Alliance for AI, IoT and Edge Continuum Innovation

Report for AIOTI Workshop on Semantic Interoperability for Digital Twins

Release 1.0

AIOTI WG Standardisation
Focus Group on Semantic Interoperability

June 2025

Report for AIOTI Workshop on Semantic Interoperability for Digital Twins

5-6 February, hosted by ERCIM in Sophia Antipolis, France

With thanks to support from INRIA, and the Nephele and SmartEdge projects, funded by the European Commission under the Horizon Europe programme. See also the workshop website[1].

[1] https://www.ercim.eu/events/aioti-workshop-on-semantic-interoperability-for-digital-twins

Executive Summary

The workshop was aimed at advancing the field of semantic interoperability and laying the groundwork for the next generation of information systems, with an emphasis on practical solutions for real-world industrial challenges. The workshop was designed as a face to face event to encourage discussion, together with some remotely presented keynotes where the programme committee felt that was appropriate.

The workshop focused on solutions for industry along with encouraging greater take-up of semantic technologies, including controlled vocabularies, taxonomies, and ontologies. Semantic interoperability is essential for supply chains as well as for repairs and recycling in relation to the plans for the circular economy as part of the EU's Green Deal.

Semantic Interoperability is the ability of computer systems to exchange data with a shared, unambiguous meaning. Achieving this is critical to opening up markets for services and data, avoiding fragmented data silos, and enabling seamless cross-domain data infrastructures. Special consideration was given to the role of Digital Twins, which represent virtual counterparts to real-world entities and provide additional capabilities like prediction, simulation, and automation.

Group photo taken in the sunshine on the first day

Table of Contents

Day 1

Session 1: Introduction

Introduction by the Host [slides]

Dominique Hazaêl-Massieux, ERCIM General Manager, welcomed the participants as host. He gave an introduction of the work of ERCIM and the support provided to W3C. This includes catering to the European system with respect to web standards, hosting W3C chapters, and participating in research projects.

Dave Raggett outlined the goals of the workshop, which are to advance the field of semantic interoperability, find practical solutions for real-world industrial challenges, and lay the groundwork for the next generation of information systems. The presentation also provides an overview of the topics that have been covered in the workshop, such as the state of the art in semantic interoperability, success stories and best practices, challenges and new technologies, and a roadmap for future development.

The workshop was organized into two days. The first day focused on the state of the art in semantic interoperability, success stories and best practices, and challenges and new technologies. The second day focused on a roadmap for future development. The workshop included a mix of invited talks, breakout sessions, and community-submitted presentations. The presentation concluded with a call to action for the audience to get involved in the workshop and help to advance the field of semantic interoperability.

Overview of AIOTI Semantic Interoperability Group and its activities, Martin Bauer [slides]

Martin Bauer introduced AIOTI, the AIOTI Semantic Interoperability Group and its activities. The AIOTI Semantic Interoperability Group has bi-weekly calls. [Contact: martin.bauer@neclab.eu if you are interested in participation. It is generally open to AIOTI members, but some exceptions for semantic experts are possible,]

Semantics in this context is about explicit agreement on meaning. The activity started out with a whitepaper on Semantic Interoperability for the Web of Things in 2016. Two more white papers have been published in 2019, one focused on standardization "Towards Semantic Interoperability Standards based on Ontologies", the other targeting developers "Semantic IoT Solutions: A Developer Perspective". A Semantic Interoperability tutorial was presented at IoT week in 2021. Current activities include the ontology landscape and a white paper on data to ontology mapping.

AIOTI Ontology Landscape, Martin Bauer [slides]

Martin Bauer presented the AIOTI Ontology Landscape. https://aiotieu.github.io/ontologylandscape/ Its goal is to encourage ontology re-use by providing an overview of relevant IoT ontologies. For the ontologies, the quality and actual usage is estimated by giving a technology readiness level (TRL) in one dimension (from TRL 4), the other dimension is the maintenance of the ontology, i.e. whether it is maintained by a single maintainer or project, an organization, an industry group or a standardization body. Ontologies are structured according to vertical IoT domains (home/building, industry, mobility, health, energy, cities, wearables, farming/agrifood, water/environment). In addition, there are vertical IoT ontologies. The idea is to further extend the ontology landscape with upper layer ontologies often used by IoT ontologies in the next version.

- *The goal is to extend the ontology landscape, for this purpose we are looking for relevant ontologies. This includes all publicly accessible ontologies fitting the above criteria, e.g. W3C ontologies.*
  Please support us in expanding the AIOTI ontology landscape!
- *We will expand beyond full formal ontologies to controlled vocabularies and taxonomies.*
- *It was suggested to link to information in the LOV / LOV4IoT ontology catalogues https://lov4iot.appspot.com/?p=ontologies, which may generally be more open with respect to which ontologies are included.*
- *It was suggested to learn from the health community. Under top level ontologies they have sub ontologies, e.g. NCI for cancer. Organization of ontologies is key*

Data to Ontology Mapping Whitepaper, Cornelis Bouter [slides]

Cornelis Bouter introduced the work on the Data to Ontology Mapping Whitepaper that is soon going to be published by AIOTI.

As part of a previous survey, we had asked the community about the most important aspects that needed to be improved for increasing the uptake of semantic technologies. The most important aspect identified in this survey was the mapping of data to ontologies.

Thus, we decided to write a whitepaper that gives an overview of existing mapping tools supporting the mapping of data to ontologies. We identified three main categories of tools: graphical user interfaces, configuration languages and programming language based tools. We compared them with respect to expressivity, accessibility/usability, scalability and maintainability. The goal is to help the user to find the most suitable tool for a given task in a given environment.

The presentation discussed the importance of data to ontology mapping for semantic interoperability. Data to ontology mapping is the process of creating a mapping between a data source and an ontology. This mapping can then be used to transform data from the data source into the ontology. The presentation also discussed a number of different approaches to data to ontology mapping, such as graphical user interfaces, configuration languages, and programming languages. The presentation highlighted the benefits of using data to ontology mapping, such as improved data sharing, increased innovation, and reduced costs. The presentation also discussed a number of challenges that need to be addressed in order to effectively use data to ontology mapping, such as the need for high-quality ontologies and the need for tools to support data to ontology mapping. The presentation concluded with a discussion of the future of data to ontology mapping. The presentation argued that data to ontology mapping has the potential to transform the way that data is shared and used.

The presentation further discussed a number of specific tools that can be used for data to ontology mapping. These tools include OntoRefine, Mapping Workbench, Semantic Treehouse, Linked Data Wizard, RML, YARRRML, R2RML, SPARQL Anything, Stardog Mapping Syntax, OnTop, Apache Jena, Carml, RDFLib, and dotNetRDF. The presentation highlighted the benefits of using these tools, such as improved accuracy, reduced costs, and increased speed. The presentation discussed a number of challenges that need to be addressed in order to use these tools effectively, such as the need for high-quality training data and the need to ensure that the generated mappings are accurate and consistent. The presentation concluded with a discussion of the future of data to ontology mapping tools. The presentation argued that these tools have the potential to make it easier to create, access, share, and reuse data.

- ***Please help us expand the survey if you know of additional tools.***
- *LinkML looks very promising.*
- *Beyond mapping data to ontology concepts, alignment between different ontologies, i.e. the mapping between terms from different ontologies, is an important topic to be addressed. This clearly goes beyond the scope of this whitepaper, but could be addressed as part of a new activity.*
- ***How could LLM's be used to support the mappings?*** *This is a relevant topic and also relates to the first Session on Day 2. Henon has promised a contribution to the whitepaper on this topic.*
- *What is the semantic representation behind the mapping languages?*
- *Semantic artefacts. This may require humans in the loop to provide additional judgement.*
- *The notion of fidelity. The level of precision and whether tight or loose concepts are appropriate.*
- *Datasets vary, making it hard to fit them to existing ontologies.*
- *How extensible are ontologies? Is it easy to extend an existing ontology to match your specific needs? SAREF as a case in point.*

Key Points from Martin's Overview and Survey:

- **Gap Analysis:** The discussion stemmed from a gap analysis exercise.
- **Survey Results:** A survey indicated that **data mapping** was the most significant issue in semantic interoperability.
- **Recommendation:** The need for a **proper graphical data mapping tool** was highlighted.
- **Criteria for the Tool:** Maintainability and requiring only limited knowledge of semantics were mentioned as important criteria.
- **Mapping Language Options:** Configuration languages, particularly RML, were discussed as options, but they require comfort with coding.
- **Trade-offs in Tooling:** The discussion sought to identify missing tools and the trade-offs between different approaches.

Vladimir Alexiev's Contributions and Concerns:

- **Missing Major Ontologies:** Vladimir pointed out the absence of major "professional" ontologies like IFC (AEC) and CIM/CGMES (electricity) in the survey's scope, suggesting that a survey might not be the most comprehensive way to identify needs in a global data landscape.
- **LOV4IoT:** Vladimir highlighted the existence of LOV4IoT (https://lov4iot.appspot.com/) with 879 vocabularies and suggested merging efforts. (Amélie later confirmed contact with Martin regarding LOV4IoT).
- **TechnoPortal.hevs.ch:** Vladimir shared the link to TechnoPortal (https://technoportal.hevs.ch), describing it as similar to the OBO portal but for industrial/engineering ontologies.
- **Bad RDF Modelling Practices:** Vladimir identified poor RDF modelling practices in ECLASS and multiple RDF shortcomings in AAS, a key point in the RDF vs. AAS discussion.
- **Integration of Semantic and Time-Series Databases:** Vladimir raised the topic of "hybrid storage", i.e. integrating semantic data with specialised databases, such as time-series databases.

Other Participants' Contributions and Perspectives:

- **Life Sciences Needs:** Someone from the life sciences domain mentioned their specific needs and their work on LinkML, which provides scripting capabilities for data mapping.
- **Ontology Alignment vs. Data to Term Mapping (PhilA & CB):** Clarification was sought on whether the focus was on mapping data to terms or aligning between two terms, highlighting the need for different tools for each.
- **LLMs for Data Mapping (CB):** The role of LLMs in data mapping was acknowledged as a topic for future discussion, with questions around accuracy and usability.
- **Semantics Behind Languages (Meyer):** A question was raised about whether the semantics behind data modelling languages themselves are being improved.
- **Terminology Alignment (AK & CB):** The discussion touched upon the lack of aligned terminology (e.g., "fidelity") in ontology mapping and the difficulty of quantifying such aspects due to their social nature.
- **Ontology Fit and Extensibility (Antonio & CB):** The issue of existing ontologies not perfectly fitting real-world data was discussed, emphasizing the importance of reusability in parts and the extensibility of ontologies (SAREF was mentioned as a good example with its layered "onion" structure).

Overall Themes:

- **Data Mapping as a Primary Bottleneck:** The survey results and subsequent discussion consistently pointed to data mapping as a major challenge hindering semantic interoperability.
- **Importance of Existing Resources:** The value of existing ontology repositories like LOV4IoT and TechnoPortal was highlighted, suggesting the need for better awareness and potential collaboration.
- **Quality of Semantic Models:** Concerns were raised about the quality of existing RDF models in industrial standards.
- **Potential of New Approaches:** Tools like LinkML and the potential role of LLMs were mentioned as areas of ongoing research and potential solutions.
- **Need for Practical Tooling:** The strong recommendation for a user-friendly graphical data mapping tool underscored the need for more accessible solutions.

Next Step Mentioned:

- **Integrating semantic and time-series databases** was identified as a future direction to explore.

Session 2: Semantic Interoperability for Digital Twins

Keynote: Asset Administration Shell (AAS), ECLASS and SAAM (Vladimir Alexiev, Ken Wenzel, Hossein Rimaz, Fraunhofer IWU) [slides] [paper]

The presentation describes challenges in the RDF representation of Asset Administration Shells (AAS), which are digital representations of physical assets.

In addition, it addressed challenges in the semantic representation of ECLASS, a product classification and data dictionary of product properties. ECLASS is used in AAS to provide shared "semantics" of product properties, but there are challenges with this approach. One challenge is that ECLASS properties are represented in AAS by copying their IRDI and name, which involves a lot of intermediate blank nodes and redundant information. Another challenge is that ECLASS class and property URLs are unstable, which makes it difficult to reuse external Linked Open Data (LOD).

The presentation also discussed the need for better Machine-Readability of AAS schemas (Submodels). Early AAS schemas were PDFs; now submodels are defined using AAS Submodel Templates. However, these templates do not constitute a strict schema with defined property cardinalities, ranges, etc. The Eclipse Semantic Modeling Framework (ESMF) and the Semantic Aspect Meta Model (SAMM) try to fill this gap. ESMF is a tool for generating machine-readable artifacts from semantic models, and SAMM is a meta-model for creating digital twin models.

The presentation argued that SAMM does not follow long-established semantic modelling standards, such as RDFS and OWL (classes, properties, subclasses, subproperties, domains, ranges, datatypes, etc.). SAMM uses RDF and SHACL to define IoT models, but semantics is used only as a means and not as an end-goal.

Here's a summary of the main points from the transcription of the talk:

Asset Administration Shell (AAS):

- Aims to provide interoperability for industrial digital twins
- Standardized as IEC 663278

- Widely adopted in Germany and expected to have a significant future, especially with initiatives like Manufacturing-X and digital product passports
- Uses submodels to contain data, with templates and instances
- Relies on external dictionaries like ECLASS and IEC CDD for semantics
- Supports various serialization formats (JSON, XML, RDF)
- AAS as RDF would be very useful for complex queries using SPARQL, but few people use it
- Facing challenges with RDF representation, including namespace versioning, ordering issues, datatype and language code handling, blank node usage (lack of URLs)

ECLASS:

- A classification and data dictionary for industrial products
- Based on the OntoML ISO 13584-32 standard
- Used for providing semantics in AAS and other industrial standards like OPC-UA
- Available in 29 languages
- Structured in four levels with classes and properties
- Has an RDF representation, but it faces challenges:
    - Unstable URLs due to version numbers: both the global ECLASS version, and change counter in IRDI identifiers
    - Non-standard use of data types and language strings
    - Complex property grouping using blocks and aspects, hindering direct querying.
    - Problems with unit of measure definitions
- The RDF representation is not following decades of experience with RDF best practices

Semantic Representation Challenges:

- The talk highlights the tension between established industrial standards and semantic web best practices.
- There's a need for better standardization and sharing of solutions for syntactic and management aspects of semantic data.
- The practical use of AAS and ECLASS reveals that even mature industrial standards require fixes in their semantic representations.
- The difficulty of querying the data because of the way that the RDF has been constructed.

QUDT is an ontology that describes units of measure and their relationships.

IRDI are used by IEC CDD and ECLASS as an identifier, and these identifiers are not resolvable.

Keynote: Asset Interfaces Description for WoT as part of the AAS, and relation to ECLASS and OPC UA (Sebastian Käbisch, Siemens)

(awaiting slides)

W3C WoT in a nutshell. Decoupling application code from the complexity of the underlying IoT protocols and data formats. The application API for things modelled in terms of properties, actions and events. Example Thing Description in JSON-LD along with reference to the W3C Recommendation for Thing Descriptions.

AAS basics. No standardised way as yet for AAS interchange. Work in progress in the Industrial Digital Twin Consortium. AAS defined by its submodels.

Asset Interface descriptions (AID) is one of those submodels that uses W3C WoT.

This talk focuses on the intersection of the Web of Things (WoT) and Asset Administration Shell (AAS), specifically regarding the "Asset Interface Description" (AID) submodel. Here's a summary of the main points from the transcription of the talk:

- **Web of Things (WoT):**
    - Aims to standardize IoT application development by addressing the "mess" of diverse protocols and data models.
    - Provides building blocks like Thing Descriptions (TDs) that act as "landing pages" for IoT devices, describing their interfaces.
    - Uses a property, action, and event architecture to abstract away protocol-specific details.
    - TDs are JSON-LD based, allowing for semantic annotations using existing ontologies.
- **Asset Administration Shell (AAS):**
    - A generic approach to standardize information exchange between stakeholders in industrial contexts.
    - Organizes asset information into submodels, each with a specific context (e.g., nameplate, documentation).
    - Submodels can leverage existing standards like ECLASS.
    - The AID submodel brings WoT's TD approach into the AAS framework.
- **Asset Interface Description (AID):**
    - A submodel within AAS that uses WoT's Thing Description to describe asset interfaces.
    - Enables plug-and-play asset onboarding and access to runtime data.
    - Based on the WoT architecture, including properties, actions, and events.
    - Aims to standardize how asset interfaces are described within the AAS framework.
    - An open source tool named "Asset Package Explorer" can be used to visualize AID submodels.
- **Key takeaways:**
    - The AID submodel bridges the gap between WoT and AAS, leveraging WoT's standardization for asset interface descriptions.
    - This integration simplifies IoT application development and asset management in industrial settings.
    - The generic nature of AAS allows it to be used in many different application domains, not just industrial.
    - The use of JSON-LD in WoT allows for the inclusion of semantic information, but it is up to the user to include that semantic information.
    - Siemens has started to publish Thing descriptions for many of its devices.

The discussion following the talk focused on the availability and semantic enrichment of data published by Siemens Energy, specifically regarding Asset Administration Shells (AAS) and Thing Descriptions (TDs). Here's a summary of the key points:

- **Siemens Energy's Data Publication:**

    - Siemens Energy has begun publishing Thing Descriptions for their devices, particularly those from the smart infrastructure business unit.
    - This data is available in a GitHub repository, with around 1000 device descriptions available for download.
- **Semantic Meaning of Thing Descriptions:**
    - While TDs provide interface descriptions, enriching them with semantic meaning is crucial.
    - TDs are based on RDF and can be serialized as JSON-LD, allowing for the integration of existing ontologies.
    - Users can add semantic annotations to TDs to provide context and meaning, especially when using devices in specific environments like building automation.
    - ECLASS annotations are being added to the thing descriptions.
- **AAS and Open Source Tools:**
    - AAS is seen as a valuable tool for exchanging diverse asset information.
    - Open-source tools like "ASA six" provide a starting point for visualizing and working with AAS data.
    - The aim is to extend the use of AAS beyond smart industry to broader applications.
- **Generic Applicability of AAS:**
    - Although originating from the industrial domain, AAS is considered a generic framework.
    - It can be applied to various sectors, including building automation.
    - AAS focuses on providing essential asset information (manuals, certificates, nameplates) that are relevant across different domains.
- **Semantic notation:**
    - There is a need to express semantics with a rich notation, rather than just through data definitions.

Keynote: Semantic interoperability in respect to Web of Things, NGSI-LD and OPC UA (Darko Anicic, Siemens) [slides]

The presentation discussed the role of standardized semantics in automation systems, focusing on the engineering aspect. Standardized semantics is essential for realizing devices, applications, and systems, both virtual and real. The presentation showcased successful implementations in the building and manufacturing domains through the NEPHELE and SmartEdge projects.

The presentation also discussed the challenges and solutions related to semantic interoperability in automation systems. One of the main challenges is heterogeneity, as various vendors and protocols make it difficult to integrate dynamically changing devices and describe their functionality in a standard way. To address this, the presentation proposed a semantic-driven engineering approach that enriches devices with additional functions and implements them via TinyML models and Complex Event Processing (CEP) rules. This approach enables checking the applicability of a rule for a device, deploying a rule to a selected set of devices, and extracting a rule from a Thing Model and executing it on a device. The presentation also highlighted the importance of W3C WoT standard for openness and interoperability of IoT devices.

This talk explored the practical application of standardized semantics in industrial automation systems, focusing on engineering devices, applications, and systems. Here's a summary of the main points:

- **Standardized Semantics in Automation:**

    - The speaker emphasized the importance of standardized semantics for interoperability and digital twins in industrial settings.
    - He presented use cases from EU research projects to demonstrate real-world applications.
    - The talk covered engineering at three levels: device, application, and system.
- **Engineering Devices (FELT Project):**
    - The project aimed to improve energy efficiency in HVAC systems by enhancing smart thermostats.
    - Challenges included device heterogeneity and dynamic function deployment.
    - Solutions involved embedding engines (microcomplex event processing, tiny machine learning) and using W3C Thing Descriptions (TDs) to expose device functionalities.
    - Thing Templates were also used to create device models, and Thing Models were used to deploy rules and machine learning models.
    - The project used semantic artifacts to automatically generate thing descriptions when new rules were deployed to the physical devices.
- **Engineering Applications (FELT Project):**
    - The project used virtual objects (based on TDs) to describe device functionalities.
    - Interoperability challenges arose when integrating devices described by different standards (TDs and NGSI-LD).
    - The speaker highlighted ongoing efforts to bridge the interoperability gap between these standards.
    - The conversion of NGSI-LD data into thing descriptions was shown.
- **Engineering Systems (SmartEdge Project):**
    - The project focused on enabling customizable production using low-code platforms and large language models.
    - OPC-UA based semantics played a crucial role in creating capabilities and skills for dynamic production.
    - Capabilities represent functional requirements, while skills represent concrete implementations.
    - The project utilized semantic models for devices, recipes, and internal concepts.
    - The OPC-UA information model was used to create semantic artifacts.
    - The project demonstrated the use of a digital twin environment to simulate and validate production recipes before deploying them to real systems.
    - A need for an RDF representation of the OPC-UA standard was identified.
- **Key Takeaways:**
    - Standardized semantics are essential for achieving interoperability and automation in complex industrial systems.
    - W3C Thing Descriptions and OPC-UA information models are valuable tools for representing device and system functionalities.
    - Challenges remain in bridging the gap between different semantic standards.
    - Semantic data integration and matchmaking are critical for dynamic production environments.
    - Digital twins enable the simulation and validation of automation processes.

