engineering proceedings

MDPI

Proceeding Paper

Traceable Intercorporation Data Exchange and Processing Using a Graph-Based Infrastructure †

Paula Ruß [1,*], Gerald Schegk [2], Deoclécio Valente [3], Jonas Jepsen [4], Malte Christian Struck [1], Oliver Bertram [5], Frank Dressel [6] and Arthur Zamfir [4]

[1] German Aerospace Center (DLR), Institute of Software Technology, 28199 Bremen, Germany; malte.struck@dlr.de
[2] German Aerospace Center (DLR), Institute of Flight Systems, 28199 Bremen, Germany; gerald.schegk@dlr.de
[3] German Aerospace Center (DLR), Institute of Software Methods for Product Virtualization, 28199 Bremen, Germany; deoclecio.valente@dlr.de
[4] German Aerospace Center (DLR), Institute of System Architectures in Aeronautics, 21129 Hamburg, Germany; jonas.jepsen@dlr.de (J.J.); arthur.zamfir@gmail.com (A.Z.)
[5] German Aerospace Center (DLR), Institute of Flight Systems, 38108 Brunswick, Germany; oliver.bertram@dlr.de
[6] German Aerospace Center (DLR), Institute of Software Methods for Product Virtualization, 01187 Dresden, Germany; frank.dressel@dlr.de
* Correspondence: paula.russ@dlr.de
† Presented at the 15th EASN International Conference, Madrid, Spain, 14–17 October 2025.

Abstract

Designing an aircraft requires multidisciplinary analysis and data processing abilities, which are often spread over various partners. Effective collaboration across organisational boundaries is difficult, but essential. As the aerospace industry becomes increasingly digitalised, ever larger volumes of data and models must be exchanged. Heterogeneous tools, data formats, and infrastructures make it difficult to exchange data and to trace it. We propose using semantic graphs for data exchange to ensure interoperability, while semantic links between data models facilitate multidisciplinary and cross-organisational collaboration. Furthermore, our approach captures comprehensive metadata that describes the creation and modification of each dataset, thereby establishing a fully traceable data provenance chain. We demonstrate its functionality via a design process for an electromechanical actuator (EMA) given requirements from a different stakeholder (simulated). Having the requirements and the EMA models translated in Resource Description Framework (RDF) graphs, we are able to create links between them. This then enables the EMA model to be automatically re-evaluated when requirements change, ensuring that it complies with them. For the data exchange, we use the DLR SemanticHub, which utilises a graph database. By providing traceability of the data results provided in different data formats and the data origins, we enable transparency and accountability across organisational boundaries, which is important for trusted collaboration and compliance in intercorporational data exchange.

Keywords: resource description framework; data integration; provenance; requirements verification; graph database; electromechanical actuator

Academic Editors: Spiros Pantelakis, Andreas Strohmayer and Gustavo Alonso

Published: 11 June 2026

1. Introduction

Complex aircraft design processes rely on integrated digital data and collaborative workflows across multiple organisations and disciplines. However, the exchange of engineering data across corporation boundaries is not straightforward. Each contributor may

use their own set of software, such that the same information can be encoded in different data formats. Furthermore, trust management poses a challenge to data exchange, since the corporation must protect its interests and enforce selective disclosure. Traditional handovers (file-based exchanges or central repositories) demand manual effort to retrieve the right data, and the flow of data over time is rarely transparent.

Our work was conducted in the context of the *SMR Aircraft Architecture and Technology Project (SMR ACAP)*, which is part of the Clean Aviation Joint Undertaking of the European Union. Our work is connected to the objective to create "a digital collaborative framework with tools, means and skills enabling to continuously link all R[esearch]&T[echnology] activities" [1]. The ultimate goal is to facilitate data-driven collaboration across organisations and disciplines while ensuring traceability, consistency, and compliance in support of next-generation aircraft development. For our contribution, we break up this complex goal into two research questions: (RQ1) How can effective data exchange across disciplines and organisations be realised? (RQ2) How can the resulting data flows be traced in detail?

We propose a graph-based infrastructure for traceable intercorporation data exchange and processing. All data, like requirements, engineering data and model artifacts, are represented as semantic graphs using their domain-specific ontology. We test our approach using a requirements-driven development process to size an electromechanical actuator (EMA). Our test development process includes two simulated stakeholders, one that provides the requirements and one that models an actuator accordingly. We transform all data models to Resource Description Framework (RDF) graphs, with which we can semantically link the requirements to the EMA properties and automatically assess the requirements. Requirement changes trigger re-evaluation of the EMA model. For data storage and exchange, we use the DLR SemanticHub, developed at the German Aerospace Center (DLR), that utilises a graph database and can store semantical graphs, such as RDF graphs [2]. This provides a powerful method of aligning data models and linking them across model boundaries, thereby enhancing interoperability and collaboration.

