf:resource="{$custom_ns}#{ss:Cell[1]/ss:Data}"/>
</xsl:if> <!-- Assigned hasParent relationship, with the condition that hasParent always has a value -->

<xsl:for-each select="following-sibling::ss:Row[ss:Cell[1]/ss:Data = current()/ss:Cell[4]/ss:Data]">
    <SIM:hasChild rdf:resource="{$custom_ns}#{ss:Cell[4]/ss:Data}"/>
</xsl:for-each> <!-- Assigned hasChild relationship -->
```

Figure 8. Extract of the XSLT script used for the transformation for the Instance based model pattern

Result: Integration of Merged Ontologies for Content Delivery

In order to establish links between metadata from different data sources, an ontology was created as part of this research project that merges metadata from the Component Content Management System (CCMS) and metadata from SPC into a standardized domain ontology consisting of main classes (Content Engineering and Product Engineering) and their respective subclasses, described in detail in [3].

Technically, the merging process was carried out at the database level of the GraphDB environment used (https://www.ontotext.com/products/graphdb/). The class structure ensures a clear assignment of the metadata even after the integration of the ontologies [1].

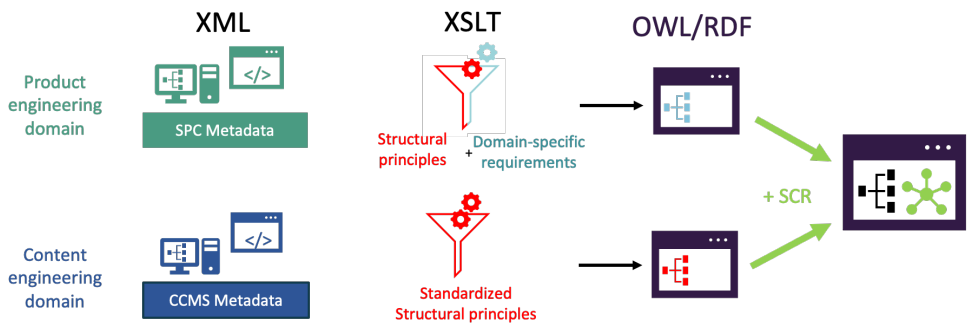

Figure 9. Visualization of the Implementation of an Standardized Ontology Merging Metadata from various domains

18 June 2026 16:39:35

Metadata and Ontology Configuration by a SIM Core

In the project we introduced the possibility for configuring the contained metadata by an additional ontology domain, called SIM Core. It serves as the framework for interpreting the metadata stemming from the underlying industry applications across different domains. This is achieved through subclass assignments, where domain-specific classes are structured based on fundamental and extended PI concepts. Additionally, the SIM Core defines the responsibilities of metadata domains and specifies the software systems that interact with the SCR-API.

Standardization within this framework involves the introduction of SIM Core classes, including domain classes for governance, environment classes for identifying system interactions, and extended PI classes [4, 6, 7] for refining domain configurations. These classes are further structured through subclass relations, enabling proper processing of equivalence relationships within the SCR-API.

The framework also supports single-domain environments, where each class is assigned to a unique domain and environment, and multi-domain environments, where classes may be associated with multiple systems utilizing the same metadata, facilitating the recognition of use case types and the generation of deep links to target systems.

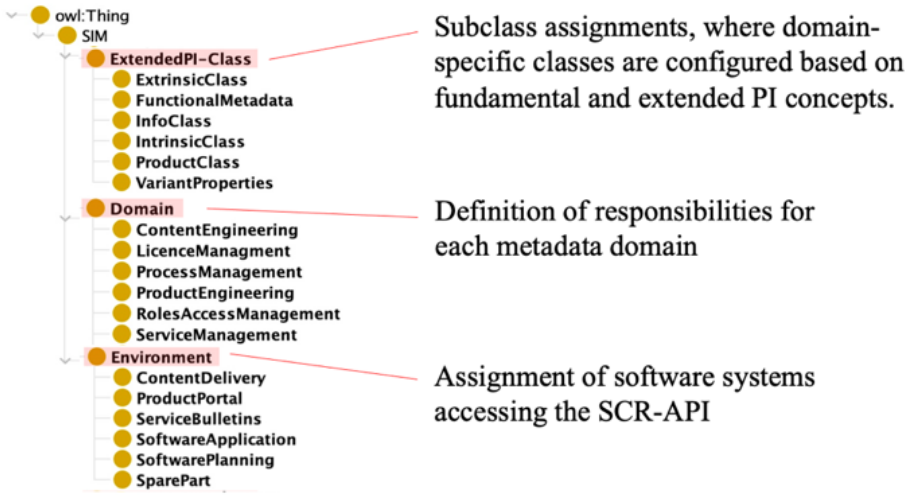

Figure 10. SIM Core Ontology

RESULTS

This project demonstrates that harmonizing metadata structures and applying a standardized approach to semantic modelling enhance data utilization and enable cross-domain integration via SCR-API access. The implemented scripting methods for ontology creation ensure consistency while providing flexibility and versioning capabilities.