The discussion following this talk highlighted several key points regarding semantics and interoperability in industrial automation:

- **NGSI-LD to W3C Thing Description (TD) Conversion:**
    - The conversion is primarily syntactic and should ideally be automated with tools.
    - Automation should only be applied when there's a precise, high-fidelity mapping of capabilities between the two standards.
- **Standardization of "Recipes" or Profiles:**
    - "Recipes," as used in the SmartEdge project, are essentially profiles that define how to use specific information models.
    - Standardizing these profiles is crucial to prevent the creation of silos and ensure interoperability.
    - There's a recognized need for standardizing engineering templates for systems, but progress has been slow.
- **Semantics vs. Data:**
    - A critical distinction is made between semantics and data.
    - Semantics is not data itself but rather a set of guidelines or rules for manipulating data or interacting with assets.
    - Semantics can be expressed using mathematical formulas (e.g., differential equations, graph theory).
    - Data should be considered an asset separate from physical assets.
- **Industry's Approach to Semantics:**
    - Industry is gradually adopting semantics, starting with domain-specific models and vocabularies.
    - There's a recognition that semantics can encompass more complex representations, such as axioms and rules.
    - Scalability and usability for engineers are essential considerations in implementing semantic technologies.
    - Industry is working towards more advanced semantic implementations, but is moving forward step by step.
- **The use of Thing Models:**
    - Thing models are used to describe rules, and those rules have semantic meaning.

Keynote: Ontology for Digital Thread (Cordelia M Ezhilarasu, SLB) [slides]

The presentation discussed the use of digital twins in the oil and gas industry, specifically in the areas of prognostics and health management (PHM). The presentation also discussed the importance of ontologies and taxonomies for digital twins. Ontologies are used to define the relationships between different concepts, while taxonomies are used to classify and organize data. The presentation argued that it is important to have clear ownership, governance, and responsibility for ontologies and taxonomies. The presentation also discussed the challenges of developing and deploying digital twins. One challenge is that oil and gas systems are often custom-made and reconfigurable. This makes it difficult to develop digital twins that can accurately represent the system. Another challenge is that digital twins need to be able to handle a wide range of data, including data from different sources and in different formats. The presentation argued that it is important to have a well-defined data strategy for digital twins.

The presentation also discussed the benefits of using digital twins. Digital twins can be used to improve the performance of oil and gas systems, reduce emissions, and improve safety. They can also be used to train operators and to plan and optimize operations. The presentation gave several examples of how SLB is using digital twins in its operations. These examples include using digital twins for PHM, for planning and optimizing drilling operations, and for training operators. The presentation concluded that digital twins are a valuable tool for the oil and gas industry. They can be used to improve safety, efficiency, and sustainability.

This talk from SLB focused on their approach to digital twins and digital thread, emphasizing the crucial role of knowledge graphs and flexible, decentralized taxonomies. Here's a summary of the main points:

- **SLB's Energy Transition:**
    - SLB is a major energy services company focusing on secure, sustainable, and affordable energy.
    - They are actively involved in decarbonization, methane reduction, and developing new energy systems (green hydrogen, geothermal, energy storage).
    - Balancing energy needs with environmental sustainability is a core concern.
- **System of Systems Approach:**
    - SLB's operations involve complex, modular technologies, requiring a "system of systems" approach.
    - This approach extends throughout their technology development pipeline, from design to deployment.
    - They aim to transfer lessons learned across different business areas.
- **Digital Thread and Knowledge Graphs:**
    - The goal of their digital thread is to enable common applications, shared lessons learned, and seamless collaboration.
    - This is achieved through:
        - **Consistent Taxonomies:** Subject matter experts own and maintain definitions, focusing on component reuse at the branch and leaf level.
        - **Compact Ontologies:** Application-focused ontologies enable specific functionalities (e.g., AI planning, risk assessment).
        - **Knowledge Graphs:** These connect ontologies and link systems, providing a holistic view for decision-making.
- **Decentralized Taxonomy Management:**
    - SLB avoids a top-down, corporate-level taxonomy.
    - Instead, they promote distributed ownership and management of taxonomies by subject matter experts.
    - This allows for flexibility and ensures that taxonomies are "fit for purpose."
- **Applications and "Twin of Twins":**
    - Ontologies enable the automatic generation of applications across various business lines.
    - The "twin of twins" concept connects multiple digital twins, allowing for system-level simulations and analyses.
    - This is crucial for understanding interactions between systems and predicting cascading failures.
- **Prognostics and Health Management (PHM):**
    - Knowledge graphs are essential for understanding relationships between systems and line replaceable units (LRUs) in PHM.
    - They enable a holistic view of system-level health, improving maintenance and troubleshooting.
- **Edge Deployment:**
    - Compute constraints at the edge require ontological patterns that support responsive reconfiguration and efficiency.
- **Core Principles:**
    - Ownership of definitions by decision-makers.
    - Responsibility close to the digital thread definition.

    - Oversight rather than top-down design.
    - The use of Knowledge graphs to tie together the different ontologies and taxonomies.

The following discussion centred on the practical application of ontologies and knowledge graphs, particularly in industrial contexts, and raised several key points:

- **Graphical Representation of Data Manipulation:**
    - There's a recognized need for graphical elements in representing data processing and manipulation.
    - This representation should be formalized and standardized to facilitate exchange and interoperability between stakeholders.
- **Generalizability of SLB's Approach:**
    - SLB's approach, while presented in the context of oil and gas, is applicable to various domains.
    - The emphasis on linking architecture, features, ontologies, and behaviour is broadly relevant.
    - The use of taxonomies as profiles or "recipes" when engineering is a key point.
    - The importance of describing behaviours with graphs was reiterated.
- **Concise vs. Large Ontologies:**
    - A trade-off exists between creating concise, application-specific ontologies and large, comprehensive ontologies.
    - While concise ontologies are useful for specific tasks, large ontologies are essential for broad semantic interoperability.
    - Examples like HL7 and FLKIF demonstrate the value of large ontologies in specific domains.
    - Mapping terms between different ontologies is a major challenge.
- **Tools for Ontology and Taxonomy Development:**
    - SLB is experimenting with various tools, including TopQuadrant and open-source Python Protégé, to determine the best solution for their needs.
- **Call for Collaboration:**
    - There's a call for SLB to participate in standardization efforts, leveraging their practical experience.
- **Semantic Interoperability:**
    - The importance of semantic interoperability was raised, and how difficult it is to achieve.
- **The use of taxonomies as recipes:**
    - The idea that taxonomies can be used as recipes was raised, and it was suggested that this idea should be leveraged.

Session 3: Semantic Interoperability for Digital Twins

Keynote: Semantic interoperability for dataspaces workshop series (Martin Kaltenböck, Graphwhise/SWC) [slides]

The presentation discussed the importance of semantic interoperability for data spaces. Semantic interoperability is the ability of systems and services that create, exchange, and consume data to have clear, shared expectations for the contents, context, and meaning of that data. The presentation also discussed the barriers, concerns, approaches, and technologies related to semantic interoperability. One of the main barriers is that interoperability is often only at the metadata level, not at the data level. This means that systems can exchange

data, but they may not be able to understand the meaning of the data. Another concern is that there is a lack of industry-specific controlled vocabularies and knowledge graphs. This makes it difficult to find and use relevant data.

The presentation also discussed the importance of interdisciplinary collaboration and centralized coordination for semantic interoperability. Interdisciplinary collaboration is important because it allows different stakeholders to come together and share their knowledge and expertise. Centralized coordination is important because it helps to ensure that interoperability efforts are aligned and that they are making progress. The presentation also highlighted the success of the Semantic Layer for Data Spaces project. This project has developed a number of tools and services that can help to improve semantic interoperability. These tools and services include a semantic annotation service, a semantic data harmonization service, and a semantic data discovery service.

This presentation summarized the activities and findings of a series of workshops focused on semantic interoperability within data spaces, highlighting key challenges and potential solutions. Here are the main points:

- **Context of Data Spaces:**
    - The discussion takes place within a framework of increasing European legislation (Data Act, AI Act) and the development of common European data spaces.
    - Organizations like the European Data Innovation Board and the Data Spaces Support Centre are crucial in fostering this movement.
    - Semantic interoperability is essential for enabling seamless data flow across these spaces.
    - Resources like Semic (Interoperable Europe) and the European Interoperability Framework are important for standardization efforts.
- **Workshop Series Overview:**
    - A series of workshops, organized by Graphwise, W3C Europe (ERCIM), Big Data Value Association (BDVA), and IDSA, has been conducted over three years.
    - These workshops aimed to identify and address gaps in semantic interoperability within data spaces.
    - Findings from each year's workshop were highlighted.
- **Key Findings:**
    - **2022:**
        - Interoperability is often limited to metadata, not data.
        - There's a lack of industry-specific controlled vocabularies and knowledge graphs.
        - Multilingual support is insufficient.
        - Data usage control mechanisms are lacking.
        - The connection between the physical and digital worlds needs improvement.
    - **2023:**
        - Interdisciplinary collaboration across industries is crucial.
        - Centralized coordination of interoperability development and standardization is needed.
    - **2024:**
        - Balancing unified models with mappings between diverse models is a challenge.
        - Reuse of existing models and cost-effective mapping approaches are essential.

        - Generative AI can assist in knowledge graph creation and interoperability.
        - Human decisions and social aspects remain important.
        - Data and vocabulary harmonization services are needed.
        - Regulatory, legal, and governance aspects are crucial.
        - Model quality and maintenance are key considerations.
        - The topic of vocabulary hubs and how to make them interoperable was discussed.
- **Vocabulary Hubs:**
    - The importance of vocabulary hubs for maintaining and resolving vocabularies was emphasized.
    - The need for interoperability between different vocabulary hubs was highlighted, with a collaboration between TNO and Graphwise being mentioned.

Underpin Success Story:

- The Underpin data space for industry, focusing on predictive maintenance, was presented as a success story.
- It demonstrates the implementation of a semantic layer for metadata annotation and data harmonization.

- **Call for Participation:**
    - The International Workshop on Semantics in Data Spaces (STS) was mentioned, with a call for participation.
    - The Endorsed conference in October in Brussels was also highlighted.
- **Ontology alignment:**
    - The connection between the workshops work, and the work of previous presentations on ontology alignment was made.

The following discussion expanded on the challenges and progress related to semantic interoperability in data spaces, particularly focusing on data quality, implementation, and future directions. Here's a summary of the key points:

- **Data Quality and Selection:**
    - The challenge of determining the "best" data set within a data space was raised.
    - The value of a data set is highly dependent on the specific use case.
    - Purpose modelling and trial data are suggested approaches for evaluating data suitability.
    - Semantic definitions and rule compliance can also aid in data selection.
- **Data Space Semantic Interoperability Meeting in October:**
    - The program for the October meeting in Brussels is not yet finalized, allowing for potential presentation submissions.
    - The W3C Data Space Community Group is a platform for coordinating related efforts.
- **Evolution of Challenges Since 2022:**
    - While the core challenges identified in 2022 remain relevant, the focus has shifted towards data interoperability rather than just metadata.
    - The lack of standardized vocabularies and knowledge graphs continues to be a significant obstacle.

    - Implementation of semantic standards is lagging behind specifications and discussions.
- **Cost of Ontology and Vocabulary Reuse:**
    - The cost associated with reusing ontologies and vocabularies is a major concern.
    - This cost, combined with the complexities of data interoperability, hinders widespread adoption.
- **Guidance and Reference Implementations:**
    - There's a strong need for practical guidance on implementing semantic interoperability in data spaces.
    - Providing reference implementations is crucial for demonstrating the feasibility and benefits of these approaches.
    - The need to move from "big bang" implementation, to incremental progress was raised.
- **Thing Descriptions vs. Data Descriptions:**
    - The discussion highlighted the fact that within the Web of Things, Thing Descriptions exist, but data descriptions are not being created. This is a problem, because there is an assumption that everyone understands JSON, but this is not the case.
- **Implementation over theory:**
    - The overall message of the discussion was that there is a need to move from theoretical discussions to practical implementations.

Keynote: Semantic Problems in Dataspaces (Vladimir Alexiev, Graphwise/Ontotext) [slides] [paper]

The presentation discussed semantic problems in dataspaces, focusing on the Eclipse Dataspace Components (EDC). The presentation argued that EDC does not use semantic technologies in its implementation, which leads to a number of problems, such as difficulty extending EDC with new semantic metadata, improper content type and context, wrong namespaces, no base; bad instance URIs, disconnected RDF, and not conforming to ODRL. The presentation also discussed a number of other dataspace semantic problems, such as problems with Sovity, IDSA, and GAIA-X. The presentation concluded with a call for the dataspace community to not ignore decades of semantic interoperability experience.

The presentation further discussed a number of possible solutions to the semantic problems in dataspaces. These solutions include using JSON-LD for semantic/RDF purposes, using W3C test approaches in the TCK underdevelopment, and emphasizing JSON-LD, not only JSON. The presentation discussed the importance of having a clear understanding of the domain for which the dataspace is being developed, the need to carefully evaluate existing ontologies before reusing them, and the importance of using a modular approach to ontology development. The presentation concluded with a discussion of the future of semantic interoperability in dataspaces. The presentation argued that semantic interoperability is essential for the development of AI-powered applications and services.

This presentation provided a critical, practical perspective on the current state of semantic metadata within data spaces, particularly focusing on the EDC (Eclipse Dataspace Components) implementation. Here's a summary of the key points:

- **Practical Experience and Concerns:**
    - The presentation was based on the speaker's first-hand experience working with EDC and data spaces.

    - It highlighted significant limitations in the semantic metadata provided by EDC, despite its widespread use.
    - The speaker called for action and feedback from the community to address these issues.
- **EDC Limitations:**
    - EDC exhibits limited support for semantic metadata, primarily due to its hard-coded Java implementation.
    - It lacks proper JSON-LD context and does not adequately declare property types, leading to disconnected RDF graphs.
    - Incorrect namespaces, property names, and URL handling were observed.
    - EDC doesn't conform to the ODRL ontology, using non-existing property *odrl:type*, and strings like *USE* instead of resources
    - It stores semantic data in a text field, and extracts only limited fields to Postgres, rather than using proper semantic representations.
    - Extending the EDC to support additional semantic representations is a difficult and time consuming process.
- **Underpin Project and Semantic Model:**
    - The Underpin project, focusing on manufacturing and maintenance data, developed a semantic model using various ontologies (CSVW, SOSA, QUDT, SKOS).
    - This model aims to describe tabular data sets with rich semantic metadata, going beyond DCAT in describing dataset columns.
    - Little of this semantic model is supported by EDC.
- **Issues and Recommendations:**
    - The speaker also raised issues on the IDSA (International Data Spaces Association) and GAIA-X GitHub repositories regarding incorrect SHACL shapes and content types.
    - They criticized the lack of dedicated semantic work in the SIMPL project.
    - A core EDC developer stated that JSON-LD is not needed, and semantic technologies make interoperability harder, prioritizing JSON schema instead.
    - The speaker recommended reusing W3C test conformance practices for dataspace testing.
    - The speaker made a point that although JSON schema is useful, the JSON that is produced, should also be valid JSON-LD, which is not currently the case.
- **Call to Action:**
    - The speaker emphasized the need for the data space community to embrace semantic technologies and learn from past experiences.
    - They urged the semantics community to actively engage with data space initiatives and provide practical examples and solutions.
    - The speaker made it clear that the semantic community needs to lead by example, and actively contribute to the solution, rather than just pointing out the problems.
- **Interoperability:**
    - The speaker made it clear that they believed that Semantics is very useful for interoperability, and not a burden to it.

The following discussion focused on the evolution of data spaces, particularly in the context of the International Data Spaces Association (IDSA) and the challenges of implementing semantic technologies in practical applications. Here's a summary of the key points:

- **Evolution of Data Spaces:**
    - Data spaces initially focused on connecting factories, but are now being repurposed for broader participation.
    - This historical focus might hinder the adoption of data spaces by a wider range of participants.
- **IDSA's Contributions and Challenges:**
    - IDSA has developed a comprehensive data model that includes classes and properties for describing data usage, ownership, and licenses.
    - However, these semantic elements are not widely used in practice.
    - This lack of practical usage is a problem.
- **The Difficulty of Semantics:**
    - Even after 20 years of development, many people still find semantics challenging.
    - This is a large hurdle to overcome.
- **Need for Collaborative Action:**
    - The speaker emphasized the importance of collaborative action from academia, industry, standardization bodies, and the European Commission.
    - Simply pointing out errors is not sufficient; detailed reporting and systematic support are needed.
    - Everyone needs to work together, because a missing part from any group, prevents progress.
- **The need to "lend a hand":**
    - The speaker made it clear that the semantic community needs to actively help with the implementation of semantics, rather than just pointing out where things are wrong.

Long Talk: Hourglass pattern for MIM and semantic interoperability (Antonio Kung, Trialog) [slides]

The presentation discussed the hourglass pattern, which is a model for achieving minimum interoperability mechanisms (MIM) and semantic interoperability. The hourglass pattern consists of three layers: applications, infrastructure, and MIM. The applications layer contains the various applications that need to interoperate. The infrastructure layer provides the necessary infrastructure for the applications to interoperate, such as networks, computing resources, and storage. The MIM layer provides the minimum set of mechanisms that are needed for the applications to interoperate, such as common data models, protocols, and APIs. The presentation also discussed the evolution of interoperability standards and architecture standards. The presentation argued that the hourglass pattern is a valuable tool for achieving interoperability in a variety of domains, such as IoT, cloud computing, and data spaces.

The presentation also discussed a number of examples of how the hourglass pattern can be used to achieve interoperability. These examples include the use of the hourglass pattern to achieve interoperability in digital twins and data spaces. The presentation concluded with a discussion of a roadmap for achieving semantic interoperability. This roadmap includes the development of new standards, the development of new tools and technologies, and the promotion of collaboration between different stakeholders. The presentation was well-received by the audience and provided a valuable overview of the hourglass pattern and its potential for achieving interoperability.

This presentation focused on the critical need for reuse and standardization across various domains to achieve effective interoperability, particularly in the context of digital twins, data spaces, and smart city applications. Here's a summary of the main points:

- **The "Hourglass" Pattern:**
    - The presentation introduced the "hourglass" pattern to illustrate the problem of siloed implementations and the need for reuse.
    - The goal is to achieve reuse at every layer of implementation to reduce costs and complexity.
- **Key Components for Interoperability:**
    - The speaker highlighted the need for:
        - Information Model Repositories (IMMO): Sharing semantic domains, ontologies, and information models.
        - Architecture Patterns: Sharing common architectural patterns across domains.
        - Minimum Interoperability Profiles (MIPs): Committing solutions to standardized profiles.
        - Open Source (or Accessible) Implementations: Promoting reusable components.
- **The Importance of Cross-Domain Standardization:**
    - The speaker emphasized the need to extend interoperability standards to support cross-domain applications.
    - This requires collaboration and agreement on reusable ontology patterns and architecture patterns.
- **Behavioural and Policy Interoperability:**
    - Interoperability should encompass not only data exchange but also behavioural and policy aspects.
    - This requires describing interactions, behaviours, and policies using semantic models.
- **The Role of Ontologies and Profiles:**
    - Ontologies should be structured into reusable patterns that can be shared across domains.
    - Ontology profiles should be used to specify constraints and ensure trustworthiness.
    - Reuse of existing ontology patterns is crucial to avoid reinventing the wheel.
- **Architecture Standards and Patterns:**
    - Architecture standards are evolving to incorporate reusable patterns.
    - Examples were given for data spaces and smart manufacturing.
    - Collaboration with relevant standardization bodies is essential.
- **Data Spaces and Interoperability:**
    - Data space interoperability requires consideration of various standards, including data standards, interoperability standards, architectural standards, security and privacy standards, and AI standards.
    - All these standards must align to achieve the goal of proper data sharing with appropriate semantics.
- **The need for working in sync:**
    - All the different profile types must be worked on in sync, or interoperability will not be achieved.
- **Call for collaboration:**
    - The speaker called for collaboration between different domains, and standardisation bodies.
- **Announcement of a workshop:**

    - The speaker announced a workshop organized by ISO on the smart city ontology.