Traceability of all processes is essential, where multidisciplinary collaboration spans multiple organisations. The aerospace industry aims to achieve virtual certification as an overarching goal, which entails demonstrating airworthiness primarily through trusted simulations and digital evidence, reducing reliance on extensive physical testing. This shift demands reliable and traceable data exchange, where rigorous and auditable tracking of data origins, transformations, and usages ensures compliance and builds regulatory confidence [3]. The complete provenance of creation, transformation, and usage events is captured with W3C PROV-O [4]. The remainder of this work is organised as follows: we summarise background technologies, present the proposed approach using the EMA use case, and then discuss implications for multidisciplinary collaboration and virtual certification.

2. Related Work

The challenge of enabling traceable, cross-organisational data exchange across disciplinary and organisational boundaries has motivated the development of multiple complementary approaches. These methods balance competing requirements such as interoperability, data sovereignty, transparency, and intellectual property protection.

2.1. Data Exchange Approaches

The first international standard for data exchange is STEP (ISO 10303) [5]. The exchange is completed via a neutral file approach: data is translated into a neutral format and then sent to the receiver, who translates it into the receiver's native format. Building upon STEP, MoSSEC (ISO 10303-243) defines an international standard that specifies a

https://doi.org/10.3390/engproc2026133196

semantic information model to support collaborative exchange and the traceability of modelling and simulation data. It specifies contextual metadata answering "Who, what, where, when, how, why" for traceability and structures distributed datasets through the MoSSEC Business Object Model (BOM). The BOM provides a web service that exchanges data in Extensible Markup Language (XML) [6]. Another approach is the International Data Spaces (IDS) framework. It focuses on data sovereignty in federated settings. IDS proposes a message-based exchange for data via IDS Connectors. These interpret and enforce the applied policies through a standardised interface. A shared information model bridges syntactic differences, called the IDS Information Model, which uses Resource Description Framework Schema/Web Ontology Language (OWL) ontologies [7].

2.2. Traceability Mechanisms

Ensuring traceability across organisational boundaries requires architectures that balance transparency and security. Li et al. [8] propose a layered blockchain-based cross-domain traceability scheme that records user privileges and access behaviours across multiple trust domains. The NIST Supply Chain Traceability Meta-Framework offers a technology-neutral model defining event-based recording and cryptographically linked traceability chains [9]. Provenance graphs further enhance traceability by capturing complete causal relationships and transformation histories, supporting reproducibility through fine-grained logs, and aligning with the W3C PROV standard and its ontology PROV-O [4,10]. Their application in aerospace demonstrates regulatory potential, as shown by Ruß et al. [3], who note that provenance-based data flow tracking aligns with virtual certification requirements of aviation authorities.

2.3. Positioning of This Work

Our contribution combines RDF-based, ontology-driven exchange across corporate boundaries and PROV-O provenance. The data models are stored in a shared graph infrastructure (DLR SemanticHub), which enables the creation of links between the models. We demonstrate its use based on a comprehensible design process involving automatic re-evaluation and analysis in response to changes in requirements, using an EMA design use case. Our process provides traceability suitable for analysis and certification evidence.

3. Technical Framework

3.1. Resource Description Framework

The Resource Description Framework (RDF) is a semantic web standard developed by W3C and intended for the interchange of data on the World Wide Web. In the RDF, data resources are linked using statements in the form of subject, predicate and object triplets. There are two types of elements: resources and literals. While resources, which are described using Internationalized Resource Identifiers (IRIs), can be used as subject, predicate and object, literals can only be used as objects and are described by a lexical form and a data type IRI. With this simple syntax, complex graphs can be built containing and linking all kinds of data.

3.2. DLR SemanticHub

The DLR SemanticHub (formerly Codex-Platform), developed at the DLR Institute for System Architectures in Aeronautics, supports the creation, exploration, inference, and collaborative curation of RDF models via a graphical user interface and a REST API [2]. It is used to exchange RDF data between processes and to link data from different domains and sources. Resources may denote native RDF assets or external artifacts (e.g., files) referenced by stable IRIs (such as file locations combined with in-file identifiers). The

platform manages access control for models and files, enabling cross-organisation sharing while maintaining governance.