The result is a standardized ontology approach suitable for knowledge graphs. In combination with Semantic Correlation Rules (SCR), it enables dynamic information delivery and cross-domain linking of relevant content through deep links in service portals. These aspects are examined in more detail in [5].

Integrating multiple OWL files into a corporate ontology and providing access via a graph database with a REST API ensures a scalable and extensible solution, allowing additional company-specific content from other domains to be incorporated. Standardized processing and semantic modelling support flexible expansion to new data sources and use cases.

A key challenge lies in maintaining metadata consistency and quality while harmonizing concepts, relationships, and properties across multiple ontologies. Quality assurance measures help detect and correct erroneous or inconsistent data. Additionally, flexible ontology designs and versioning mechanisms are essential to adapt to evolving metadata requirements.

18 June 2026 16:39:35

OUTLOOK

Based on the findings of this study, the following outlook can be derived for productive implementation in an industrial environment: The next steps should focus on establishing a central Semantic Information Model (SIM) for metadata management. This will ensure that metadata remains consistent, scalable, and of high quality in the long term. The approach developed within this project serves as a preparatory step toward implementing a centralized SIM for metadata governance. Interdisciplinary collaboration has led to the development of a standardized process and script.

Given the findings outlined in this study, it can be assumed that this approach is also applicable to other company-specific content, demonstrating both technical robustness and practical usability. The application of standardized modelling principles will ensure an efficient and adaptable solution, optimizing internal processes and enhancing the user experience.

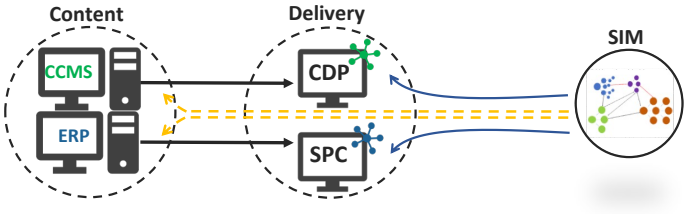

Figure 11. Representation of Future Potentials through the SIM

ACKNOWLEDGEMENTS

We would like to thank the three companies (Dieffenbacher, Smiths Detection and Siemens) for providing the basic data and acting as contact persons throughout the project. Also, we would like to thank Maurice Daum for technical service and support. And finally, we would like to thank the International Office of Karlsruhe University of Applied Sciences, the AStA and the KMM program for travel and conference support.

REFERENCES

1. L. Tempelfeld, C. Wiedewilt, A. Kruse and W. Ziegler, Transforming Domain Data into Semantic Models for Content Delivery Applications. Proceedings of the ETLTC ACM Chapter International Conference. Aizuwakamatsu, Japan. (2024)
2. F. Mueller and W. Ziegler, Overcoming Fuzziness with Semantic Modeling and AI. Proceedings of the ETLTC ACM Chapter International Conference. Aizuwakamatsu, Japan. (2024)
3. A. Kruse, C. Wiedewilt and L. Tempelfeld, Implementing Semantic Correlation Rules in the Domain of Energy Technology for Transportation Systems. Proceedings of the ETLTC ACM Chapter International Conference. Aizuwakamatsu, Japan. (2024)
4. Florian Rommel, Wolfgang Ziegler, A Microservice-based Infrastructure for the Academic Research on Semantic Knowledge Graphs and Derived Applications. SHS Web Conf. 139 02002 (2022). DOI: 10.1051/shsconf/202213902002
5. H. Paproth, S. Sommer, S. Kopf and W. Ziegler, Challenges and Opportunities of Semantic Information Management and Cross Domain Linking in the Context of Manufacturing. Proceedings of the ETLTC ACM Chapter International Conference. Aizuwakamatsu, Japan. (2025)
6. W. Ziegler, Basic Concepts Which Support the Management and Delivery of Intelligent Content, Frontier, Official Journal of Japan Technical Communicators Association JTCA, p. 84-91, (2017)
7. W. Ziegler, The Evolution of Content Management towards Intelligent Delivery Systems for Technical Communication, Frontier, Official Journal of Japan Technical Communicators Association JTCA, p. 6875, (2017)
8. W. Ziegler, Semantic Correlation Rules as a Logic Layer between Content Management and Content Delivery. Proceedings of the ETLTC ACM Chapter International Conference. Aizuwakamatsu, Japan. (2021) https://doi.org/10.1051/shsconf/202110202007

18 June 2026 16:39:35