The following discussion centred on the crucial topic of aligning different methodologies and resources to ensure effective interoperability, particularly in the context of data spaces and ontologies. Here's a summary of the key points:

- **Alignment with Cornelius's Four Mapping Aspects:**
    - The discussion highlighted the importance of aligning the four mapping aspects presented by Cornelius in the white paper.
    - It was emphasized that misalignment among these aspects could hinder interoperability.
- **Regular Meetings and Collaborative Solutions:**
    - The proposed solution involved establishing regular meetings to discuss and resolve interoperability challenges.
    - Leveraging resources from TNO (Netherlands Organisation for Applied Scientific Research) and other data space experts was considered essential.
- **Evaluation of Maturity (TRL):**
    - The focus shifted towards evaluating the maturity (Technology Readiness Level - TRL) of different methodologies.
    - The goal was not simply to integrate everything but to assess the progress and suitability of existing ontologies and tools.
    - The idea was to use community tools to evaluate the progress of already existing ecosystems.
- **Avoiding "Glue and Paste" Integration:**
    - It was stressed that integration should not be a simple "glue and paste" approach.
    - Instead, a more strategic and evaluated approach was favoured.
- **Leveraging TNO's Treehouse Feature:**
    - The discussion recognized the valuable feature of TNO's Treehouse for aligning ontologies with upper ontologies.
    - It was suggested that this feature could be used to transform ISO PDF documents into more usable formats.
- **Vocabulary Hub:**
    - The vocabulary hub was referenced as moving in the right direction, but that a wider picture needed to be taken into account.
- **The need for a high level integration:**
    - The need for a high level integration of the different resources was raised.

Talk: Semantic Interoperability for Environmental Digital Twins and Data Spaces (Arne Berre, SINTEF AS) [slides]

The presentation discussed the challenges and opportunities of using semantic interoperability for environmental digital twins and data spaces. Environmental digital twins are digital representations of natural processes and functions, such as the ocean, the water cycle, weather/climate, and biodiversity. Data spaces and data lakes are used to store and manage the data that is used by environmental digital twins. Semantic interoperability is the ability of different systems to exchange and understand data with the same meaning. This is essential for environmental digital twins and data spaces to effectively interoperate.

The presentation highlighted a number of use cases where semantic interoperability is being used to support environmental digital twins and data spaces. These use cases include the European Green Deal, the Digital Twin of the Ocean, and the AquaINFRA EOSC project. The presentation also discussed a number of tools and technologies that can be used to support semantic interoperability, such as ontologies, knowledge graphs, and semantic APIs. The presentation concluded with a call for more research and development in the area of semantic interoperability for environmental digital twins and data spaces.

This presentation highlighted the growing importance of environmental digital twins, particularly within the Destination Earth and Mission Ocean programs, and the significant role semantic interoperability plays in their development. Here's a summary of the key points:

- **Environmental Digital Twins:**
    - The presentation distinguished between asset-oriented digital twins (e.g., AAS) and environmental digital twins (e.g., Destination Earth).
    - Environmental digital twins aim to model and predict natural phenomena like climate change, extreme weather, and ocean dynamics.
    - Destination Earth is a European program creating digital twins for various aspects of the natural world.
    - Mission Ocean is a related program focused on digital twins of the ocean.
- **Destination Earth and Mission Ocean:**
    - Destination Earth's platform includes a data lake connected to various data spaces (e.g., energy, agriculture, green deal) and digital twin engines.
    - Mission Ocean's architecture follows a similar pattern, with a data lake, digital twin engine, and services.
    - Both programs rely on data from diverse sources, including sensors, observations, and satellite data.
- **Semantic Interoperability Challenges:**
    - Semantic interoperability is crucial for integrating data from different sources into the data lakes.
    - The earth observation and geospatial ocean communities currently rely on controlled vocabularies and terminology definitions (e.g., climate forecasting convention) rather than semantic technologies.
    - Introducing semantic technologies into these communities is a significant challenge.
    - There is a need to combine elements from multiple ontologies to create common information models (e.g., ocean information model).
- **Data Spaces and Data Lakes:**
    - Data spaces play a role in handling data from various sources and feeding it into the data lakes.
    - The green deal data space is particularly relevant for ecosystem and biodiversity data.
    - The European Open Science Cloud and Copernicus satellite data are also important data sources.
- **Beyond Data Semantics:**
    - Interoperability should encompass not only data semantics but also the semantics of services, processes, and policies.
    - There is a need to unify and create synergies among the various semantic tools being used in these projects.
- **The need for Multi Domain wise agreement:**

    - The speaker made it clear that there is a need to have a multi domain wise agreement of how to handle the semantic interoperability.
- **Digital twin definitions:**
    - Digital twins are defined as models of reality that are synchronized two ways at a certain frequency and fidelity in time and space.

Talk: Ontologies and the Data Continuum and their use in the Edge Intelligence Industry Applications (Martin Serrano, Insight SFI Research Centre) [slides]

The presentation discussed the importance of ontologies and the data continuum for semantic interoperability in data spaces. Ontologies are formal representations of knowledge that can be used to define the meaning of data. The data continuum is the idea that data should be collected, classified, shared, and made interoperable. The presentation also discussed a practical use case of using ontologies and the data continuum for automotive diagnostics. In this use case, ontologies were used to represent the knowledge of automotive experts. The data continuum was used to collect, classify, share, and make interoperable data from a variety of sources, including sensors, repair manuals, and diagnostic tools. The presentation concluded that ontologies and the data continuum are essential for semantic interoperability in data spaces. They can be used to improve the accuracy and efficiency of data analysis and decision making.

The presentation also discussed the role of semantic web technologies in IoT. Semantic web technologies can be used to collect, exchange, and share IoT data. The presentation highlighted the benefits of using semantic web technologies for IoT, such as the ability to share and reuse structured and already designed domain knowledge, interconnect datasets, describe data with common vocabularies, and enable machine reasoning to interpret sensor data. The presentation also discussed the IoT stack and the data space interoperability vision. The IoT stack is a layered architecture that describes the different components of an IoT system. The data space interoperability vision is the idea that data should be accessible to everyone, regardless of where it is stored. The presentation concluded with a discussion of the future of semantic interoperability in data spaces. The presentation argued that semantic interoperability is essential for the development of AI-powered applications and services.

The IoT interconnects devices to improve services. Linked data and Semantic Web as a basis for efficient stream processing, data management and smart applications. This enables machines to interpret data, reusing domain knowledge.

This talk and subsequent discussion focused on the practical application of ontologies in the digital twin era, emphasizing the need for engineering principles, data interconnectivity, and machine understanding. Here's a summary of the key points:

- **Ontologies as Enablers:**
    - Ontologies are seen as crucial for enabling semantic interoperability in digital twin applications.
    - The focus should shift from pure modelling to ontology engineering, emphasizing practical application.
- **Data Flow and IoT Stack Evolution:**
    - The IoT stack has evolved from four to six layers, including business and knowledge graph layers.
    - This evolution reflects the increasing importance of data interconnectivity and machine understanding.
    - Data lakes and cloud technologies are central to this evolution.
- **Machine Understanding and AI:**

    - The goal is to enable machines to interpret and understand data, reducing reliance on human intervention.
    - Large language models (LLMs) and knowledge graphs play a vital role in this process.
    - AI-powered services and applications are transforming the IoT landscape.
- **Practical Applications:**
    - The speaker presented a video demonstrating an AI-powered application that uses ontologies to diagnose car problems based on audio analysis.
    - This example highlights the potential of edge-based semantic applications.
    - The discussion also raised the topic of using acoustic sensors to monitor agriculture.
- **Challenges and Considerations:**
    - Real-world knowledge is often imprecise, requiring fuzzy mapping and context sensitivity.
    - The need for a feedback loop to improve models and understanding was emphasized.
    - The evolution of digital twins, from representing physical elements to complex systems that react and model themselves, was discussed.
- **Wake-Up Calls and Standardization:**
    - The discussion addressed the need for a collaborative approach to standardization, involving both top-down and bottom-up perspectives.
    - The importance of learning from other communities, such as the biomedical community, was highlighted.
    - The need for better testing and compliance mechanisms for semantic technologies was raised.
    - The necessity for unification of tools was raised.
    - The need for awareness of how to correctly contribute to international standards.
- **Compliance and Incentives:**
    - The possibility of creating compliance benefits or incentives for adhering to semantic standards was explored.
    - The role of regulatory bodies, such as the European Commission, in promoting standardization was discussed.
- **The need for working together:**
    - There was a clear message that the semantic community needs to work together to improve the application of semantics.

Session 4: Building Digital Twins and Applications

Keynote: OMG standards relevant to digital twins (Claude Baudoin, Object Management Group) [slides]

This presentation provided an overview of the Object Management Group (OMG) and its various activities, highlighting its relevance to semantic interoperability and digital twin technologies. Key points include:

- **OMG Overview:**
    - OMG is a 35-year-old consortium involved in various standardization efforts.
    - Its scope includes systems modelling, ontologies, product lifecycle management (PLM), middleware, agent and event modelling, and AI.

    - OMG is structured with domain and platform task forces, special interest groups, and the Digital Twin Consortium.
    - OMG holds quarterly meetings for members, with opportunities for invited guests.
- **SysML v2:**
    - SysML v2 is a recent standard with a stronger semantic foundation than its predecessor.
    - It is grounded in formal semantics and based on model libraries.
    - The metamodel is no longer constrained by UML.
    - It includes the Kernel Modelling Language (Kernel) with both graphical and textual representations.
    - Kernel is intended to make SysML v2 more machine readable.
- **Ontology definition metamodel (ODM)**
    - ODM is part of a family of specifications. Manufacturing technology and industrial systems (ManTIS).
- **Ontology Work:**
    - OMG has an Ontology Special Interest Group that develops ontology-related specifications.
    - It has contributed to domain ontologies, including those for finance, basic coding systems, and retail.
    - A robotic service ontology is also under development.
- **Manufacturing and PLM:**
    - OMG's Manufacturing, Technology, and Industrial Systems (Mantis) task force focuses on PLM and related standards.
    - It is currently working on Cascade, a standard for collaborative artifact specification, context, and data exchange.
    - Cascade is heavily involved in the digital thread of product information through the supply chain.
- **OMG's Standardization Process:**
    - OMG uses a request for proposal (RFP) process, where members submit proposals for standards.
    - Adopted standards require commitments from members for implementation.
    - The goal is to avoid "shelfware" standards that are not put into practice.
- **Other Relevant Activities:**
    - OMG has developed the Data Distribution Service (DDS) for middleware.
    - It is also involved in enterprise resource metadata standards and agent and event modelling.
    - OMG is also working on standardizing neural network definitions.
- **Call to Action:**
    - The speaker encouraged attendees to learn more about OMG and its activities.
    - He offered to provide guest passes to OMG meetings and information about trial memberships.
    - He provided contact information, and website information.

Keynote: OMG Commons Ontology Library and related standards (Elisa Kendall, OMG, remote from California) [slides]

The presentation discussed the importance of standard ontologies for cross-domain reuse. Ontologies are formal representations of knowledge that can be used to define the meaning of data. The presentation also discussed the OMG's Multiple Vocabulary Facility (MVF). MVF is a standard that enables modelers to use multiple vocabularies in their models. The presentation

highlighted the benefits of using standard ontologies and MVF, such as improved interoperability, reduced development costs, and increased innovation. The presentation also discussed a number of use cases where standard ontologies and MVF are being used, such as in the financial industry, the pharmaceutical industry, and the retail industry. The presentation concluded with a discussion of the future of standard ontologies and MVF. The presentation argued that these technologies are essential for the development of AI-powered applications and services.

The presentation also discussed the OMG's work on the Commons Ontology Library. The Commons Ontology Library is a library of common ontologies that can be used by different industries. The presentation highlighted the benefits of using the Commons Ontology Library, such as improved interoperability and reduced development costs. The presentation also discussed a number of new ontologies that are being added to the Commons Ontology Library, such as ontologies for tensors and vectors, sites and facilities, addresses, reporting, statistical measures, and products. The presentation concluded with a discussion of the future of the Commons Ontology Library. The presentation argued that the library is essential for the development of semantic interoperability.

Financial Industry Business ontology (FIBO) is used in banking. It is semantically rich, forward looking with pattern driven design supporting machine learning and AI. These patterns are useful across other domains. We've tested that for an ontology for pharmaceuticals (IDMP-O). The common names are insufficient to identify the drugs with sufficient precision. The aim of the new ontology is to eliminate ambiguities. Companies are also working on their own extensions using the design patterns provided. The same patterns are applicable to digital twins when integrated with top level ontologies.

OMG seeks to enable people to use their own human language and nomenclature for existing standards (multiple vocabulary facility - MVF). Commons Ontology Library specification version 1.3 with definitions aligned to W3C's OWL.

This presentation focused on the work being done at the Object Management Group (OMG) in developing reusable ontology patterns. Here's a summary of the key points:

- **FIBO and Pharmaceutical Ontology:**
    - FIBO is a large, long-lived ontology focused on financial data, being increasingly used for wholesale and commercial banking.
    - The pharmaceutical ontology aims to standardize data related to medicinal products, addressing ambiguity in ISO standards.
    - Both projects highlight the need for consistent patterns to enable machine learning and interoperability.
- **Commons Ontology Library:**
    - The work on FIBO and the pharmaceutical ontology led to the development of the Commons Ontology Library at OMG.
    - This library provides reusable patterns for various domains, including controlled vocabularies, designations, classification schemes, parties, roles, organizations, and locations.
    - The library is designed for extension and welcomes contributions.
- **Multiple Vocabulary Facility (MVF):**
    - MVF addresses the limitations of modelling tools by allowing users to swap in different languages and nomenclatures.
    - It includes standard ontologies for representing controlled vocabularies and mapping them to ontology terms.
    - MVF facilitates mappings between different vocabularies.

- **Benefits of Reusable Patterns:**
    - Reusing patterns has saved significant development time and facilitated interoperability.
    - Consistent use of patterns provides building blocks for machine learning, natural language processing, and large language models.
    - The patterns have been widely tested and adopted by various organizations.
- **Future Plans:**
    - Plans for Commons 1.3 include adding ontologies for tensors and vectors, quantities and units, sites and facilities, addresses, statistical measures, and products.
    - The creation of an ontology service, similar to bio portal, is being discussed.
- **Call to Action:**
    - The speaker encouraged attendees to use the Commons Ontology Library and provide feedback and requirements.
    - They also noted that OMG welcomes contributions to the library.
- **Top Level Ontologies:**
    - The speaker recommended that if someone uses a top level ontology, that they need to fully understand the commitments that the ontology forces them to make.
- **The success of the patterns:**
    - The speaker made it clear that the patterns developed at OMG are being used, and that they are saving significant amounts of development time.

The following discussion focused on clarifying aspects of OMG's work, addressing questions about mapping standards, ontology libraries, and practical experiences with ontology development and implementation. Here's a summary of the key points:

- **Mapping Standards:**
    - There are mappings from MOF to RDF available, and MOF to link ML mappings were also developed.
    - The OMG website is the primary source for these resources, and direct contact with the speaker was offered for further assistance.
- **Ontology Libraries and Industrial Ontologies:**
    - The importance of ontology libraries was highlighted, as opposed to isolated ontologies.
    - The Industrial Ontologies Foundry (IOF), led by NIST and standardized by OAGI, was mentioned as a relevant effort.
    - IOF focuses on areas like supply chain, maintenance, and planning, which are relevant to digital twins.
    - The Caesar Posca Association's work in developing ISO standards for the energy industry was also acknowledged.
- **IDMP Extension and Return on Experience:**
    - The discussion explored how the IDMP ontology is extended and used internally by pharmaceutical companies.
    - OMG uses a collaborative approach, involving use case development, regression testing, and feedback from pharma companies.
    - Implementation progress is monitored, and the ontology is refined based on real-world experiences.
    - The speaker offered to share more details about this process in offline discussions.

- **Addressing RDF Representation Problems:**
    - The issue of RDF representation problems in product libraries like eCl@ss and digital twin descriptions like AAS was raised.
    - The speaker emphasized the importance of collaborative environments for ontology development.
    - She described the EDM Council's innovation laboratory, which facilitates collaborative work and addresses specific use cases.
    - The speaker suggested that motivated people and a collaborative development environment can accelerate progress.
- **Collaboration:**
    - The discussion highlighted the importance of collaboration between different organisations, and different domains, to improve the development, and implementation of ontologies.

Talk: On Digital Twin Semantics (Jan B. de Meer, Object Management Group) [slides]

The presentation discussed the use of graph theory to model the behaviour of industrial digital twins. Graph theory is a branch of mathematics that studies the relationships between objects. The presentation argued that graph theory can be used to model the behaviour of industrial digital twins by representing the different components of the digital twin as nodes in a graph. The relationships between the components are represented by edges in the graph. The presentation also discussed the use of graph grammars to model the evolution of industrial digital twins. Graph grammars are a type of formal grammar that can be used to generate and manipulate graphs. The presentation argued that graph grammars can be used to model the evolution of industrial digital twins by defining rules that specify how the graph representing the digital twin can be transformed over time. The presentation concluded with a discussion of a number of applications of graph theory and graph grammars to industrial digital twins. These applications include the design of digital twin architectures, the development of digital twin simulations, and the analysis of digital twin data.

The presentation also discussed the use of UML to design SGAM Type Graph and Graph Grammar. UML is a modelling language that can be used to design and develop software systems. The presentation argued that UML can be used to design SGAM Type Graph and Graph Grammar by providing a graphical notation for representing the different components of the system and their relationships. The presentation concluded with a discussion of a number of benefits of using UML to design SGAM Type Graph and Graph Grammar. These benefits include improved communication between stakeholders, reduced development costs, and increased innovation.

This talk proposed a formal semantic representation using graph theory, emphasizing the need for both human understanding and machine execution. Here's a breakdown of the key points:

- **Formal Semantics Representation:**
    - The speaker argued for a twofold representation of formal semantics:
        - One that is understandable and expressible by humans.
        - One that is executable by machines.
    - A clear distinction between the "thing itself," its linguistic representation, and its conceptual understanding is crucial.
    - The speaker expressed concern about informal semantic definitions based on context and use cases.
- **Graph Theory Approach:**
    - The proposed solution utilizes graph theory, representing states and events.

    - Events are defined as relations between states, e.g., an event "A" interconnecting states "Y1" and "Y2."
    - Artifacts are represented as variables, and processes/computations are represented as edges, forming signal flow graphs.
- **State Changes and Graph Grammars:**
    - Modelling state changes of basic artifacts (object capabilities) is essential.
    - Graph grammar rewriting rules (left-hand side/right-hand side) are used to define modifications.
    - The left-hand side identifies matching patterns, and the right-hand side specifies changes to the graph.
    - This approach enables modelling behaviour and self-modifying algorithms.
- **Proof of Concept: Smart Grid Example:**
    - A proof of concept was presented using a smart grid example with a power management system and consumer automation system.
    - The objective was to maintain energy balance in the network to prevent stability problems.
    - OML notation was used to represent the graphs and their modifications, enabling declarative and operational semantics.
- **Semantic Specification:**
    - Semantic specifications are twofold:
        - Typecrafts: Basic topological functions of systems.
        - Modifications: How the system's behaviour changes.
    - The speaker gave an elevator example to show how the system is extended to represent modifications.

Talk: Virtualization of IoT devices and the development of Digital Twins software stack in the NEPHELE project (Nikolaos Filinis, National Technical University of Athens) [slides]

The presentation discussed the NEPHELE project, which is developing a software stack to achieve interoperability across the continuum for IoT devices. The presentation also discussed the contributions of the National Technical University of Athens to the W3C WoT Python Runtime. These contributions include the development of a three-file format for deploying virtual objects (VOs), the integration of different deployment models, the automatic instantiation of VOs, and the orchestration of VOs and their integration in application graphs. The presentation also discussed a number of features of the W3C WoT Python Runtime, such as security schemes, SQLite and InfluxDB integration, and generic functions.

The presentation further discussed the NEPHELE architecture. The NEPHELE architecture is a hierarchical approach that facilitates deployment through dashboard and HDAR environments. It also provides compute/network interplay mechanisms and a unified way of managing resources under the SMO. The presentation discussed the Digital Twin Development Framework. This framework comprises three layers: the physical layer, the digital twin layer, and the application layer. The physical layer comprises IoT and edge devices. The digital twin layer serves as the virtual counterpart of the physical entities. The application layer supports the development and orchestration of distributed applications. The presentation also discussed the VOStack architecture. This architecture provides a unified way of accessing data and simplifies application deployment. The presentation concluded with a discussion of the different deployment types for VOs.