3.3. Provenance

Provenance records the lineage of data from creation through subsequent transformations and uses. We adopt the W3C PROV family, using its OWL serialisation (PROV-O) [4]. PROV distinguishes *entities* (states of things), *activities* (processes/events that use or produce entities), and *agents* (responsible parties). Core relations include `prov:used`, `prov:wasGeneratedBy`, `prov:wasDerivedFrom`, `prov:wasAssociatedWith`, and `prov:wasAttributedTo`. These create provenance graphs, which are acyclic, as causality in derivation imposes a temporal ordering. Our provenance queries rely on these causal constraints to support auditability and reproducibility [3].

4. Application and Methodology

4.1. Intercorporation Data Exchange

For designing an aircraft, many disciplines and parties need to collaborate, but each discipline and organisation works with a different software stack and data formats. This can sometimes make the exchange of data difficult.

Our use case is a requirements-based design of an EMA for the control surface of an aircraft. The EMA is designed by a stakeholder which receives a set of requirements from the Requirements Manager. All data exchange between the aircraft designer and the EMA designer is done through the DLR SemanticHub [2]. Figure 1 illustrates the approach for intercorporation data exchange via the DLR SemanticHub applied to our use case. It shows both stakeholders: the Requirements Manager on the left and the EMA Designer on the right-hand side. At the beginning of the use case, the Requirements Manager creates the requirements using the Cameo Systems Modeler, a Model-Based Systems Engineering (MBSE) environment. During the creation of the requirements, the use of conditioning terms, such as “lower than” or “equal to”, for requirements postulating concrete values leads to an additional entry in the exported Requirements Interchange Format (ReqIF) file. This additional entry contains the comparator and the numeric value without unit. As a consequence, extracting values for automated requirement verification does not require parsing the full requirement text. However, it is necessary to either agree on a predefined set of units or to ensure unit consistency during requirement allocation. Missing unit alignment can lead to a delay in the development process or, in the worst case, to a catastrophic event after product launch. To mitigate this, organisations can agree on a specification defining all units in use, and additionally cross-check models and documents. ReqIF is defined by an XML Schema Definition (XSD) file and widely used in industry. The implementation of a converter allows for translating of XML data into the RDF/OWL by considering the corresponding XSD, which ultimately converts the requirements into the RDF. The created requirements are uploaded to the DLR SemanticHub in an RDF format, so the EMA Designer can fetch the requirements. The EMA Designer then generates the design of the EMA described as follows, thus enabling automated requirements verification in a later step. The EMA model is designed in SysML v2. This SysML v2 model is then converted into the Ontological Modeling Language (OML) using an OML vocabulary for SysML v2 [11,12]. The correctness of this conversion is verified by the OML reasoner. The OML representation is subsequently translated into OWL and exported as a turtle file. The following mapping of the created EMA model to the provided requirements is executed by the EMA Designer to allow for automated requirements verification, as described in Section 4.2. Each requirement and SysML v2 element carries a unique identifier (ID), allowing unambiguous links to be established. The mapping

must connect each requirement with all model elements it defines or constrains. This mapping is essential for automated execution, as it provides explicit information on which requirements must be compared with which model elements. Without it, this information would have to be extracted from the requirement text, making the process more error-prone. The current implementation does not include an automatic addition/removal of new/obsolete requirements and model elements to the mapping. To perform such automatic actions, an algorithm requires a knowledge base, e.g., in the form of a knowledge graph, in both the requirement and EMA domain to first understand the requirement and, second, find the correct counterpart in the EMA model. This lack of automation does not have a major impact on the scalability of this method, because the addition or removal of requirements coincides with a necessary review of the model. This review is conducted by the EMA Designer and allows for the mapping of both the requirements and the model in the same step. The RDF data generated by the automated requirements verification is then provided to the provenance wrapper (see Section 4.3). The provenance wrapper enriches the provided RDF data with recorded provenance data stored in a provenance graph, ensuring traceability. These graphs are then uploaded to separate models on the DLR SemanticHub. The Requirements Manager can then retrieve the information from the DLR SemanticHub to check whether the requirements are fulfilled or not. If the requirements change, the whole process is started again.