This presentation focused on a project developing a software stack for the virtualization of IoT devices, enabling their integration and management within a hierarchical system of systems. Here's a summary of the key points:

- **Virtualization of IoT Devices:**

    - The project aims to virtualize IoT devices, creating digital twins for easier management and integration.
    - A software stack has been developed to enable this virtualization.
- **Hierarchical System of Systems:**
    - The architecture utilizes a hierarchical approach, integrating cloud and edge cloud components.
    - This allows for distributed deployment and management of IoT devices and applications.
- **Architecture Components:**
    - The system includes a development environment for creating application graphs.
    - An orchestrator is responsible for constructing deployment plans and placing services.
    - Virtual objects (digital twins) are placed close to the physical devices.
- **Logical Separation of Concerns:**
    - The project addresses interoperability at the physical layer, handling diverse IoT device protocols.
    - The digital twin layer adds functionality to enhance device capabilities.
    - The application layer enables data consumption and orchestration on platforms like Kubernetes.
- **Virtual Objects and Composite Virtual Objects:**
    - Virtual objects are virtualized instances of IoT devices.
    - Composite virtual objects aggregate multiple virtual objects.
- **Technology Stack:**
    - The project utilizes the Web of Things (WoT) Thing Descriptor.
    - Support for Lightweight Machine-to-Machine (LwM2M) semantic models is also included.
    - Tools are used to translate NGSI-LD models to WoT Thing Descriptors.
- **Deployment Types:**
    - Devices with computing capabilities can run the software runtime directly.
    - Devices without computing capabilities send data to their virtual counterparts.
- **Implementation Details:**
    - The project uses JSON Thing Descriptors and a script runner for deployment.
    - Docker images and Helm charts facilitate integration with orchestration platforms.
- **Contributions:**
    - The project simplifies deployment and integration on orchestration platforms.
    - It adds features like protocol support, security schemes, and database integration.
    - The project also maintains and improves an existing Python WoT implementation.
- **Visual clarity:**
    - The presenter was praised for the clarity of the visual aids.
- **Real world implementation:**
    - It was stated that the European Commission is utilising the visual aids produced by this project.

Talk: Semantic interoperability approach of the AURORAL H2020 project (Raúl García-Castro, Universidad Politécnica de Madrid) [slides]

The presentation discussed the AURORAL project, which is developing a decentralized semantic interoperability platform for smart communities and in rural areas. The presentation also discussed the AURORAL ontology, which is a core component of the AURORAL platform. The AURORAL ontology is a formal representation of the knowledge in the AURORAL domain. The presentation highlighted the benefits of using the AURORAL platform, such as improved data sharing, increased innovation, and reduced costs. The presentation also discussed a number of use cases where the AURORAL platform is being used, such as in the areas of tourism, energy, and mobility. The presentation concluded with a discussion of the future of the AURORAL project. The presentation argued that the AURORAL platform has the potential to transform the way that data is shared and used in smart communities and wider rural areas.

The presentation also discussed the AURORAL node for interoperability. The AURORAL node is part of a middleware that enables different systems to interoperate with each other. The presentation highlighted the benefits of using the AURORAL node, such as improved security, increased privacy, and reduced complexity. The presentation further discussed a number of features of the AURORAL node, such as its ability to discover, access, and understand data, as well as the ability of enforcing privacy policies following the W3C ODRL standard.

The presentation concluded with a discussion of the AURORAL ecosystem. The AURORAL ecosystem is a network of AURORAL nodes that are used to share data and services. The presentation argued that the AURORAL ecosystem has the potential to create a more sustainable and equitable future for smart communities and wider rural areas.

This presentation emphasized the need to go beyond basic semantic interoperability, focusing on a decentralized, privacy-aware ecosystem for smart rural communities developed in the AURORAL H2020 project. Here's a summary of the key points:

- **Beyond Semantic Interoperability:**
  - The speaker argued for moving beyond simple semantic interoperability to address broader interoperability challenges in decentralized environments.
- **AURORAL Project:**
  - The AURORAL project focuses on smart communities in rural areas with a fully decentralized scenario.
  - It emphasizes user control over data access and discovery.
- **Decentralized Scenario:**
  - The project addresses interoperability among numerous actors in a decentralized setting.
  - It involves "vertical interoperability" (homogenizing formats, APIs, and data models).
  - A network of ontologies (based on those from standards such as W3C Web of Things and ODRL) is used to homogenize data models across diverse use cases.
- **Auroral Nodes:**
  - Auroral nodes are crucial components that handle data normalization, discovery, validation, and privacy policy enforcement.
  - They enable federated data discovery across the network.
  - They use ODRL to define and enforce privacy policies in real-time.
- **Horizontal Services:**
  - Horizontal services, built on top of the AURORAL nodes, provide added value to the data.

    - These include domain-specific services (e.g., demand response for energy) and cross-domain services (e.g., knowledge graph builder, AI service).
    - The AURORAL conformance node checks that all the nodes in the ecosystem satisfy the AURORAL interoperability profile.
- **Auroral Interoperability Profile:**
    - The interoperability profile defines the set of applications that every node must satisfy.
    - It monitors interoperability according to different facets (transport, syntax, semantics, behaviour, policy).
    - It covers different tasks involved in interoperability (discovery, access, understanding, validation).
- **Project Outcomes:**
    - The project created a privacy-aware ecosystem based on standards.
    - It enabled fine-grained and federated data discovery.
    - It supported data validation and privacy policy enforcement.
    - It created a conformance node to check for interoperability.
- **Resources:**
    - The project provides ontologies, GitHub repositories, and an AURORAL Academy course for learning how to use the middleware.

The following discussion centred on the implications and future directions stemming from the Auroral project presentation, particularly regarding the practical implementation and expansion of interoperability. Here's a summary of the key points:

- **Project as a Playground for Reality:**
    - The AURORAL project was recognized as a valuable "playground" for demonstrating desired interoperability features in a realistic, though controlled, setting.
    - The goal is to move beyond the project's scope and generalize its findings for broader application.
- **Addressing the "Not Sufficient" Claim:**
    - The core question was how to translate the project's insights into real-world, widespread interoperability.
    - The speaker emphasized the need to "engineer interoperability" by:
        - Identifying interoperability points (internal, external, transversal).
        - Defining interoperability profiles tailored to specific needs.
        - Ensuring compliance with these profiles.
- **Identifying Gaps and Future Challenges:**
    - The project revealed existing gaps in interoperability solutions, particularly in areas like:
        - Discovering privacy policies.
        - Other areas needing further research and industrial development.
    - The project was praised for showing the stable elements of interoperability, and also the gaps in the current state of interoperability.
- **Future Work:**
    - The discussion made it clear that the project has provided a great deal of material to work from, for future interoperability projects.

Talk: Semantic Technologies to Enhance Interoperability Between Multi-Energy Networks (Claudio Sesto, RSE S.p.A.) [slides]

The presentation discussed the use of semantic technologies to enhance interoperability between multi-energy networks. The presentation also discussed the IEC CIM standard for the electricity sector and how it can be used to model data for other energy provision, such as natural gas, district heating, and hydrogen. The presentation highlighted the benefits of using semantic technologies for multi-energy networks, such as improved data sharing, increased innovation, and reduced costs. The presentation also discussed a number of use cases where semantic technologies are being used for multi-energy networks, such as in the areas of smart grids, energy efficiency, and demand response. The presentation concluded with a discussion of the future of semantic technologies for multi-energy networks. The presentation argued that semantic technologies have the potential to transform the way that energy is generated, distributed, and consumed.

The presentation also discussed the MESP-DT (Multi-Energy Semantic Platform Digital Twin). MESP-DT is a semantic, modular, open-source platform based on international standards to manage and study multi-energy networks. The presentation highlighted the benefits of using MESP-DT, such as improved interoperability, reduced development costs, and increased innovation. The presentation also discussed a number of features of MESP-DT, such as its ability to integrate data from different sources, its support for different kinds of energy provision, and its ability to perform geospatial analysis. The presentation concluded with a discussion of the next steps for MESP-DT. These steps include developing CIM-derived ontologies to model data for other energy vectors, such as natural gas, district heating, and hydrogen.

This presentation focused on leveraging the Common Information Model (CIM) to enhance interoperability within the energy sector, expanding its application beyond electricity to include other energy vectors. Here's a summary of the key points:

- **Standardized Communication:**
    - The increasing number of devices and actors necessitates standardized communication protocols.
    - The CIM provides a systematic representation of power system components, facilitating data exchange.
    - RDF/XML serialization enhances machine readability and simplifies software development.
- **Expanding CIM Beyond Electricity:**
    - The presentation described the development of ontologies for natural gas and district heating distribution networks, based on the CIM's electrical ontology structure.
    - This demonstrates the CIM's adaptability to other energy sectors.
    - The benefits include consistent data management and integrated analysis.
- **Life Cycle Asset Management:**
    - Combining the CIM with Building Information Modelling (BIM) enables information exchange across design and operational phases.
    - This integration utilizes the IFC format.
- **Real-Time Data Management:**
    - Streaming data is exposed as RDF data through a virtual knowledge graph.
    - This data is enriched using static knowledge graphs and transaction history for versioning and storage.
    - This approach enables the tracking of knowledge graph configurations over time.
- **Geospatial Extension:**

    - GeoSPARQL is used to extend the ontology with geospatial information.
    - Geographic locations of geo-feature subclasses are modelled.
    - This enables geospatial visualization and analysis.
- **Multi-Energy Semantic Platform:**
    - All presented use cases are implemented within an open-source multi-energy semantic platform.
- **Conclusions:**
    - The CIM effectively addresses interoperability in the electricity sector.
    - Developing new ontologies extends the CIM to other energy vectors.
    - Harmonizing the CIM with other standards enriches the ontology with new information.
    - The plan is to add support for other standards in the future.

The following discussion focused on the practical applications and future directions of the presented work, particularly regarding IFC-CIM integration and its relevance to ongoing projects. Here's a summary of the key points:

- **IFC-CIM Integration:**
    - There was significant interest in the integration of IFC (Industry Foundation Classes) and CIM (Common Information Model).
    - A participant mentioned a project involving the conversion of IFC substation models to CIM, expressing a desire for collaboration and knowledge sharing.
    - The importance of combining energy and building information models was highlighted as a crucial future challenge.
    - The participants agreed to collaborate on this topic.
- **SAREF Integration:**
    - A question was raised about the relationship between the presented work and SAREF (Smart Appliances REFerence ontology).
    - The presenter acknowledged that SAREF integration is considered future work.
    - The presenter offered to follow up via email with a more detailed answer.
- **Symphony Software:**
    - The graphical representation on the right-hand side of a slide was identified as being generated by the Symphony software, developed by Open Grid.
    - The knowledge graph developed using the CIM standard was uploaded to the Symphony platform for visualization.

Talk: Design of the TIPS4PED Project Semantic Interoperability Layer (Anna Geraci, LINKS Foundation) [slides]

The presentation discussed the TIPS4PED project, which is focused on designing, developing, and testing a Digital Twin-based platform to support municipalities in the implementation of Positive Energy Districts (PEDs). The presentation also discussed the Semantic Interoperability Layer (SIL), which is a core component of the TIPS4PED platform. The SIL is responsible for ensuring interoperability across data sources, modules, and interfaces. The presentation highlighted the benefits of using the TIPS4PED platform, such as improved data sharing, increased innovation, and reduced costs. The presentation also discussed a number of use cases where the TIPS4PED platform is being used, such as in the areas of energy efficiency, demand response, and renewable energy integration. The presentation concluded with a discussion of the future of the TIPS4PED project. The presentation argued that the TIPS4PED platform has the potential to transform the way that municipalities plan and implement PEDs.

The presentation also discussed the TIPS4PED platform architecture. The TIPS4PED platform architecture is a modular architecture that comprises a number of different modules, such as a data ingestion module, a data processing module, a data storage module, and a data visualization module. The presentation highlighted the benefits of using a modular architecture, such as improved flexibility, increased scalability, and reduced complexity. The presentation further discussed a number of features of the TIPS4PED platform, such as its support for different data formats, its ability to integrate data from different sources, and its ability to perform semantic reasoning. The presentation suggested further steps for the TIPS4PED project. These steps include developing new modules for the platform, improving the performance of the platform, and deploying the platform in a number of pilot sites.

This presentation focused on a semantic interoperability layer developed within the Tips4PED project, designed to facilitate data exchange and integration for energy optimization in smart cities. Here's a summary of the key points:

- **TIPS4PED Project:**
    - The project aims to design, develop, and test a digital twin investment platform for positive energy districts (PEDs).
    - It was implemented in Turin and will be replicated in Cork.
    - The platform includes 13 modules for energy optimization and management.
- **Semantic Interoperability Layer:**
    - This component acts as a data exchange enabler, retrieving data from various sources and distributing it to platform services.
    - It merges static data (building specifications, energy statistics) and dynamic data (IoT sensors, monitoring systems).
    - It addresses the challenge of handling data in diverse formats.
- **Data Format Conversion and Brick Ontology:**
    - The layer performs data format conversion, transforming IoT sensor data into RDF-serialized format.
    - All data is integrated into a single semantic model compliant with the Brick ontology.
    - Brick ontology was chosen because it was already in use within the project's digital twin, and because it is well suited to building applications.
    - The brick ontology is used as a hub for the data.
- **Data Distribution and Protocol Support:**
    - A broker facilitates data exchange with platform services using various protocols (MQTT, AMQP, HTTP).
    - The platform supports multiple communication protocols.
    - While Brick is promoted, services are not strictly mandated to use it.
- **Flexibility and Reusability:**
    - The component allows data providers and service developers to maintain their preferred data formats and communication protocols.
    - The layer converts these inputs into a unified Brick model for data reusability and understanding across services.
    - The goal is to allow data to be reusable and understood across all the services of the platform.
- **Links Foundation:**
    - The Links Foundation is a research company active in energy systems, smart cities, and industry.
    - The department that presented the information is active in data management systems, integrations, and semantic interoperability.

The following discussion focused on the technical details and functionality of the semantic interoperability layer and its interaction with other components within the TIPS4PED project. Here's a summary of the key points:

- **JSON to RDF and RDF to JSON Conversion:**
  - The presentation mentioned JSON to RDF and RDF to JSON conversion.
  - The team is using a directive programming Python library for this conversion.
  - They intend to integrate these conversion functions into data flows using an open-source tool.
  - A need was identified for tools that can perform RDF to CSV conversions, to allow for the return of data to CSV format after changes to the knowledge graph.
- **Data Storage:**
  - The system utilizes a centralized database to store data compliant with the Brick ontology.
  - SPARQL is used to query and retrieve data, which can then be converted to desired formats and protocols.
- **Digital Twin Viewer:**
  - The digital twin viewer is a separate component developed by a project partner.
  - It visualizes the district being modelled, including various buildings and energy-related data.
  - It allows users to monitor parameters like temperature, humidity, and energy production.
  - It also enables sending commands to actuators for actuation.
  - The digital twin viewer is mainly for visualisation and interaction.

Talk: Digital Twins for waters (Franco Crivello, CETaqua, Water Technology Center) [slides]

The presentation discussed the IDEATION project, which aims to prepare the development of the Digital Twin of the inland waters (rivers and estuaries, lakes, reservoirs, wetlands, groundwater, coastal, snow and ice, urban water). The presentation also discussed the concept of the Digital Twin of the inland waters, the reference scenarios, and the OpenKIWAS knowledgebase. The presentation highlighted the benefits of using the Digital Twin of the inland waters, such as improved water management, increased innovation, and reduced costs. The presentation also discussed a number of use cases where the Digital Twin of the inland waters is being used, such as in the areas of flood forecasting, drought management, and water quality monitoring, among others. The presentation concluded with a discussion of the future of the Digital Twin of the inland waters and its potential to transform the way that water is managed.

The presentation also discussed the IDEATION architecture design. Its design is based on a number of principles, such as shared data, data quality assurance, geolocated data, open API, real-time integration, security, modular and scalable architecture, user interfaces, multi-format support, workflow manager, collaborative development, and documentation and tutorials. Moreover, the architecture design also considered core components from EDITO-Infra and EDITO-Model Lab. The presentation further discussed a number of features that IDEATION architecture design should consider, such as its support for different data formats, its ability to integrate data from different sources, and its ability to be interoperable with further solutions, like those provided by the Digital Twin Ocean. The presentation concluded with a discussion of the next steps for the IDEATION project.

This presentation provided an overview of the IDEATION project, which aims to prepare for the development of a digital twin for inland waters, integrated and interoperable with the existing Digital Twin Ocean (DTO). Here's a summary of the key points:

- **IDEATION Project Goals:**
    - To lay the groundwork for a digital twin of inland waters (rivers, lakes, wetlands, groundwater, etc.).
    - To ensure interoperability with the Digital Twin Ocean (DTO).
    - The project is a Coordination and Support Action (CSA), focusing on planning and road-mapping, not direct development.
- **Project Approach:**
    - Multi-stakeholder forums across Europe to gather user needs and requirements.
    - Development of OpenKIWAS, a knowledge database, gathering policies, projects, open data, models and tools that will provide support to the inland waters digital twin final user.
    - Mapping users' needs to existing tools and technologies.
    - Developing a reference architecture and roadmaps for building the future digital twin.
- **Multi-Stakeholder Forums:**
    - Used to gather requirements, feedback and key points to take into account for developing the digital twin roadmap.
    - Helps to identify gaps in existing technologies and data.
- **OpenKIWAS Database:**
    - Aggregates information from funding streams (FP7, Horizon 2020, Horizon Europe), water management policies, and scientific papers.
    - Uses AI and keyword analysis to filter and select relevant information.
    - Validates findings against existing water-related project lists (ICT4Water).
- **Architecture Design:**
    - Draws core components from the DTO's infrastructure and modes.
    - Emphasizes interoperability of datasets and auxiliary data sources.
- **Project Principles:**
    - Semantic interoperability and shared data.
    - Data quality assurance.
    - FAIR data principles (Findable, Accessible, Interoperable, Reusable).
    - Security and data access control.
    - Modular and scalable architecture.
- **Project Timeline:**
    - The project has a duration of two years, and at the time of the presentation was in month 9.
    - The project is still open to input, and requests feedback from the audience.

The following discussion highlighted the importance of the IDEATION project's work on principles for digital twins, and emphasized the need for collaboration and standardization in this area:

- **Standardization Efforts:**
    - It was mentioned that there is an ongoing standardization effort related to this area.
    - The speaker, who is an editor of the standard, encouraged collaboration and input.
- **Importance of Principles:**

    - The project's focus on principles for digital twins was recognized as significant.
    - It was noted that while there are principles for data spaces, there is a lack of established principles for digital twins.
- **Call for Collaboration:**
    - The speaker urged the IDEATION project team to continue working on the principles and to share them with the wider community.
    - There was a call for collaboration to reach a consensus on these principles.
- **Interoperability as a Key Focus:**
    - Interoperability was emphasized as a crucial aspect of the principles, and of digital twins in general.

Talk: Ontologies for digital twin: JUNON project use case (Fatma Chamekh, BGRM - French geological survey)

(slides unavailable)

This presentation outlined an ongoing project within the Juno research program, focusing on developing digital services for environmental monitoring and natural resource management. Here's a summary of the key points:

- **Juno Research Program:**
    - Aims to provide digital services for improved environmental understanding and natural resource management.
    - Funded by the region.
- **Digital Twin Architecture:**
    - Employs a four-part digital twin architecture: water quantity, water quality, air, and soil.
    - Each digital twin comprises several functional object models.
    - These digital twins share various data sources.
    - Use cases are designed that may utilize one or more of the digital twins.
- **Interoperability Challenges:**
    - The need for interoperability is crucial due to shared data sources and vocabularies among the digital twins.
    - Different digital twins require descriptions of shared data sources.
    - Shared vocabularies exist between various digital twin objectives.
- **Data Set Description:**
    - The project proposes describing each data set using metadata.
    - SOSA/SSN vocabulary is used for sensor data.
    - DCAT vocabulary is used for other data sources.
- **Communication Protocol:**
    - The project addresses the importance of communication protocols for message organization among architecture components.
    - Data story telling is being explored to organize messages.
    - Domain ontologies are used to standardize vocabulary and messages.
- **Linking Metadata and Ontologies:**
    - Ontological properties and concepts are linked to metadata elements.
    - Ontological concepts are linked to terminology and metadata element values.
    - This model is proposed to connect terminology to data set descriptions.
- **Proof of Concepts:**
    - A first proof of concept has been conducted.

    - A second proof of concept is underway, incorporating the communication protocol aspect.
- **French Geological Survey:**
    - The presentation was given by a representative of the French Geological survey.

The following discussion revolved around the concept of "data storytelling," its potential connection to data provenance, and broader reflections on the day's presentations, including the need for future collaboration and standardization. Here's a breakdown of the key points:

Data Storytelling:

- **Definition:** Initially described as a technique to organize the message between the conceptualization of an architecture and the visualization of results.
- **Importance:** Recognized as crucial for communication between different components of a system and for explaining the "why" behind data and engineering efforts.
- **Connection to Data Provenance:** The idea was raised that data provenance could be a form of data storytelling, focusing on the history and authenticity of data. However, it was noted that provenance typically describes the past, while storytelling can also encompass future implications and context.
- **Potential for Standardization:** A suggestion was made to consider developing a standard for data storytelling, particularly within the environmental digital twin community, which is actively exploring this concept for communicating climate change information to various stakeholders.
- **Benefits:** Storytelling can provide engineering context, motivation, and justification for data needs and quality requirements, especially in regulated industries.

Data Provenance:

- **W3C PROV:** Mentioned as an existing W3C vocabulary for expressing data provenance, initiated by NASA.
- **Verifiable Credentials:** Linked to data provenance as a way to secure assertions and ensure the authenticity of the "story" of the data.
- **Compliance Logs:** The combination of provenance and verifiable credentials was suggested as a potential tool for automating compliance and reducing bureaucratic burdens.
- **Ongoing Standardization:** An ISO project (SC 27 WG 4) on data provenance is currently in the Committee Draft (CD) stage, inviting participation.
- **Graph Manipulation:** A connection was made to graph manipulation methodologies as a formal way to describe data provenance.

Reflections on the Day and Future Directions:

- **Time Constraints:** A feeling that there wasn't enough time for in-depth understanding of the presentations.
- **Need for Cadence:** A suggestion to hold more regular (annual, semi-annual, or quarterly) events to allow for deeper engagement.
- **Exploratory Nature:** The day was seen as more exploratory, showcasing various initiatives without a focused problem-solving discussion.
- **Ontology Library Organization:** The idea of needing upper ontologies within vertical domains to better organize and discover sub-ontologies was raised.

- **Implementation Challenges:** The importance of discussing the implementation of approaches, including the need for agreement from Standards Developing Organizations (SDOs) and coordination between committees, was highlighted.
- **Extending and Reusing Ontologies:** The challenges of extending existing ontologies without creating silos and the need for clear documentation on extension methodologies were discussed, using the IDMP example.
- **Need for Extension Storytelling:** The idea of having a clear "extension storytelling" to guide how ontologies can be adapted was proposed.

Overall Takeaway:

The discussion highlighted the evolving concept of "data storytelling" as a crucial element for communicating the value and context of data, potentially complementing data provenance efforts. There was a strong consensus on the need for continued collaboration, standardization, and deeper discussions to move beyond simply showcasing individual initiatives towards building a more cohesive and interoperable semantic web ecosystem. The challenges of ontology reuse, extension, and the coordination of standardization efforts were also emphasized as key areas for future work.

Day 2

Session 5: Semantic Interoperability in Relation to Generative AI

Keynote: Using generative AI to support mapping data to ontologies (Martin Bauer, NEC Laboratories Europe) [slides]

The presentation discussed the use of generative AI to support data mapping and transformation. Generative AI can be used to find a suitable data model or ontology for a given data set, map data to a specific data model or ontology, and transform data from one data model or ontology to another. The presentation also discussed the use of generative AI to map data models and ontologies to each other, map concepts, and transform instances from one data model or ontology to another. The presentation highlighted the benefits of using generative AI for data mapping and transformation, such as improved accuracy, reduced costs, and increased speed. The presentation also discussed a number of challenges that need to be addressed in order to use generative AI for data mapping and transformation effectively, such as the need for high-quality training data and the need to ensure that the generated mappings are accurate and consistent. The presentation concluded with a discussion of the future of generative AI for data mapping and transformation. The presentation argued that generative AI has the potential to revolutionize the way that data is mapped and transformed.

The presentation further discussed the use of generative AI to support instance and entity mapping. Instance and entity mapping is the process of mapping data instances, such as entities, to each other. This is a challenging task, as data instances can be represented in different ways in different data sets. The presentation argued that generative AI can be used to address this challenge by learning how to map data instances from different data sets to a common representation. The presentation discussed a number of benefits of using generative AI for instance and entity mapping, such as improved data integration, increased data sharing, and reduced data redundancy. The presentation concluded with a discussion of the future of generative AI for instance and entity mapping. The presentation argued that generative AI has the potential to make it easier to integrate data from different sources and to create a more unified view of data.

The discussion focused on leveraging Generative AI, particularly Large Language Models (LLMs), to enhance semantic interoperability, with Martin providing an overview and sharing NEC's work in this area.

Key Points:

- **Semantic Interoperability as a Bridge:** Semantic interoperability is crucial for overcoming data silos.
- **Generative AI's Potential:** Generative AI can produce diverse media, including structured data and code, making it a potential tool for semantic tasks.
- **Types of Mapping:** The discussion touched upon different types of mapping relevant to semantic interoperability: data to ontology, data and ontologies to each other, and data instance to data instance.
- **ITU-T Report:** Mentioned the existence of an ITU-T report exploring the implications of Generative AI for semantic interoperability.
- **NEC's Work (Data Transformation with LLMs):**
    - **Goal:** To use LLMs to automatically generate Python code for transforming data (e.g., from CSV) into a target schema.
    - **Process:**
        - Starts with source data and a desired target schema.
        - LLM helps select a suitable target data model.
        - LLM generates Python code for the transformation.

        - Incorporates human feedback and automatic testing to improve the generated code.
    - **User Involvement:** Users can provide additional information to refine data handling in the generated code when the AI's output isn't perfect.
    - **Current Focus:** Primarily on structured data.
- **Limitations and Future Directions (Raised by Questions):**
    - **Unstructured Data:** The current demos focus on structured data; handling unstructured data is a future challenge.
    - **Ontology to Ontology Mapping:** This is not yet a focus of NEC's work.
    - **Retraining the Model:** NEC has not yet explored retraining the LLM for this specific task.
- **Impact of Accurate Information:** Providing more accurate information to the AI improves its performance in generating mapping code.
- **Declarative Mappings vs. Code Generation (Pierre-Antoine Champin):** Generating declarative mapping specifications would be preferable to code generation in many use cases.
- **Relation to Graph Mapping Algorithms:** The work has connections to existing graph mapping algorithms.
- **Hard Cases for Mappings (Dave):** The discussion acknowledged that mapping becomes difficult when the source and target schemas or ontologies are not well-aligned.
- **Mapping from Ontology with AI:** There was a question about directly mapping from an ontology using AI.
- **Temporal Variability of Answers:** A concern was raised about whether the LLM's output could vary depending on the time of day, but this was not tested.
- **Generative Mapping for Testability:** The idea of generating declarative mappings that are easier to test was suggested as an alternative to code generation.

In essence, the discussion explored the promising role of LLMs in automating data transformation for semantic interoperability, particularly in generating code based on target schemas. However, it also highlighted current limitations (unstructured data, ontology mapping) and raised important considerations about the nature of the generated mappings (declarative vs. code) and the need for robust testing and validation.

Long Talk: Reuse of ontologies and impact of generative AI (Henon Mengistu Lamboro, Institut Polytechnique de Paris, and Trialog) [slides]

The presentation discussed the challenges and opportunities of reusing ontologies in the age of generative AI. Ontology reuse is the process of selecting, adapting, and integrating existing ontologies into new knowledge systems. This can be a complex and time-consuming process, but it can also lead to significant benefits, such as improved interoperability and reduced development costs. The presentation also discussed the impact of generative AI on ontology reuse. Generative AI can be used to automate some of the tasks involved in ontology reuse, such as finding and evaluating ontologies. However, there are also some challenges associated with using generative AI for ontology reuse, such as the need to ensure that the generated ontologies are accurate and consistent. The presentation concluded with a discussion of the future of ontology reuse in the age of generative AI. The presentation argued that generative AI has the potential to revolutionize the way that ontologies are reused. However, it is important to address the challenges associated with using generative AI for ontology reuse in order to realize its full potential.

The presentation further discussed a number of lessons learned from EU projects on ontology reuse. These lessons include the importance of having a clear understanding of the domain for which the ontology is being developed, the need to carefully evaluate existing ontologies before reusing them, and the importance of using a modular approach to ontology development. The presentation also discussed a number of best practices for ontology reuse, such as documenting the ontology reuse process and using a version control system. The presentation concluded with a discussion of the future of ontology reuse. The presentation argued that ontology reuse is an essential part of building interoperable systems.

The discussion covered various aspects of ontology reuse, modularity, and the potential of Generative AI in ontology engineering, along with mentions of specific projects and resources.

Key Points on Ontology Reuse and Modularity:

- **Mathematical Description of Reuse:** Ontology reuse can be mathematically described through operations like "as is," extraction, extension, and combination.
- **Modularity Levels:** Ontologies can be modularized based on granularity (foundational, domain, application/task ontologies, and hybrids).

Generative AI for Ontology Engineering:

- **Potential Applications:** Generative AI can assist in:
    - Creating ontologies from text.
    - Semantic mapping across ontologies.
    - Ontology validation and error detection.
    - Competency question generation.
- **Current Limitations (Henon's Claim):** No single existing Generative AI system currently supports all these steps comprehensively.
- **Ontology Generation from Text:** Tools like DeepSeek and ChatGPT can generate ontologies from textual input, but ensuring the quality of these generated ontologies requires significant effort from human experts.

Specific Projects and Standards:

- **WoT and SAREF:** These standards can be integrated, with the service area/definition of SAREF acting as a point of connection.
- **CIM (Common Information Model IEC):** CIM has adopted RDF for semantic interoperability, enabling cross-domain data exchange. Its use by DSOs can be improved through CIM profiles tailored for specific applications like energy consumption.
- **Maria Poveda's Tool:** A tool exists to transform diagrams created in draw.io into OWL ontologies.
- **Semaptic:** This tool addresses the challenges of semantic interoperability for time series data, as discussed by Vladimir and Rigo.
- **NIST Privacy Workforce Taxonomy:** A relevant resource for privacy-related concepts.
- **IoT Catalogue:** This platform publishes auroral use cases, which could be valuable for understanding real-world applications.

Call for Continued Communication and Collaboration:

- **Need for Communication Continuation:** There's a recognized need for ongoing communication among the workshop participants, potentially through a Slack group or similar platform.
- **Contribution to Standards:** A suggestion was made to contribute all use cases discussed during the meeting to ISO TC 42 (Artificial Intelligence).

Long Talk: LLM Toolchain for Open Data and NGSI-LD (Danilo Pillitteri, Parco Scientifico e Tecnologico della Sicilia (PSTS)) [slides]

The presentation discussed the Data-Highway project, which is a Linked Open Data platform for the development of innovative applications for tourism and cultural promotion. The presentation also discussed the use of Large Language Models (LLMs) for data analysis and the use of NGSI-LD for data interoperability. The presentation highlighted the benefits of using LLMs and NGSI-LD for data management, such as improved data quality, enhanced entity extraction, and increased automation. The presentation also discussed a number of challenges that need to be addressed in order to use LLMs and NGSI-LD effectively, such as the need for high-quality training data and the need to ensure that the generated data is accurate and consistent. The presentation concluded with a discussion of the future of LLMs and NGSI-LD for data management. The presentation argued that these technologies have the potential to revolutionize the way that data is managed.

The presentation further discussed a number of specific applications of LLMs and NGSI-LD for data management. These applications include the use of LLMs to generate sentences that can be used to improve entity extraction, the use of LLMs to structure relationships between entities, and the use of NGSI-LD to manage entities and relationships. The presentation discussed a number of benefits of using LLMs and NGSI-LD for data management, such as improved data quality, enhanced entity extraction, and increased automation. The presentation concluded with a discussion of the future of LLMs and NGSI-LD for data management. The presentation argued that these technologies have the potential to make it easier to create, access, share, and reuse data.

Key Points on Data Harmonization using LLMs:

- **Data-Highway Project:** Collects data from portals and aims to map it to NGSI-LD using LLMs.
- **Two-Step LLM Approach:**
    - **Graph Generation:** An LLM is used to generate a graph from the text, representing entities and relationships. The prompt defines the specific details required for this graph.
    - **Model and Prompt Creation:** A model and a prompt are created for the next stage of mapping to NGSI-LD.
- **Focus on Temporal Information:** The project specifically targets temporal information related to events in the source documents.
- **LLM Effectiveness:** LLMs are generally effective at extracting named entities and relationships from text.
- **LLM Comparison:** Different LLMs were tested, with varying results:
    - **GPT-4o:** Performed best overall.
    - **DeepSeek:** Good at conforming to the NGSI-LD schema.
    - **Llama:** Fast but less accurate.
    - **Phi-4:** Best of the free models tested.
- **Graph-Based Approach:** The graph-based approach for initial representation seems to be more effective.

- **Alternative Approach:** A second approach yielded more specific entity definitions but at the cost of increased complexity and potential information loss.
- **Semantic vs. Syntactic Harmonization:** Semantic data harmonization aims for universal understanding without ambiguity (e.g., including units of measure), unlike syntactic harmonization.
- **Labour-Intensive Process:** The traditional 5-stage semantic harmonization process is labour-intensive, hindering adoption.

Key Points on Entity Matching:

- **Entity Matching Problem:** Determining if two entities are the same is crucial for duplicate removal and effective querying.
- **LLM Techniques Explored:** The project has investigated fine-tuning of LLMs and few-shot learners for entity matching. This combines prompt engineering with retrieval augmented generation.
- **Challenges with LLMs:** Numerous challenges exist when using LLMs for this task.
- **Comparison of Techniques:** Several entity matching techniques were compared: string-based, vector-based, prompt engineering, and fine-tuning.
- **Cost of Training Data:** Creating training examples for fine-tuning is costly, necessitating learning from a limited number of examples (few-shot learning).

Discussion Points and Questions:

- **Reasoning for Ontology Alignment:** A question was raised about using reasoning techniques for ontology alignment (in the context of the first talk), and the response was that it hasn't been tried but is worth exploring.
- **Role of Location in Entity Matching:** Francisco questioned whether building location helps in determining if entities are the same. It was noted that location can be misleading (e.g., car rental agencies at airports).
- **Prompt Tuning for Context:** Tuning the LLM prompt can help to account for misleading information like location, but relying on "common sense" with LLMs is unreliable.

Long Talk: Role of LLMs for data integration and semantic data interoperability, focusing on entity matching (Francisco Martin-Recuerda, SINTEF AS) [slides]

The presentation discussed the use of Large Language Models (LLMs) for data integration and semantic data interoperability. The presentation also discussed the challenges and opportunities of using LLMs for data integration and semantic data interoperability. The presentation highlighted the benefits of using LLMs for data integration and semantic data interoperability, such as improved accuracy, reduced costs, and increased speed. The presentation also discussed a number of challenges that need to be addressed in order to use LLMs for data integration and semantic data interoperability effectively, such as the need for high-quality training data and the need to ensure that the generated data is accurate and consistent. The presentation concluded with a discussion of the future of LLMs for data integration and semantic data interoperability. The presentation argued that LLMs have the potential to revolutionize the way that data is integrated and made interoperable.

The presentation further discussed a number of specific applications of LLMs for data integration and semantic data interoperability. These applications include the use of LLMs to map data to ontologies, the use of LLMs to transform data from one data model or ontology to another, and the use of LLMs to generate sentences that can be used to improve entity extraction. The presentation discussed a number of benefits of using LLMs for data integration and semantic data interoperability, such as improved data quality, enhanced entity extraction, and

increased automation. The presentation concluded with a discussion of the future of LLMs for data integration and semantic data interoperability. The presentation argued that LLMs have the potential to make it easier to create, access, share, and reuse data.

Session 6: Best Practices for Vocabulary Development

Keynote: SAREF Suite of Ontologies (Cornelis Bouter) [slides]

The presentation was titled "SAREF at the Intersection of Academia, Industry, Standardisation & Government." The presentation discussed the Smart Applications Reference Ontology (SAREF), which is a suite of ontologies that can be used to model the knowledge of different domains. The presentation also discussed the SAREF European Norm (EN), which is a standard that specifies a process for making IoT smart applications and products interoperable in compliance to the SAREF framework. The presentation highlighted the benefits of using SAREF, such as improved interoperability, reduced development costs, and increased innovation. The presentation discussed a number of use cases where SAREF is being used, such as in the areas of energy, environment, and smart cities. The presentation concluded with a discussion of the future of SAREF. The presentation argued that SAREF has the potential to become a de facto standard for semantic interoperability in the IoT domain.

The presentation also discussed the different levels of interoperability, such as technical interoperability, syntactic interoperability, and semantic interoperability. The presentation argued that semantic interoperability is the most important level of interoperability, as it enables different systems to exchange data with unambiguous, shared meaning. The presentation further discussed the role of ontologies in semantic interoperability. Ontologies are formal representations of knowledge that can be used to define the meaning of data. The presentation highlighted the benefits of using ontologies for semantic interoperability, such as improved data sharing, increased innovation, and reduced costs. The presentation concluded with a discussion of the future of semantic interoperability. The presentation argued that semantic interoperability is essential for the development of AI-powered applications and services.

This talk provided a comprehensive overview of the SAREF (Smart Applications REFerence Ontology), its history, principles, and its connections with academia, industry standardization, and government initiatives. Here's a summary of the key points:

- **Semantic Interoperability as a Means to an End:** SAREF aims to achieve semantic interoperability (mutual understanding between systems) to facilitate organizational interoperability (achieving common goals, e.g., in digital twins).
- **SAREF Definition and Scope:** SAREF is a formal ontology describing the Internet of Things domain, focusing on devices, their measurements of objects, tasks, and procedures. It provides an abstraction layer over specific standards.
- **Core Ontology and Reference Patterns:** SAREF offers a core ontology defining fundamental IoT concepts and a set of reference ontology patterns for guiding usage, reuse, and extensions for various vertical domains (energy, agriculture, water, etc.).
- **Fundamental Principles:** SAREF's development is based on principles of reuse and alignment, modularity, extensibility, and maintainability, following good design and modelling practices (managed in Git).
- **SAREF Model Overview:** The core model centres around devices and their measurements of objects, including services, functions, commands, and states. It is heavily based on and has overlap with the Semantic Sensor Network (SSN) ontology.
- **Resources and Access:** Documentation for the core and vertical extensions is available online (labs.etsi.org). The GitLab environment hosts the ontology code (Turtle) and additional documentation.

- **History and Evolution:** SAREF originated from a European Commission request for a smart appliance standard and has evolved over the past decade to encompass broader smart applications and various domains. Version 4.1 is currently under development with further harmonization and functionalities.
- **Government and Industry Adoption:**
    - The European Commission's initial request spurred SAREF's development.
    - SAREF is used in the JRC Code of Conduct for Smart Energy Smart Appliances, where several manufacturers voluntarily commit to exposing device data using SAREF for interoperability in energy efficiency and flexibility use cases.
- **ETSI EN Standard for SAREF Extension:** A European Norm (EN) has been developed to prescribe the extension process of SAREF, offering guidance on reusing the core and creating new extensions for interoperable IoT applications.
- **SAREF Extension Process:** The process is based on the Linked Open Terms methodology, involving requirement analysis, data source and standard identification, alignment with existing models, model creation, and consideration of SAREF integration.
- **Options for Contributing and Aligning with SAREF:**
    - **New SAREF Extension:** Submitting a new domain model to ETSI to be recognized as an official SAREF extension (pros: visibility, strong involvement; cons: long process, high entry threshold).
    - **Merging with Existing Extension:** Integrating a new model with an existing SAREF extension if significant overlap exists (pros: faster, leverages existing applications; cons: requires alignment with existing design, prevents extensions from becoming too large).
    - **SAREF Compliance (Independent Model**): Aligning a compliance model with SAREF for interoperability but maintaining governance independently without formal submission to ETSI (pros: full governance control, faster development; cons: lacks formal standardization, requires ongoing alignment with SAREF updates).
- **Collaborative Effort:** The development of SAREF is a result of the work of various individuals over the past ten years. The core development team includes Laura, Cornelis, Raul, Maria, and Maxim.

The following discussion raised several important points regarding SAREF's structure, alignment with other initiatives, and practical application:

- **Semantic Interoperability Layering:** A suggestion was made to reconsider the layering in the initial slide, proposing that "semantic" should be a transversal attribute applied across technical, syntactic, organizational, behavioural, policy, and product layers, rather than being a distinct top layer itself. The presenter was open to discussing and refining this.
- **Extending Beyond Ontology:** The need to move beyond just defining the ontology (as in the Code of Conduct) to the next stages of taxonomy, phases, profiles, and integration with real-world practices was emphasized. The ETSI EN standard was seen as a step towards this, but further integration is needed.
- **Alignment with Other Ontologies:** Questions were raised about SAREF's relationship with other relevant ontologies like SOSA/SSN, QUDT, UMBEL, DOLCE, top-level ontologies, and domain-specific ontologies.