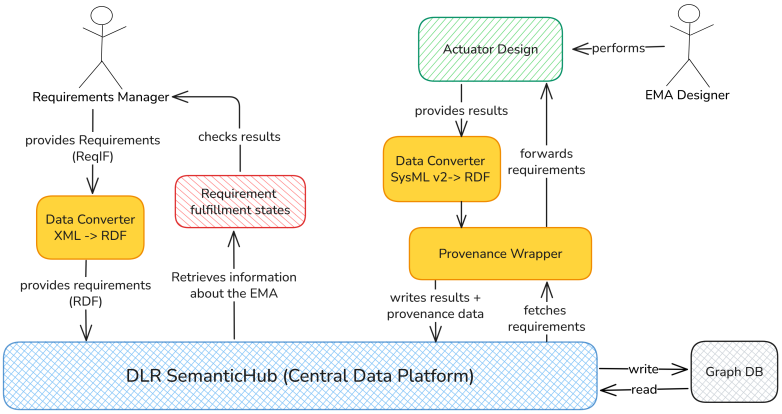

Figure 1. Illustration of our test design process for the EMA designing use case.

4.2. Requirement Verification and Model Selection

This section describes the requirement verification process mentioned in Section 4.1. The goal of this process is to adapt a model to requirements and verify it in an automated manner. This process needs three prerequisites for automated execution: requirements in an RDF format, a model in an RDF format, and a map connecting requirements and model. The process itself will be described using the use case example, a design of an EMA. The output of this process is a model linked to requirements with information describing the verification status. The whole process is shown in Figure 2.

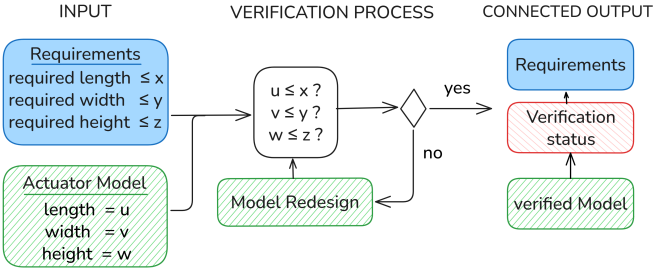

Figure 2. Automated model adaption and requirements verification process.

Four requirements are considered: three define maximum values for the actuator's overall length, width, and height, while a fourth states that the actuator must fit into a given compartment. The first three requirements are of a comparative type, since they define maximum values for specific properties. The fourth requirement, however, is non-comparative, because it does not specify a numerical limit. Instead, it requires a dedicated verification rule, as no direct value comparison can be applied. The setup of the process requires the prerequisites described earlier. Based on the requirements, the EMA model must provide an architecture, functions, and properties that reflect the constraints, including explicit properties for overall length, width, and height. The mapping then links requirement IDs with the IDs of the corresponding model elements in the RDF. For example, the maximum length requirement is mapped to the model property for overall length. Equivalent mappings exist for width and height. The fourth requirement, demanding that the EMA fits into the compartment, is mapped to the entire EMA part since this requirement applies to the whole actuator and no single property provides a direct correspondence.

With these inputs in place, the verification process begins by importing both the model and requirements into a database, as seen on the left-hand side in Figure 2. Within this environment, the requirements and model elements are connected through the predefined mapping. This integration makes it possible to identify the set of requirements defining the EMA model within a larger knowledge graph and enables targeted queries. The actual verification is carried out in an iterative loop that runs until either all requirements are fulfilled or a termination condition is met, as seen in the centre of Figure 2. Termination may occur if a defined maximum number of iterations is exceeded or if certain requirements are classified as unsatisfiable. In each iteration, all requirements containing a comparator and a value are queried, and the corresponding model properties are retrieved. These values are then compared according to the mapping and the comparator specified by each requirement. Depending on the outcome, each requirement is marked as fulfilled or not fulfilled. Non-comparative requirements, such as the "fit into compartment" condition, are checked separately by means of predefined rules. In this specific case, the rule states that the requirement is satisfied if all three comparative requirements for length, width, and height are fulfilled. If all requirements are marked as fulfilled, the loop ends successfully. Otherwise, the EMA model is adapted, and the loop starts again, continuing until either compliance is achieved or a break condition is triggered. The iterative verification loop is followed by a compliance classification step. Each model element is annotated with its compliance status relative to the linked requirements. Elements satisfying all requirements are assigned to a class of compliant elements, while elements violating one or more requirements are assigned to a class of non-compliant elements. This systematic classification enhances traceability by allowing for direct identification of the parts in the model that meet the requirements and those which require further adaptation.

The result of the process is an adapted EMA model enriched with explicit compliance information. Both the verified model and the verification data are exported and made available within a central database, ensuring consistency and providing a reliable foundation for subsequent engineering processes.