    - **SOSA/SSN:** SAREF aims for compatibility, not competition, with SOSA/SSN. There's ongoing work and personnel overlap to align the ontologies. A mapping between SAREF and SSN exists and will be included in the next SSN version.
    - **QUDT:** SAREF uses QUDT as a recommended vocabulary for units of measure, advocating for its reuse rather than creating its own.
    - **General Approach:** SAREF's strategy is to reuse existing well-established ontologies where appropriate, using placeholders in its model and recommending external vocabularies in its documentation.
- **Relationship with Web of Things (WoT) Thing Description:** SAREF has been developed with consideration for WoT. There's involvement in the WoT working group, and projects successfully mix both. WoT Thing Description is seen as a way to describe services, which can be linked to SAREF's service concept.
- **Alignment with SOSA/SSN (Revisited):** A direct question about whether SAREF is totally aligned with or just based on SOSA/SSN, and why it doesn't directly use the SSN namespace, was posed.
    - While there is significant overlap and collaboration, some differences exist, particularly in the service aspect.
    - The ongoing evolution of both SAREF and SSN, with cross-participation in working groups, suggests a move towards greater alignment.
- **Reusability vs. Compatibility:** A key point was made that the goal should be higher than just compatibility; it should be about reusability, which inherently includes compatibility.
- **Future Work:** The discussion highlighted the ongoing nature of ontology development and the need for continuous alignment and integration with other relevant initiatives and real-world applications.

In essence, the discussion acknowledged the value of SAREF while emphasizing the importance of its place within a broader ecosystem of ontologies and standards. The need for practical application, alignment with existing work, and a focus on reusability were key takeaways.

Keynote: Smart Data Models, NGSI-LD (Alberto Abella) [slides]

The presentation discussed the Smart Data Models initiative, which is an open-source project that provides data models for a variety of domains, such as smart cities, smart energy, and smart agriculture. The presentation also discussed the NGSI-LD standard, which is a standard for exchanging data between different systems. The presentation highlighted the benefits of using Smart Data Models and NGSI-LD, such as improved interoperability, reduced development costs, and increased innovation. The presentation also discussed a number of best practices for developing vocabularies, such as using a modular approach, reusing existing vocabularies, and documenting the vocabulary. The presentation concluded with a discussion of the future of Smart Data Models and NGSI-LD. The presentation argued that these technologies have the potential to transform the way that data is shared and used.

The presentation further discussed the governance of Smart Data Models. The Smart Data Models initiative is governed by a Steering Board, which comprises representatives from a number of different organizations. The presentation highlighted the benefits of having a strong governance model, such as improved decision-making, increased transparency, and reduced risk. The presentation also discussed a number of tools that are available to support the development of Smart Data Models, such as a self-testing tool, a draft tool, and a merging contexts tool. The presentation concluded with a discussion of the future of Smart Data Models. The presentation argued that Smart Data Models have the potential to become a de facto standard for data interoperability.

This talk presented the Smart Data Models initiative, emphasizing its close relationship with NGSI-LD and its pragmatic approach to creating and promoting the use of data models. Here's a summary of the key points:

- **Focus on Usage:** The primary goal of Smart Data Models is to create data models that are not only well-defined but actively used in real-world applications.
- **Open and Collaborative:** The initiative is led by the FIWARE Foundation, IUDX (Government of India), TM Forum, and OASC (Open & Agile Smart Cities). All data models are open licensed, allowing free use, customization, and sharing.
- **User-Driven Contribution:** Smart Data Models doesn't have a dedicated group for creating models. Instead, contributions come from users implementing solutions in various domains (logistics, robotics, transportation, etc.) who share their working models.
- **Large and Growing Repository:** There are currently around 1000 official data models and over 250 more in the queue, covering 75 subjects. Mappings exist for SAREF (v3) and the Common Information Model (CIM).
- **Technical Foundation:** The "single source of truth" is JSON Schema, chosen for its accessibility to non-data modelling experts. A Python package provides tools for working with the data models and their semantic information.
- **Multi-Format Support:** The JSON Schema is automatically exported to other formats like SQL, DDL, and potentially RDF, making adoption easier for various users.
- **Mandatory Examples and Automated Generation:** Every data model requires real-world examples, which are also exported in multiple formats. Specifications are automatically generated in eight languages (and more to come), with updates taking only minutes after approval.
- **Community and Adoption:** Over 120 organizations and 200 individuals have contributed. While the exact adoption rate is hard to track, over 330 documented adapters suggest significant usage.
- **Simplified Contribution Process:** The initiative provides tools on its website to lower the barrier for contributions, including self-testing, drafting based on examples, and automatic inclusion of existing attribute descriptions.
- **Mapping to Other Ontologies:** While creating unique URIs, Smart Data Models also provides tools for mapping to other ontologies like schema.org, facilitating integration with existing systems.
- **Enabling Real-Time Data Interchange:** The ultimate goal is to enable near real-time data exchange between real organizations using standardized, customizable, and real-world tested data models.
- **Agile and Pioneer-Led Approach:** The initiative prioritizes speed and real-world usage over lengthy consensus processes. It follows pioneers who demonstrate working solutions.
- **User-Led Development:** New data models are created based on user requests and demonstrated needs, not based on theoretical usefulness.
- **Key Best Practices Learned:**
    - Act quickly or don't do it (time to market is crucial).
    - Don't reinvent the wheel (reuse existing open standards).
    - Focus on normalized real use cases.
    - Help contributors generalize their models.
    - Data models are intentionally kept relatively flat (limited deep inheritance) based on user demand.

    - The initiative is self-funded and ongoing, not a finite project.
- **Lack of Conflicts in Model Approval:** Surprisingly, there have been few conflicts in approving new data models, possibly due to the pragmatic approach, easy extensibility, and the fact that the models are based on working solutions.
- **Mapping Imperfect Models:** The initiative acknowledges and accepts that it often maps "imperfect" real-world models with existing inconsistencies (e.g., same names for different concepts or attribute types) to ensure practical adoption. Unique URIs help disambiguate, but short names might differ across domains.
- **Emphasis on Time to Market:** The automated processes ensure that new or updated data models can be published and made available quickly.
- **Adoption:** Tracking the actual adoption rate remains a significant challenge despite efforts.
- **Lowering User Barriers:** The initiative actively works to make contribution and usage accessible to domain experts who may not be data modelling experts.

In conclusion, Smart Data Models offers a practical and agile approach to semantic interoperability by focusing on real-world usage, lowering barriers to contribution, and embracing a degree of imperfection to facilitate widespread adoption.

The discussion following Alberto's presentation on Smart Data Models focused on practical aspects of adoption, tooling, and alignment with more formal semantic approaches. Here's a summary of the key points:

- **Tooling for JSON Schema:**
    - A question was raised about whether custom code or libraries like LinkML were used to generate JSON schemas. The answer was that plain JSON Schema with descriptive text fields is used.
    - The simplicity of JSON Schema was highlighted as a key reason for its choice, making it easier for non-experts to contribute and validate data.
- **Addressing Non-Expert Contributors:**
    - The challenge of experts in domains (e.g., marine logistics) not being familiar with JSON Schema or YAML was acknowledged.
    - Existing tools are recommended, including an external online JSON Schema editor.
    - A "Data Model Editor" tool on the Smart Data Models website allows users to paste example data (e.g., JSON payloads) to automatically generate a corresponding JSON schema. This tool aims to lower the barrier for contribution.
    - A spreadsheet template is also provided as a more understandable interface for non-experts to draft data models.
- **Practicality vs. Formal Semantics (Ontologies):**
    - A participant praised the practical and somewhat "provocative" approach of Smart Data Models, focusing on immediate usability.
    - The idea of a continuum of practice was introduced, where practical data models can coexist and potentially align with more formal ontologies over time, especially when regulation and certification are involved.
    - A call for collaboration was made to explore how Smart Data Models could interact with and potentially derive from existing ontologies (e.g., for privacy, AI Act compliance) without disrupting its successful practical approach.

    - The European Commission was suggested as a potential facilitator for finding ways to ensure interaction, synchronization, or alignment between practical data models and more formal semantic standards.
- **Leveraging Existing Standards:**
    - Smart Data Models prioritizes reusing and mapping to existing open standards and vocabularies whenever possible (e.g., GTPS from the Open Connectivity Foundation).
    - A service for mapping to external ontologies is available on the Smart Data Models website, encouraging users to leverage vocabularies like schema.org or others as needed.
- **Addressing Immediate Needs:**
    - Smart Data Models aims to fill the gap of quickly providing data models (within weeks) for immediate project needs, even if they are extensions of existing ontologies.
- **Alignment with Formal Standardization Bodies:**
    - Collaboration with larger standardization bodies (like DDC EDTC 25) is welcomed to ensure alignment and prepare for future requirements.
- **Importance of Examples and Tooling:**
    - The discussion reinforced that simply publishing an ontology or vocabulary is insufficient. Providing examples, software tools, and lowering the barrier to adoption are crucial for practical use.
- **Spreadsheet as an Accessible Tool:**
    - The use of a spreadsheet for creating basic data models highlights the focus on making contributions accessible to individuals who may not have technical data modelling expertise.

In summary, the discussion centred on the practical success of Smart Data Models, the tools provided to ease adoption and contribution, and the potential for future collaboration and alignment with more formal semantic standards and ontologies to address evolving needs and regulations. The emphasis remained on usability and lowering the barriers for domain experts to participate in data standardization.

Keynote: CIM Model (Electricity/Energy), (Svein Olsen - CIM expert) [slides]

The presentation discussed the IEC Common Information Model (CIM), which is a standard for modelling and exchanging information about electrical power systems. The presentation also discussed the benefits of using CIM, such as improved data sharing, increased innovation, and reduced costs. The presentation highlighted the importance of semantic interoperability for digital twins in the power system domain. The presentation also discussed a number of challenges that need to be addressed in order to achieve semantic interoperability in this domain, such as the need for a common understanding of the data and the need for tools to support data exchange. The presentation concluded with a discussion of the future of CIM and semantic interoperability in the power system domain. The presentation argued that CIM has the potential to become a de facto standard for semantic interoperability in this domain.

The presentation further discussed a number of specific applications of CIM. These applications include the use of CIM for power system planning, power system operations, and power system maintenance. The presentation discussed a number of benefits of using CIM, such as improved data quality, enhanced data analysis, and increased automation. The presentation concluded with a discussion of the future of CIM. The presentation argued that CIM has the potential to make it easier to create, access, share, and reuse data in the power system domain.

This talk provided a detailed overview of the Common Information Model (CIM) within the power system domain, its history, challenges, and future directions towards greater semantic interoperability. Here's a summary of the key points:

- **CIM Usage and Challenges:** CIM is widely used in the power industry, which is a benefit but also creates challenges regarding backward compatibility when updates are needed.
- **Speaker's Background:** The speaker is a computer system engineer with extensive experience in energy domain applications and is involved in IEC TC 57 working groups responsible for CIM.
- **CIM Scope:** CIM (IEC 61970, 61968, 62325 series) provides a common information model for power systems, covering energy management systems (EMS), distribution management systems (DMS), asset management, system analysis, and market perspectives. The lines between transmission and distribution are blurring, leading to a broader scope.
- **Cross-Domain Interoperability:** A key focus is achieving interoperability across different domains and building a framework for digital twins that are not monolithic but can collaborate.
- **CIM and Semantic Web History:** CIM was an early adopter of RDF/XML around 2000, which was progressive at the time. A more human-readable dialect led to N3 and Turtle. While graph-based modelling exists, full linked data mainstream adoption is still a goal.
- **Modernization Efforts:** The current focus is on modernizing CIM using JSON-LD, SHACL for validation, and federated graphs to improve usability and leverage current semantic web technologies. Balancing semantic interoperability with object-oriented toolchains is a consideration.
- **Architectural Alignment:** Efforts have been made to align CIM with architectural frameworks like ArchiMate, defining business functions and objects to better integrate with IT systems.
- **Modelling Timeframes:** CIM models various timeframes, from nanoseconds in power electronics to years in long-term planning, requiring equivalent and interoperable models for different simulation purposes.
- **CIM in Operation:** CIM underpins critical power system operations like state estimation (e.g., in Statnett, producing a grid state every 10 seconds based on thousands of observations), which helps identify model or data inaccuracies.
- **Drivers for Upgrade:** The need for more efficient grid operation, cross-country coordination, and the integration of renewables are driving the upgrade of CIM towards more advanced semantic web technologies.
- **Alignment with Semantic Web Standards:**
    - **JSON-LD & SHACL:** Being adopted for data representation and validation.
    - **DCAT:** Used for describing datasets and their relationships.
    - **PROV-O:** Considered for describing data provenance.
    - **DX Profile:** Utilized for describing profiles, important with many variations.
    - **QUDT:** Aiming for alignment on units of measure.
    - **GeoSPARQL:** Querying for geospatial data, despite some community disagreements.
    - **SSN (Semantic Sensor Network):** Seeking alignment to address web of things integration and core concept alignment for better interoperability.

    - **IDO (Industrial Data Ontology):** Investigating as a potential upper ontology to help understand different concepts (functional model, product model, assets, location).
    - **BIM (Building Information Modelling):** Exploring semantic ways to move between CIM and BIM (though not necessarily limited to IFC).

In conclusion, the power industry through CIM has a long history with semantic modelling. The current efforts are focused on modernizing CIM by adopting the latest semantic web standards to enhance interoperability, data sharing, and integration with other domains, ultimately supporting the evolution of digital twins in the energy sector.

This discussion following the CIM modernization talk raised several key points about implementation, standardization, and future directions:

- **Standardization Push:** A question arose about whether the CIM modernization is driven by individual efforts or a broader push from standardization bodies. The answer indicated a mix, with individual initiatives (like Statnett's) aligning with broader IEC trends, but progress can be slow due to the need for consensus and backward compatibility.
- **Alignment with Ontology Libraries (IDO Example):** The potential for CIM to align with established ontology library structures (like IDO with top-level, domain-independent, and domain-dependent ontologies) was discussed. While acknowledging the challenges of aligning existing terminology, IDO was seen as a potentially helpful resource, particularly for understanding different conceptual levels.
- **Importance of Profiles:** The concept of "profiles" of CIM was highlighted as crucial for practical implementation. These profiles, whether defined by individuals, organizations, or standardization bodies, provide a specific and usable subset of CIM. Publishing and promoting the use of such profiles, along with supporting tools, is seen as key to adoption.
- **SIM Model Version:** A question clarified that the modernization efforts are an ongoing process, not solely based on a specific older version of the CIM model.
- **CIM Query Language (WQL) and OWL Compatibility:** Concerns were raised about the CIM's query language (WQL) being UML-based and potentially incompatible with OWL. The response clarified that IEC 61970-501 describes how UML is serialized as RDFS for CIM, including for data systems (like the asset catalogue in IEC 61970-13). While acknowledging the historical networking focus, ongoing work aims to extend this to data assets, which is seen as a significant and challenging but necessary step.
- **Graph-Based Approach vs. Web of Things:** A clarification was sought on whether a graph-based approach in CIM is being replaced by Web of Things (WoT). The answer was no; the aim is to harmonize. The existing CIM data model for SCADA acquisition (dating back to the 80s) is not ideal for IoT. Future updates aim to align CIM with WoT principles and standards (potentially including IEEE approaches and aligning with IEC 61850) to better represent IoT devices and data, while still leveraging the underlying graph database structure of CIM.
- **Flexibility Use Cases (IEC 62746):** The upcoming IEC 62746 standard to support flexibility-related use cases (aligned with European regulations and demand response) was mentioned as a significant development.
- **Interoperability Testing:** An upcoming interoperability test in early March among energy data spaces and projects (like AD) will leverage a simple CIM profile for customer data

access, using both CIM profiles and ontology syntax with the support of the EU-SysFlex project.

- **Extending CIM to Distribution System Operators (DSOs):** The importance and challenge of extending CIM usage and ontologies to the numerous DSOs in Europe (far more than the transmission system operators) was emphasized, highlighting the need for tooling and support tailored to their needs.

In summary, the discussion underscored the ongoing evolution of CIM towards greater semantic interoperability by embracing modern semantic web technologies and aligning with other relevant standards and initiatives like WoT. The importance of practical CIM profiles, tooling for adoption, and addressing the specific needs of both transmission and distribution system operators were key themes. The challenges of backward compatibility and the need for community agreement remain significant factors in this evolution.

Keynote: GS1's experience with managing vocabularies and anchoring semantics (Phil Archer) [slides]

The presentation discussed GS1's experience with managing vocabularies and anchoring semantics. GS1 is a global standards organization that develops and maintains a number of standards related to the identification and tracking of products, services, and locations. The presentation also discussed the GS1 Web Vocabulary, which is a set of RDF vocabularies that can be used to describe GS1 identifiers. The presentation highlighted the benefits of using GS1 vocabularies, such as improved interoperability, reduced development costs, and increased innovation. The presentation also discussed a number of challenges that need to be addressed in order to effectively manage vocabularies, such as the need for a clear governance model and the need for tools to support vocabulary development and maintenance. The presentation concluded with a discussion of the future of GS1 vocabularies. The presentation argued that GS1 vocabularies have the potential to become a de facto standard for semantic interoperability in a number of domains.

The presentation further discussed the use of GS1 Digital Link URIs. GS1 Digital Link URIs are URIs that contain GS1 identifiers. These URIs can be used to link to data about products, services, and locations. The presentation highlighted the benefits of using GS1 Digital Link URIs, such as improved data sharing, increased innovation, and reduced costs. The presentation discussed a number of use cases where GS1 Digital Link URIs are being used, such as in the areas of supply chain management, product authentication, and consumer engagement. The presentation concluded with a discussion of the future of GS1 Digital Link *URIs. The presentation argued that GS1 Digital Link URIs have the potential to transform the way that data is shared and used.*

Phil: when I was in W3C I was involved in the work on SOSA, it was hard work and it is gratifying to see the current level of interest. I am now at GS1, which is a standard organisation for bar codes used across many industries. We're transitioning from serial to 2D bar codes given the demand for richer identifiers.

The bar code encodes a number: global trade identifier number. There is a mapping to URIs, where the domain name may vary depending on the task in hand.

We're aligning with verifiable credentials.

The bar code links to a JSON object and in turn to further information. We have long term stability in the identifiers. We have recently worked on RDF link types as part of the GS1 web vocabulary. This is modelled as an extension to schema.org and available from https://ref.gs1.org/voc

GS1 barcodes become an entry point for linked data about products.

We have a similar standards process to other SDOs with open working groups. We publish a log of changes with clearly identified versions, including links to the latest version.

Everyone says "we should align our standards". They mean you should use their standards, which is problematic in many cases. Different communities have different needs. We can use tools to support mappings and are happy to support those. We are happy to collaborate.

Phil Archer from GS1 presented their work, emphasizing their role as the barcode people but also their broader presence in pharmaceuticals, warehousing, and supply chains. The key points of his talk were:

- **GS1's Transition to 2D Barcodes:** Driven by the need for more information (batch numbers, serial numbers, expiry dates, etc.) in supply chains, GS1 is undergoing a significant transition from the traditional one-dimensional barcode to two-dimensional barcodes.
- **GS1 Digital Link URI:** This is a key mechanism for embedding GS1 identifiers (GTINs) into URLs, along with other attributes, creating a structured syntax that can be used online and increasingly by scanners. The domain name in the URI is not considered significant by GS1.
- **Moving from Identifier to Information:** GS1 aims to use the identifier in the data carrier (barcode, RFID, etc.) to link to a wealth of information through a "link set" (a JSON structure with typed links). This can point to product information pages, certifications, ownership details, maintenance information, and more.
- **Semantic Web Integration:** GS1 is building a system where things have identifiers that link to information using RDF predicates (link relation types), effectively creating a network of things with discoverable and semantically understandable information.
- **GS1 Web Vocabulary:** This vocabulary, originally an extension of schema.org, takes GS1's product attributes and turns them into a semantic web vocabulary, standardizing the link types used.
- **Stable and Persistent Hosting of Standards:** GS1 has established a dedicated repository (ref.gs1.org) for persistently hosting their standards and machine-readable assets, managed by technical staff to ensure stability and prevent accidental deletion.
- **Version Management:** New versions of the vocabulary come with a machine-readable log of changes, linking to the work requests that prompted them. Both the latest version (at a versionless URL) and immutable versions (with version identifiers) are available, following W3C best practices.
- **Alignment of Standards is Complex:** The common refrain of "we should align our standards" often means "you should use my standard." GS1 is not going to force its standards on others or adopt standards that don't fit their community (e.g., AAS).
- **Focus on Interoperability through Mapping:** Instead of forced alignment, GS1 is willing and keen to work with other vocabulary developers by including well-created cross-links (mappings) between their vocabulary and others, to enhance the accuracy and efficiency of AI-powered mapping tools.
- **Best Practices for Vocabulary Management:** GS1 emphasizes stable production, stable posting, and a documented change management process for their vocabulary.
- **Connecting Physical Things to the Web:** While not directly in everyone's domain, GS1's work is fundamentally about connecting physical things to information on the web using identifiers that are already ubiquitous.