4.3. Traceability

Building on data exchange via the DLR SemanticHub and automated requirements verification (Sections 4.1 and 4.2), we tackle RQ2 and explain how we capture the data lineage. Certification of model-based systems depends on *defensible* provenance: what was produced, by whom, with which configuration, and how it evolved [3]. To address these aspects, we track the data flow and generate provenance data from the outputs

of the EMA design process. This work captures provenance from the EMA designer's perspective because the Requirements Manager acts as an external stakeholder, and they might not want to expose internal processes to other partners. As a result, the provenance record fully describes the EMA-side workflow, but it describes the requirement authoring process only partially. Future work could involve extending the provenance chain across organisations by connecting stakeholder-specific provenance exports through shared IDs, as described in [3]. When the requirements are fetched from the DLR SemanticHub, we record the retrieval event and associated metadata in the provenance graph. We track the data flow at tool level, capturing the input and output produced by the EMA designing tool. The workflow also records the subsequent upload of the derived artifacts to the DLR SemanticHub and transforms it into a provenance graph that represents higher-level workflow.

To trace the provenance of the EMA design process, we created a single tracing workflow that turns heterogeneous engineering data into a consistent evidence record. This is shown in Figure 3. First, we use the aligned inputs: requirements (ReqIF) and design models (SysML v2) are normalised to RDF graphs, providing a governed starting point for traceability. By adding some metadata about, e.g., the responsibilities and the tool version, we can enrich these graphs into a PROV-O evidence model [4]. Key items (requirements, models, files) become *entities*, engineering steps are *activities*, and tools or people are *agents*. We use qualified associations/roles where appropriate, and link each resulting artifact to its producing activity and responsible agents. The core result is the provenance graph created in PROV-O ontology. Further, we obtain a visualisation of the provenance graph, as well as traceability matrices between the mapped requirements and the model parts. If provenance graphs of a prior run are given, then we manage changes via semantic differencing rather than raw text comparison: volatile fields are ignored, and we report additions, removals, and updates. At the end, we obtain difference summaries with respect to the previous run, as well as a visualisation of the differences in the provenance graphs.

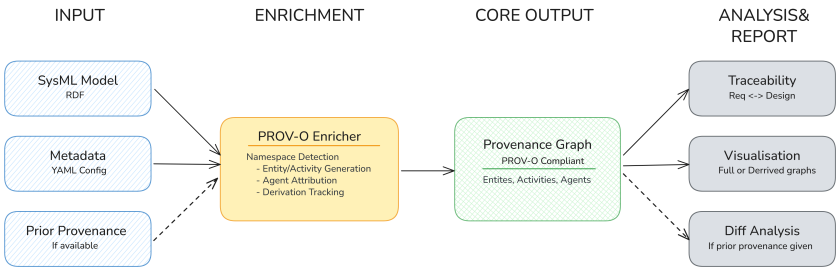

Figure 3. Workflow from SysML v2 to a PROV-O evidence graph for the actuator design process.

In sum, providing the coarser data flow provenance graph alongside the graph generated from the actuator design tool's outputs in machine- and human-readable formats provides traceability without the need to adapt the EMA design process. This ensures an auditable lineage suitable for lifecycle and regulatory use.

5. Conclusions

Our proof-of-concept highlights the benefits of using semantic graphs to facilitate collaboration between different stakeholders in a collaborative design system. The employed industrial data exchange standards are transformed into graphs without loss of information. Through the creation of semantic links between requirements model graphs and actuator model graphs in our central platform DLR SemanticHub, the automatic verification of actuator requirements is enabled. Crucial to transparent collaboration, our approach systematically captures and manages provenance information within the data flow established

https://doi.org/10.3390/engproc2026133196

during the actuator model selection and verification lifecycle. By being non-intrusive, i.e., not requiring the adaptation of specific tools and software, implementing the traceability mechanism is kept straightforward and has a low implementation threshold. Intercorporational data exchange and processing facilitated by interlinking semantic graphs, as well as significantly increasing the transparency of the process through the provision of provenance information, has great potential to facilitate and accelerate aircraft engineering. It might prove to be key to enabling the virtual certification of the next generation of aircraft.

Limitations and Future Work

Our current work revolves around a single but representative system and its stakeholders. Therefore, validity and practicability for larger, more complex design processes remain to be shown. Capturing provenance introduces overhead, and PROV-O/RDF graphs can grow rapidly. Limiting granularity, storing provenance separately, and using efficient multi-graph queries help contain this, though scalability remains critical for industrial use. Furthermore, the integration of this process into existing MBSE/PLM tool chains needs to be evaluated.