In essence, Phil Archer highlighted GS1's evolution from traditional barcodes to a system leveraging linked data principles to connect products to a rich ecosystem of information,

emphasizing stable and persistent standards management and a pragmatic approach to interoperability through mapping rather than forced alignment.

The discussion following Phil Archer's presentation on GS1 raised several practical questions and highlighted potential collaborations:

- **GS1 and SC 41 (IoT Standardization):** It was noted that GS1 has a liaison category A with SC 41, positioning them within the IoT standardization landscape. This opens opportunities for collaboration, particularly in the area of policy and behavioural interoperability.
- **Call for GS1 Input on Interoperability Standard:** SC 41 is working on the next version of a standard on policy and behavioural interoperability and invited GS1 to provide input and real-world application examples (like those in water management or other sectors GS1 is involved in). This would demonstrate the importance of interoperability and showcase widely used applications adhering to potential rules.
- **QR Codes for Equipment Nameplates/Tags:** The power industry is exploring using QR codes on equipment nameplates to access information and links, moving away from limited text abbreviations.
- **Information Directly in QR Code vs. Linked Information:** The discussion clarified that QR codes should ideally contain minimal essential information, primarily a unique identifier (like a GS1 Digital Link URI). This URI then acts as a key to access a much broader range of information online.
- **Offline Access in Critical Situations:** For scenarios where internet connectivity might be unreliable (e.g., a technician on a wind turbine in harsh weather), the information needed immediately should be encoded in a way that can be extracted offline from the identifier (like the structured data within a GS1 Digital Link URI itself).
- **Limitations of QR Codes (Size and Security):** The size limitations of QR codes restrict the amount of direct information they can hold. Furthermore, QR codes themselves are inherently not secure and can be easily copied. Claims of "secure QR codes" are misleading.
- **Potential of Wireless Codes (NFC):** Wireless technologies like NFC were discussed as having the potential to exchange more information quickly and offer enhanced security features (due to the possibility of embedded computing power for generating signatures or one-time codes). While GS1 doesn't widely support NFC for current supply chain operations due to cost considerations, it's a promising technology for the future, especially in applications where cost is less of a barrier (like wind turbines) and security is crucial.
- **Security Lies in the Reader/System, Not the QR Code:** The security of interacting with a QR code comes from the software and systems that read and process the information it points to, not from any inherent security within the QR code itself.

In summary, the discussion explored practical applications of GS1 standards in other industries, highlighted the limitations and best practices for using QR codes, and considered the potential of more advanced wireless technologies like NFC, particularly in scenarios requiring offline access and enhanced security. The opportunity for GS1 to contribute to IoT interoperability standards was also emphasized.

Open Discussion: Best Practices for Vocabulary Development, Moderators: Phil Archer, Dave Raggett.

Dave Raggett shows the following in a slide to spark discussion ...

What are the key questions around best practices for creating and updating vocabularies, covering controlled vocabularies, taxonomies, ontologies, and semantic models?

- Some vocabularies are static whilst others need to be continually extended, e.g. ECLASS
- How to support legacy applications whilst evolving a vocabulary?
- How to measure maturity?
- Practices relating to transparency and accountability of the development process
- What's needed to establish trust in a vocabulary and the process used to maintain it?
- How to relate similar vocabularies developed by different communities with different perspectives and needs?

Does it make sense to try to align the terminology and metrics across standards development organisations? The following are some initial ideas to seed discussion and can be expected to evolve considerably:

✳ Informal work by a handful of people/organisations on a vocabulary of mutual interest

✳✳ Work on vocabularies with limited usage, but with an open process and documentation

✳✳✳ Draft standard for vocabularies with a well-defined process for measuring usage

✳✳✳✳ Standard for widely used vocabularies with a well-defined process for revisions, governance, and community feedback

✳✳✳✳✳ SDO standard for a mature, stable vocabulary

This rating system aims to provide a clear path for the development and adoption of vocabularies, fostering trust and collaboration across industries and ensuring semantic interoperability is achieved effectively. To encourage re-use rather than re-invention, we need a common framework for describing vocabularies in terms of agreed metrics and metadata. That includes criteria for selecting vocabulary terms.

The ensuing discussion focused on best practices for developing and managing vocabularies and standards, covering versioning, change management, maturity, reusability, testing, and the role of use cases.

Key Points on Development Best Practices:

- **Semantic Versioning:** Using semantic versioning (for human readability and machine understanding) was suggested as a good practice for indicating the nature of changes in vocabularies.
- **Machine-Readable Change Descriptions:** The need for machine-readable descriptions of vocabulary updates was emphasized to enable automated understanding and communication of changes to users and systems. Using a controlled vocabulary to describe changes could further aid this.
- **Addressing Versioning Challenges:** The difficulties in migrating between versions of vocabularies (as seen in BIM and smart appliances) highlight the importance of clear change management and user support.
- **Vocabulary Maturity:** Maturity can be domain-specific (a term is well-defined within a domain) but not necessarily cross-domain (the same term has different meanings in

other domains). Carefully structuring vocabularies and considering cross-domain consistency is important.
- **Reusability and Avoiding Reinvention:** Whenever possible, existing open standards and vocabularies should be reused to avoid redundant work. Mapping to established vocabularies is crucial.
- **Data Usage Tracking:** While challenging, understanding who is using a vocabulary or dataset can provide valuable feedback and indicate its relevance. Server logs alone are insufficient for this.
- **Ambiguity of Natural Language:** The inherent ambiguity of natural language in defining terms should be acknowledged, and formal, declarative definitions should be preferred when possible.
- **Criteria for Vocabulary Validation and Qualification:** The discussion touched upon the need for criteria to assess the usability and quality of vocabularies beyond just publication (e.g., on GitHub).
- **Importance of the Whole Semantic Stack:** True maturity involves having ontologies, well-defined semantics (including rules and inference capabilities), and practical data availability.
- **Semantic Triangle:** The concept of a "semantic triangle" (ontology, physics/engineering, and mathematical semantics) was introduced, emphasizing the need for all three for comprehensive semantic understanding. Efficient mappings between these representations are also crucial.

Role and Management of Use Cases:

- **Importance of Use Cases:** Gathering and curating use cases is a fundamental best practice for developing standards and vocabularies to ensure they address real-world needs.
- **Balancing Specificity and Generalization:** Finding the right balance between specific, initial use cases and broader applicability is a challenge. Standards that aim for longevity (like SOSA/SSN) need to be robust and often developed within formal standardization organizations.
- **AI for Use Case Curation:** The potential of using AI to help curate and analyse use cases was suggested.
- **Pragmatic, User-Driven Approach:** Starting with user needs and simple solutions (like JSON schema) and then gradually incorporating more formal semantics can be a successful approach, especially when driven by immediate data exchange requirements.
- **Standardized Use Case Methodologies and Repositories:** Efforts are underway (e.g., at IEC and in Europe) to standardize use case methodologies and create repositories for sharing and leveraging existing use cases.
- **Use Cases at Different Levels:** Different levels of use cases (user stories, specific use cases in the specification, test use cases for conformity) are needed for various audiences and purposes throughout the standardization process.
- **Testing as a Mandatory Part of Development:** Testing (including test suites and validation against example data) should be an integral part of the vocabulary and standard development process, ideally starting early.

Learning from Other Communities:

- **Biomedical Community as a Model:** The biomedical community was highlighted as a successful example of building and using interoperable ontologies (e.g., through BioPortal), and other domains can learn from their practices.

Final Thought:

- **Formalization Aids AI:** Formalization and standardization of semantics are beneficial for artificial intelligence and machine understanding.

In summary, the discussion emphasized a pragmatic yet rigorous approach to vocabulary and standard development, stressing the importance of user needs, practical implementation, thorough testing, and learning from established communities, while also acknowledging the complexities of versioning, cross-domain consistency, and achieving true semantic interoperability.

Session 7: Tooling for Vocabulary Development and Discovery

Martin Serrano made some introductory remarks outlining the content and goals of the upcoming session on semantic interoperability for digital twins. The key points were:

- **Session Focus:** The session, co-chaired, will focus on ontology discovery and the best practices and tooling for vocabulary generation.
- **Keynote on OntoPortal:** A keynote speaker from InfoAI will discuss their use of the OntoPortal for ontology discovery in the agricultural sector, building on previous mentions of the platform.
- **Presentations on Best Practices and Tooling:** Following the keynote, there will be three longer presentations from academics and industry representatives (SMEs and large companies). These talks will cover practical aspects of vocabulary generation and the tools used.
- **Context of Digital Twins:** The session emphasizes that the discussed tooling and best practices need to be considered within the context of data interoperability using semantics, ultimately aiming to represent the physical world in the virtual as digital twins. This means existing tools might need to evolve or be contextualized for this specific purpose.
- **Variety of Implementations:** The three presentations will showcase a good variety of implementations and examples of using vocabulary generation tools.

Keynote: Ontology discovery with ontoportal.org, as used by agroportal (Naouel Karam, InfAI) [slides]

The presentation discussed the use of ontology repositories to address the challenges of distributed ontologies, misalignment, and the ontology wasteland. The presentation also discussed the OntoPortal Alliance, which is a community of organizations that are working together to develop and maintain ontology repositories. The presentation highlighted the benefits of using ontology repositories, such as improved alignment, increased collaboration, and enhanced discoverability. The presentation also discussed a number of features of OntoPortal, such as ontology browsing, versioning, content visualization, and mapping. The presentation concluded with a discussion of the future of ontology repositories. The presentation argued that ontology repositories have the potential to transform the way that ontologies are developed and used.

The presentation further discussed the BiodivPortal, which is an ontology repository that is used to support biodiversity research. The presentation highlighted the benefits of using BiodivPortal,

such as improved data sharing, increased collaboration, and enhanced discoverability. The presentation also discussed a number of features of BiodivPortal, such as its support for different ontology formats, its ability to perform semantic reasoning, and its ability to generate SPARQL queries. The presentation concluded with a discussion of the future of BiodivPortal. The presentation argued that BiodivPortal has the potential to become a central hub for biodiversity data.

This talk presented the BiodivPortal, a biodiversity-focused instance of the OntoPortal technology, highlighting the need for ontology repositories and the features of OntoPortal. Here's a summary of the main points:

Why Ontology Repositories are Needed:

- **Misalignments:** Conflicting definitions and lack of consensus among ontologies.
- **Forgotten Ontologies:** Many ontologies exist only in publications and are not publicly available, leading to lost work and information.
- **Ontologies Trapped in Spreadsheets:** Valuable knowledge is locked in non-interoperable formats like Excel.
- **Versioning Problems:** Duplicated or conflicting versions of ontologies across different systems.
- **Standardization Bottleneck:** The slow and laborious process of creating ontologies within formal committees.
- **Synchronization Issues:** Lost updates between distributed ontology systems.

OntoPortal as a Solution:

- **Alliance of Repositories:** OntoPortal is a technology underpinning a network of openly accessible ontology repositories maintained by a consortium of research and data infrastructures (e.g., AgroPortal, EcoPortal, EarthPortal, Material Portal, and Biodive Portal).
- **Based on BioPortal Technology:** The technology has evolved from the NCBI BioPortal, an open-source project since 2005.
- **Federation:** Key portals (AgroPortal, EcoPortal, EarthPortal, and BiodivPortal) are now federated, allowing users to discover ontologies across these thematic areas.

BiodivPortal Implementation:

- **Instance for Biodiversity:** BiodivPortal is the German national research infrastructure (NFDI) instance for biodiversity.
- **Successful Deployment:** It has been running for over a year, hosting 50 ontologies and attracting 29 users.
- **Key Features:**
    - **Ontology Browsing:** Filtering by format, category, groups, etc. Supports OWL, SKOS, and UMLS.
    - **Ontology Details:** General information, identifiers, creators, and importantly, **versioning** with the ability to download specific versions and view differences (diffs) automatically calculated for ontologies linked to a specific location.
    - **Content Visualization:** Browsing classes, properties, and instances.
    - **Ontology Mappings:** Displaying existing mappings between ontologies (e.g., many mappings from NCBI Taxonomy to other BiodivPortal ontologies).
    - **Federation:** Displaying and linking to relevant ontologies from other federated OntoPortal instances.

- **Additional Features:** Text annotation service, ontology recommender, content negotiation, linked data publication, community features (comments, notes), enhanced SKOS support, fairness assessment tool, and multilingualism (English, French, with German and Italian in progress).

Standardization Efforts:

- **MOD Standard (Metadata for Ontologies):** Developing metadata standards to represent ontologies, including concepts for semantic artifact catalogues and services.
- **MOT API (Ontology Repositories and Terminology Services API):** Working on standardizing APIs for ontology repositories and terminology services.

Further Information:

- Websites provide documentation and links to the open-source GitHub repository.
- Publications (ISWC 2023 for OntoPortal technology and a specific publication for BiodivPortal) offer more technical details.

In summary, the BiodivPortal demonstrates the practical application of the OntoPortal technology for managing and discovering biodiversity-related ontologies, addressing key challenges in ontology management and fostering FAIR (Findable, Accessible, Interoperable, Reusable) principles within the biodiversity research community. The ongoing standardization efforts for metadata and APIs aim to further enhance the interoperability and usability of ontology repositories.

This discussion following Nauel's presentation on BiodivPortal raised questions about content negotiation, ontology recommendation, and the rationale behind federated ontology repositories.

Content Negotiation:

- **URI Namespace Management:** A question was raised about how BiodivPortal manages URI namespaces for content negotiation, particularly when multiple developers contribute ontologies.
- **Portal-Specific Base URIs:** The approach is for projects developing ontologies within the BiodivPortal context to use a base URI specific to the portal. This ensures that ontologies developed within the project have a clear namespace.
- **Original Ontologies Retained:** Original ontologies from external sources retain their original URIs when ingested into the portal. If a portal-specific URI is used for a version within BiodivPortal, both the original and the portal's version can be displayed.

Ontology Recommendation:

- **Content-Based Recommendation:** The recommender service allows users to paste text and receive suggestions for relevant ontologies based on the text's content. This is primarily for annotation purposes.
- **FAIR Assessment as Guidance:** The FAIR assessment tool evaluates ontology metadata (mapping to other ontologies, adherence to FAIR principles, etc.) to help users decide on reusability.
- **Community Interaction:** Recommendations are also heavily influenced by direct communication with the user community through workshops and other interactions. The community-specific nature of the portals is key here.

Federation of Ontology Portals:

- **Conflict Resolution and Redundancy:** A question addressed potential conflicts and redundancies arising from federating thematically similar portals (AgroPortal, EcoPortal, EarthPortal, BiodivPortal).
- **Canonical Portal Designation:** For ontologies common across federated portals (around 30 out of 450), a "canonical portal" is designated as the responsible entity for curation (metadata, updates, etc.). Other portals then pull information from the canonical source, ensuring consistent versions.
- **Rationale for Multiple Portals vs. a Single Marketplace:** The decision to have multiple, community-focused portals that are then federated is driven by the need for:
    - **Dedicated Curation:** Each domain (agriculture, ecology, biodiversity, earth science) benefits from focused curation efforts by experts in that area.
    - **Stronger Community Links:** Individual portals can build closer relationships with their specific user communities, understanding their needs and curating relevant ontologies.
    - **Broader View through Federation:** Federation then provides a wider perspective on available ontologies across related domains, offering a balance between focused curation and comprehensive discovery.

In summary, the discussion clarified the technical aspects of URI management and content negotiation within BiodivPortal, explained the multi-faceted approach to ontology recommendation (combining content analysis, FAIR assessment, and community interaction), and justified the federated model of OntoPortal instances as a way to balance domain-specific curation with broader ontology discovery.

Talk: Best practices and tooling for vocabulary development and publishing (María Poveda-Villalón) [slides]

The presentation discussed the FAIR principles for ontology development. FAIR stands for Findable, Accessible, Interoperable, and Reusable. The presentation also discussed a number of tools that can be used to support FAIR ontology development, such as Linked Open Vocabularies (LOV), Chowlk, and Widoco. The presentation highlighted the benefits of using FAIR ontologies, such as improved data sharing, increased innovation, and reduced costs. The presentation also discussed a number of challenges that need to be addressed in order to develop FAIR ontologies effectively, such as the need for high-quality metadata and the need for tools to support ontology evaluation. The presentation concluded with a discussion of the future of FAIR ontologies. The presentation argued that FAIR ontologies have the potential to transform the way that data is shared and used.

The presentation further discussed the LOT (Linked Open Terms) methodology, which is a lightweight and agile process for ontology development. The LOT methodology is based on a number of principles, such as the use of competency questions, the use of intermediate representations, and the identification and definition of activities. The presentation highlighted the benefits of using the LOT methodology, such as improved collaboration, reduced development costs, and increased innovation. The presentation discussed a number of tools that can be used to support the LOT methodology, such as a template for ontology requirements specification and a tool for ontology evaluation. The presentation concluded with a discussion of the future of the LOT methodology. The presentation argued that the LOT methodology has the potential to make ontology development more accessible and efficient.

María's talk focused on best practices and tools for developing FAIR (Findable, Accessible, Interoperable, Reusable) ontologies. Here are the main points:

- **Ontology Development Expertise:** María is an experienced ontology developer working across various domains and also researches methods and tools to empower domain experts to build ontologies.
- **Leveraging Existing Methodologies:** While new languages like OWL exist, established ontology building methodologies from the past (pre-OWL) still hold valuable principles.
- **Applying FAIR Principles to Ontologies:** Following the emergence of FAIR principles for data in 2016, the community started applying these principles to ontologies and vocabularies as research outputs. Several publications provide guidelines for making ontologies more FAIR.
- **LOT (Linked Open Terms) Methodology:** This methodology, published after the FAIR principles, specifically considers publication and documentation in the context of the Semantic Web, providing an iterative pipeline with tools, recommendations, and tips for each step. It also includes lessons learned from past project mistakes.
- **Integrating FAIR Throughout the Ontology Lifecycle:** Instead of just checking FAIR compliance at the end, the FAIR Impact project explored how to integrate FAIR considerations into each stage of the ontology development process (reuse, conceptualization, encoding, documentation, publication).
- **Essential Tools for FAIR Ontology Development (Desert Island Scenario):**
    - **Ontology Reuse: Linked Open Vocabularies (LOV)** and domain-specific portals like **OntoPortal** and **BioPortal** for finding existing ontologies to improve interoperability and reuse FAIR vocabularies.
    - **Ontology Conceptualization and Encoding: Chowlk** (a UML-based framework and notation adapted for OWL ontologies) used with **diagrams.net (draw.io)** for visual modeling and automatic OWL code generation (via an online converter), allowing collaboration with domain experts without exposing them to Protégé.
    - **Ontology Documentation: Withdoko** for generating HTML documentation, including Chowlk diagrams and examples for better findability, interoperability, and reusability.
    - **Ontology Publication: Ontology Pitstop** for publishing ontologies under W3C IDs with built-in content negotiation or for generating bundles for self-hosting with predefined content negotiation.
    - **FAIR Assessment: F-O-O-D (FAIR Ontology Observatory Dashboard)**, a simple tool to check the FAIRness of a published ontology by evaluating metadata, content negotiation, HTML provision, version URI compliance, and more, providing a pass/fail output and suggestions for improvement.
- **Conclusion and Future Directions:** While existing tools can be leveraged for FAIR ontology development, there is a need for more user-friendly tools for metadata updates and maintenance. Checking FAIRness should ideally happen throughout the ontology lifecycle, not just at the end.

Long Talk: SEMAPTIC - A new tool for fast and easy adoption of semantic interoperability (Carlos Pereira, INESC-TEC) [slides]

The presentation discussed SEMAPTIC, a new tool for fast and easy adoption of semantic interoperability. Semantic interoperability is the ability of different systems to exchange data with unambiguous, shared meaning. The presentation also discussed the benefits of semantic interoperability, such as improved data sharing, increased innovation, and reduced costs. The presentation highlighted the challenges of semantic interoperability, such as the complexity of the technology and the lack of standards. The presentation concluded with a discussion of the

future of SEMAPTIC. The presentation argued that SEMAPTIC has the potential to make semantic interoperability more accessible and efficient.