Author Contributions: Conceptualisation, P.R., G.S., D.V., J.J. and M.C.S.; methodology, P.R., G.S., D.V., J.J., M.C.S. and A.Z.; software, P.R., G.S., D.V., J.J., M.C.S. and A.Z.; investigation, P.R., G.S., D.V. and J.J.; writing—original draft preparation, P.R., G.S., D.V., J.J. and O.B.; writing—review and editing, P.R., G.S., D.V., J.J., M.C.S., O.B. and F.D.; visualisation, P.R., G.S., D.V. and J.J.; supervision, M.C.S. and F.D.; project administration, P.R. and M.C.S. All authors have read and agreed to the published version of the manuscript.

Funding: Co-funded by the European Union under grant agreement number 101101955. The views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the Clean Aviation Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data belongs to the SMR ACAP project under grant agreement no. 101101955 and cannot be shared without explicit consent of the consortium.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. European Commission. SMR Aircraft Architecture and Technology Integration Project [Grant agreement ID: 101101955]. Available online: https://cordis.europa.eu/project/id/101101955 (accessed on 15 February 2026). [CrossRef]
2. Jepsen, J.; Zamfir, A.; Boden, B.; Zamboni, J.; Moerland, E. On the Development of a Collaborative Knowledge Platform for Engineering Sciences. In Proceedings of the 15th International Conference on Knowledge Engineering and Ontology Development, KEOD 2023, Rome, Italy, 13–15 November 2023; pp. 208–215.
3. Ruß, P.N.; Schreiber, A.; Struck, M.C.; Weinert, A. Supporting Virtual Aircraft Certification via Provenance. In Proceedings of the ProvenanceWeek 2025, New York, NY, USA, 27 June 2025; pp. 85–90. [CrossRef]
4. Lebo, T.; Sahoo, S.; McGuinness, D.L.; Belhajjame, K.; Cheney, J.; Corsar, D.; Garijo, D.; Soiland-Reyes, S.; Zednik, S.; Zhao, J. PROV-O: The PROV Ontology. In *W3C Recommendation*; World Wide Web Consortium: Cambridge, MA, USA, 2013.
5. Pratt, M.J. Introduction to ISO 10303—The STEP Standard for Product Data Exchange. *J. Comput. Inf. Sci. Eng.* **2001**, *1*, 102–103. [CrossRef]
6. *ISO 10303-243:2021*; ISO Central Secretary. Application Protocol: For Modelling and Simulation Information in a Collaborative Systems Engineering Context (MoSSEC). International Organization for Standardization: Geneva, Switzerland, 2021.

7. Bader, S.; Pullmann, J.; Mader, C.; Tramp, S.; Quix, C.; Müller, A.W.; Akyürek, H.; Böckmann, M.; Imbusch, B.T.; Lipp, J.; et al. The International Data Spaces Information Model—An Ontology for Sovereign Exchange of Digital Content. In *Proceedings of the Semantic Web—ISWC 2020*; Pan, J.Z., Tamma, V., d'Amato, C., Janowicz, K., Fu, B., Polleres, A., Seneviratne, O., Kagal, L., Eds.; Springer: Cham, Switzerland, 2020; pp. 176–192.
8. Li, J.; Cao, L.; Wang, X. Cross-domain Joint Traceability Scheme Based on Layered Blockchain. In *Proceedings of the 2025 4th International Conference on Bigdata Blockchain and Economy Management (ICBBEM 2025)*; Atlantis Press: Dordrecht, The Netherlands, 2025; pp. 335–355. [CrossRef]
9. Pease, M.; Wallace, E.; Reed, H.; Martin, V.; Granata, S. *Supply Chain Traceability: Manufacturing Meta-Framework*; Technical Report; National Institute of Standards and Technology: Gaithersburg, MD, USA, 2025. [CrossRef]
10. W3C. PROV-Overview. W3C Note. 2013. Available online: https://www.w3.org/TR/prov-overview/ (accessed on 15 February 2026).
11. Opencaesar. An OML Representation of the SysML2 Vocabulary from OMG. Available online: https://github.com/opencaesar/sandbox/tree/master/sysml2 (accessed on 15 February 2026).
12. Bertram, O. Exploring the Synergy of SysML v2 and OML in Model-Based Aircraft Systems Design. In Proceedings of the 2nd International Forum on Ontological Modeling and Analysis, La Cañada Flintridge, CA, USA, 5–6 May 2025.