The presentation further discussed the SEMAPTIC architecture. The SEMAPTIC architecture is based on a number of principles, such as the use of a semantic map, the separation of data and metadata, and the use of a variety of semantic technologies. The presentation highlighted the benefits of using the SEMAPTIC architecture, such as improved interoperability, reduced development costs, and increased innovation. The presentation discussed a number of features of SEMAPTIC, such as its ability to generate semantic maps, its support for different data formats, and its ability to perform semantic reasoning. The presentation concluded with a discussion of the next steps for SEMAPTIC. These steps include developing new features for SEMAPTIC, improving the performance of SEMAPTIC, and deploying SEMAPTIC in a number of pilot sites.

Carlos Pereira from INESC-TEC presented SEMAPTIC, a practical approach to semantic interoperability designed to be an easy entry point for adoption. Here's a summary of the main points:

- **Semantic Interoperability is Versatile:** There isn't one single way to achieve it. Examples include agreed-upon vocabularies within a cluster, controlled frameworks like the Interconnect project, and generic approaches for data producers to add semantic meaning for consumers.
- **Why Adoption Remains a Challenge:** Despite being defined for almost 30 years, semantic interoperability faces challenges in widespread adoption due to the perceived complexity and the initial learning curve.
- **Traditional Semantic Interoperability Workflow (and its drawbacks):**
    - **Producer:** Chooses ontologies, selects a semantic output format (e.g., RDF), and transforms data, often leading to data size inflation (e.g., one data point requiring multiple RDF triples).
    - **Consumer:** Receives semantic data (e.g., RDF) and may be able to perform inferences, but often needs to revert the process to extract the original data format required by their existing services and databases. This requires synchronization.
- **Semaptic's Approach (Key Principles):**
    - **Simple Data Model:** Compliant with semantic interoperability definitions.
    - **Simple Specification:** Easy to understand and implement.
    - **Simple Tooling:** To aid in creating the semantic map.
    - **No Data Transformation:** The original data is not altered.
    - **No Data Size Inflation:** The original data size remains the same.
    - **Clear Separation:** Between the raw data and its semantic specification.
- **Semaptic Workflow:**
    - **Producer:** Builds a "semantic map" (once per data structure) by identifying relevant ontologies and mapping their data parameters to concepts within those ontologies.
    - **Embedded Semantic Map:** The semantic map is embedded directly within the original data (e.g., JSON).
    - **Semantic Map Structure:** Consists of three blocks:
        - **Prefix:** Declares the ontologies/vocabularies used.
        - **Context (Modelling):** Maps each data parameter to a concept (class or property) in the declared ontologies.
        - **Annotations:** Includes extra metadata to enrich the service.

    - **Data Exchange:** The producer provides the original data with the embedded semantic map to the consumer.
    - **Consumer Options:**
        - **No Reasoning Needed:** Use the original data directly.
        - **Parameter Understanding:** Consult the embedded semantic map for clarity on data parameters.
        - **Advanced Reasoning:** Serialize the data (including the semantic map) into a format suitable for their semantic tools (e.g., RDF).
- **Semaptic Tool:** A web-based tool facilitates the creation of the semantic map:
    - Allows users to select or upload relevant ontologies.
    - Extracts keys from pasted JSON data.
    - Enables mapping each data key to classes and properties from the chosen ontologies (with dropdowns filtered by the selected ontology).
    - Automatically generates the semantic map.
    - Provides options to copy the semantic map and visualize it graphically.

In summary, SEMAPTIC offers a lightweight and practical approach to semantic interoperability by embedding a semantic map directly within the original data, avoiding data transformation and inflation. This allows data consumers to understand the data's meaning without necessarily requiring complex semantic tools, while still providing the necessary information for those who do need advanced reasoning capabilities. The accompanying web tool simplifies the creation of these semantic maps.

Long Talk: The Local Digital Twin Toolbox: Catalysing Smart City Transformation (Antonio Jara, Libelium) [slides]

The presentation discussed the EU LDT Toolbox, which is a set of tools that enables cities to simulate and predict scenarios of local digital twins (LDT). The presentation also discussed the benefits of using the EU LDT Toolbox, such as improved decision-making, enhanced community engagement, and increased innovation. The presentation highlighted the importance of semantic interoperability for digital twins in the urban domain. The presentation concluded with a discussion of the future of the EU LDT Toolbox. The presentation argued that the EU LDT Toolbox has the potential to transform the way that cities are managed.

The presentation also discussed a number of specific applications of the EU LDT Toolbox. These applications include the use of the EU LDT Toolbox for mobility planning, energy optimization, urban health, urban planning, and sustainability. The presentation also discussed a number of benefits of using the EU LDT Toolbox, such as improved data quality, enhanced data analysis, and increased automation. The presentation concluded with a discussion of the future of the EU LDT Toolbox. The presentation argued that the EU LDT Toolbox has the potential to make it easier to create, access, share, and reuse data in the urban domain.

This talk introduced the **European Local Digital Twin (LDT) Toolbox**, a European Commission initiative focused on creating capabilities for digital twins specifically for **smart cities**, with a strong emphasis on the **Green Deal** and **energy efficiency**. It includes a collection of relevant models, e.g. mobility, pollution, energy forecasting, etc., and supports the use of AI to convert data into knowledge, and to visualise the impact.

Here are the main points:

- **Local Digital Twin Concept:** A digital replica of a city or parts of it, integrating data, algorithms, and models to represent urban behaviour.
- **Toolbox Purpose:** To provide a set of tools and components to enable the creation and use of LDTs in cities with varying levels of maturity.

- **Key Capacities of the Toolbox:**
    - **Data Acquisition:** Primarily supporting NGSI-LD.
    - **City Modelling & Insight:** Exploring AI models.
    - **Interconnection:** Using NGSI-LD to communicate with other systems.
    - **City Verse (Visualization):** 3D visualization capabilities.
- **User-Centric & Regulatory Focused:** Designed to be user-friendly and supports regulatory compliance (AI Office involvement).
- **Leveraging Existing Initiatives:** Aims to build upon existing European activities like data spaces (Simple Finance, etc.) and previous projects (FIWARE).
- **Open Source:** The entire toolbox is open source.
- **Key Components:** Data platform, data space connectors, toolbox itself, AI notebook, and a marketplace.
- **Emphasis on Interoperability:** Regulatory compliance and interoperability are central, leveraging "Minimal Interoperable Mechanisms" (MIMs) and standards (ETSI, ITU).
- **European Local Digital Twin Infrastructure Consortium:** A consortium of 14 EU members to leverage infrastructure and best practices for wider adoption.
- **Initial Assets:**
    - **European Building Database (Person University):** Detailed data on European buildings.
    - **AI Models:** For energy demand, efficiency, pollution, mobility, and reconstruction costs (already used in Valencia and Ukraine).
- **Key Use Cases (Inspiration for New Calls):** Urban health, urban planning, sustainability, energy optimization, and mobility planning.
- **Further Collaboration Opportunities:**
    - Securing infrastructure.
    - Developing transformation plans for applying models.
    - Creating specific simulations.
- **Toolbox Workflow:** Gather data -> Transfer data to knowledge (AI models) -> Simulate scenarios -> Visualize impact -> Share with citizens -> Support decision-making.
- **Example: Low Emission Zones:** Illustrates how the toolbox integrates data from various sources (air quality, traffic, weather, noise), AI models, scenario planning, and visualization.
- **Specific Tools Mentioned:**
    - **Data Modeler:** Crucial for converting and modelling data (needs integration and extension). Currently using Smart Data Models.
    - **Data Space Connectors:** Based on FIWARE and Eclipse connectors.
    - **Data Platform:** Oriented towards FIWARE experience.
    - **AI Notebook:** For modelling and editing AI models, federated learning.
    - **Marketplace:** For data gathering and potentially other services.
    - **Use Case & Scenario Design:** Low-code framework (based on Node-RED).
    - **Visualization:** Open Standard for CityGML, Open 3D Streamlines, Grafana for charting.
- **Funding Opportunity: €17 million in cascade funding** available this year for including use cases, extending tools, and experimenting with new ideas (e.g., federated learning at the edge for privacy).
- **Timeline:** The first version of the tools will be released in **June of this year**.
- **Call for Collaboration:** Interested parties are encouraged to reach out for more information and to explore collaboration opportunities.

- **Vision for the Future:** The European Commission aims to interconnect initiatives like Simple Finance, the LDT Toolbox, and Cityverse to create a comprehensive ecosystem for data exploitation and digital twins.

Panel: Tooling for Vocabulary Development and Discovery - Moderators: Martin Serrano, Pierre-Antoine Champin

The panel discussion session focused on identifying challenges and future directions for semantic interoperability, building on the presentations from Carlos Perella and María Poveda, and including input from Antonio Jara.

Key Challenges and Future Directions:

- **Reluctance to Transform Data:** A significant challenge, highlighted by Carlos' experience, is the unwillingness of data providers (e.g., DSOs) to transform their data into semantic formats, even within research projects with semantic data sharing requirements. Semaptic's approach of separating data and semantic specification aims to address this.
- **Debate on Deriving Semantics from Syntactic Structure:** Alberto raised a critical point questioning whether semantics can be truly derived solely from the structural analysis of syntactic data. He argued that semantics should be clear and independent of representation and should encompass behaviour and processes, not just structural mappings. Carlos clarified that their current requirement is primarily to remove ambiguity, and the transformation to richer semantic formats for reasoning is a separate, ongoing process.
- **Effectiveness of Semantic Discovery Engines:** María highlighted that current semantic discovery engines often act as centralized registries rather than enabling true discovery of new knowledge.
- **Role of Agents in Discovery:** Maria suggested exploring agent-based systems for discovering and mapping semantic resources, rather than relying solely on centralized registries.
- **Generative AI (LLMs) for Connecting Data to Ontologies:** The potential of LLMs to bridge the gap between data and ontologies (e.g., for mapping and virtualizing queries) was acknowledged as a strong trend.
- **LLMs vs. Ontologies: Complementary, Not Competitive:** Antonio Jara and a later commenter emphasized that LLMs and ontologies are likely to be complementary rather than competitive. LLMs can assist in mapping, data collection, and making semantic annotation more accessible, while ontologies provide the structured knowledge, reasoning capabilities, traceability, and reduced hallucination necessary for reliable AI applications.
- **Importance of Ontologies for LLM Training:** It was noted that ontologies can be used to structure the datasets that LLMs learn from, potentially leading to better learning and representation of data.
- **Separating Operational Speed from Strategic Semantic Thinking:** Antonio highlighted the need to differentiate between the need for quick operational solutions and the strategic importance of semantic alignment, which requires more in-depth discussion and vocabulary/ontology development. He suggested that workshops should consider this distinction.
- **Accessibility as the Key to Semantic Interoperability:** The panel concluded that accessibility is a crucial factor in achieving semantic interoperability. Making data

accessible and adhering to FAIR principles will facilitate greater semantic understanding and exchange.
- **Federation Between Domains:** The increasing trend of federation between different domains that share common elements demonstrates a move towards broader semantic interoperability.

In summary, the panel discussion identified the ongoing challenge of data transformation reluctance, debated the foundations of semantic derivation, explored the limitations of current discovery methods and the potential of agent-based approaches, and strongly emphasized the complementary roles of LLMs and ontologies in advancing semantic interoperability. The importance of accessibility, FAIR principles, and strategic semantic thinking were also highlighted as key enablers for future progress.

Session 8: Bringing It All Together

Session Chair/Moderator: Claude Baudoin

Claude sets the scene. Thanks to all the presenters - there was a lot of insightful material!

1. [Share] best practices on vocabularies
2. "Reuse, align, be inspired" [by ontologies developed by others]
3. If you don't reuse, justify it
4. Validation
5. Scenarios
6. Need to give more time to present [these initiatives], dig deeper, monitor their evolution
7. Hold more focused events (e.g., one workshop dedicated to the Asset Administration Shell)
8. [Emphasize] reusability across domains
9. Think of the life cycle: ontology to taxonomy to … (similar to what SLB presented)
10. Look at the European Commission's BRIDGE Data Management Working Group
11. Ideas should be handled by SDOs, so try to hold joint meetings with them
12. Publish a calendar of related events, so we can attend them
13. Write joint report(s), distributed to multiple target organizations
14. Socialize this [workshop's outcomes]
15. Form a smaller group to solve issues – could be a breakout session in a future workshop
16. Develop a case study (e.g., in traffic, transportation, or water management) [domains where owners of digital twins are not in competition with each other]

The wrap-up session aimed to identify how to bring together the discussions and insights from the workshop and define concrete next steps. Here's a summary of the key points raised:

Challenges and Considerations:

- **Ontology Interoperability Beyond Structure:** Emphasized the need to consider data rules, constraints, and formulas alongside conceptual models (classes, attributes, relations) for practical ontology interoperability. The Mars climate orbiter failure was cited as a real-world example of the dangers of ignoring unit mismatches.
- **Collaboration in the Research Community:** A plea for greater collaboration among academic and research institutions to avoid duplication of effort and maximize the impact of their work for end-users.
- **Bridging Research to Concrete Results:** The need to translate research into practical, usable outcomes and tools.

- **Intellectual Property and Competitiveness:** Acknowledged the challenges of encouraging organizations to share ontologies due to IP concerns and competitive pressures.

Proposed Initiatives and Actions:

- **Best Practices Handbook for Ontologies:** Develop a resource similar to best practices in geospatial data, with a mantra of "Reuse, align, or be inspired." Include a set of prepared reusable ontologies and a requirement to justify not using them.
- **Ontology for Ontologists (Validation):** Create a schema or constraints for ontologies themselves to enable validation and ensure quality from the outset. Use tools like DX probe for this.
- **Regular Events and Enhanced Communication:** Suggestions for making the workshop an annual event (potentially hosted at different locations like IHP in Frankfurt/Oder [Eastern Germany on the border to Poland]), creating a Slack group or similar forum for ongoing discussion, and sharing contact information among participants (with consent).
- **Leveraging the Local Digital Twin Initiative:** Use this initiative to drive cross-domain reusability of ontologies.
- **Integration in the Engineering Lifecycle:** Define "recipes" for when and how to use ontologies throughout the engineering lifecycle, from strategic planning to implementation.
- **Engagement with Policymakers and Standardization Bodies:** Work with policymakers and standards organizations to integrate ontologies effectively into regulations and standards themselves.
- **Joint Events with Standardization Organizations:** Organize joint workshops or sessions in conjunction with standardization body meetings (e.g., OMG in Leeds, ETSI ISG CIM plenary in Berlin) to foster collaboration and bring the research community closer to standardization processes.
- **Publishing a Calendar of Relevant Events:** Create a calendar of technology-related events across various organizations to improve awareness and encourage participation.
- **Leveraging AIOTI Mechanisms for Reports and Input to Standards:** Utilize AIOTI's existing mechanisms for producing reports and contributing to standards organizations (ISO TC 41, ISO TC 42, ISO TC 211, OGC) on topics like digital twins, data spaces, and geospatial ontologies. Enhance the semantic aspects of these contributions.
- **More Focused Workshops on Specific Problems:** In addition to broader events, organize smaller, more focused workshops to deep-dive into specific semantic interoperability challenges and work towards solutions (e.g., addressing the AAS problem).
- **Case Studies for Cross-Sector Integration:** Develop concrete case studies that demonstrate cross-sector integration of data and the need for semantic interoperability in areas like smart cities, traffic management, and water management. This could be a collaborative working effort for the next half-year.
- **Leveraging Existing Standards and Initiatives (e.g., ITS in the US - Arc IT, Smart Data Models):** Explore and build upon existing successful models for semantic interoperability in specific domains, potentially through collaboration with experts like Antonio Jara on transportation use cases.
- **Careful Engagement with Other Communities:** When engaging with communities that may not be deeply familiar with semantic web technologies, approach them

respectfully, explain the benefits and trade-offs, and avoid a purely critical "TikTok style" approach. Highlight the potential for mutual benefit.
- **Emphasis on the Standards Development Lifecycle:** Highlight the importance of continuous feedback and revision processes within standards development organizations (like OMG's finalization and revision task forces).

Concrete Case Studies Proposed:

- **Traffic Management:** Focusing on the specific challenges and opportunities for semantic interoperability in this domain (potentially leveraging US ITS examples).
- **Water Management:** Using the digital twin of Valencia's water management system as a potential example.

The overall tone of the wrap-up was action-oriented, emphasizing the need to move beyond discussion towards concrete initiatives, collaborations, and engagement with both standardization bodies and specific use cases to demonstrate the value of semantic interoperability.

Conclusion and Recommendations for Next Steps

The workshop successfully highlighted the critical importance of semantic interoperability for realising the full potential of Digital Twins and addressing real-world industrial challenges.

A key bottleneck to wider adoption of semantic technologies is the complexity and labour-intensiveness of data to ontology mapping. There is a rich landscape of ongoing research, projects, and standardisation efforts related to semantic interoperability, but better coordination and focus are needed to maximise impact.

Generative AI, particularly LLMs, offers promising avenues for automating and improving aspects of semantic interoperability, but further research is needed to address current limitations and ensure reliability.

The workshop underscored the need for both practical tools and high-level strategic alignment to foster semantic interoperability across diverse domains and stakeholders.

Recommendations and Next Steps:

This workshop covered a lot of ground and the participants generally felt it would be well worth following up with further workshops. Some topics for exploration are listed below:

- **Prioritise Development of User-Friendly Data Mapping Tools:** Given the identified importance of data mapping, a key recommendation is to focus on the development of user-friendly, graphical data mapping tools. This should be a priority for future research and development efforts. These tools should aim to lower the barrier to entry for users with limited semantic expertise.
- **Establish a Semantic Interoperability Roadmap:** Develop a roadmap that outlines key research directions, standardisation activities, and practical steps for advancing semantic interoperability, with specific milestones and measurable outcomes.
- **Foster Collaboration and Knowledge Sharing:**
    - Create a sustainable platform (e.g., an online forum, knowledge base) for continued communication, collaboration, and knowledge sharing among researchers, industry practitioners, and standardisation bodies.

    - Organise regular workshops, conferences, or joint events with relevant organisations (e.g., W3C, OMG, ISO) to facilitate dialogue and promote collaboration.
- **Investigate and Develop LLM-Based Solutions:**
    - Further research and development are needed to explore the potential of LLMs for various semantic interoperability tasks, including data mapping, ontology creation, and ontology alignment.
    - Address challenges related to the reliability, accuracy, and testability of LLM-generated outputs in this context.
- **Develop Best Practice Guidelines:**
    - Create and disseminate best practice guidelines for ontology development, data mapping, and semantic interoperability, drawing on successful case studies and lessons learned.
    - These guidelines should address issues such as ontology quality, modularity, extensibility, and alignment.
- **Focus on Cross-Domain Use Cases:**
    - Develop and demonstrate concrete cross-domain use cases that highlight the benefits of semantic interoperability and the value of ontologies in solving real-world problems.
    - These use cases should involve stakeholders from different domains and address specific challenges, such as data integration, system interoperability, and knowledge sharing.
- **Engage with Standardisation Bodies:**
    - Actively engage with relevant standardisation bodies (e.g., ISO, IEC, W3C, ETSI) to ensure that semantic interoperability considerations are integrated into relevant standards and specifications.
    - Contribute research findings, best practices, and use case examples to inform the development of standards.
- **Address Data Governance and Trust:**
    - Recognise and address the importance of data governance, trust, and security in semantic interoperability.
    - Explore how semantic technologies can support data provenance, access control, and privacy-preserving data sharing.

Contributors

Editor:

Dave Raggett, W3C / ERCIM
Martin Bauer, NEC Laboratories Europe

Reviewer:

Damir Filipovic, AIOTI Secretary General

Contributors:

Alberto Abella, FIWARE Foundation
Vladimir Alexiev, Graphwise/Ontotext
Darko Anicic, Siemens
Pierre-Antoine Champin, W3C
Phil Archer, GS1
Claude Baudoin, Object Management Group (OMG)
Martin Bauer, NEC Laboratories Europe
Arne Berre, SINTEF AS
Cornelis Bouter, TNO
Fatma Chamekh, Bureau de Recherches Géologiques et Minières
Franco Crivello, CETaqua, Water Technology Center
Jan B. de Meer, Object Management Group (OMG)
Cordelia M Ezhilarasu, SLB
Nikolaos Filinis, National Technical University of Athens
Anna Geraci, LINKS Foundation
Dominique Hazaêl-Massieux, ERCIM
Antonio Jara, Libelium
Sebastian Käbisch, Siemens
Martin Kaltenböck, Graphwhise/SWC
Naouel Karam, Institute for Applied Informatics
Elisa Kendall, Object Management Group (OMG)
Antonio Kung, Trialog
Francisco Martin-Recuerda, SINTEF AS
Henon Mengistu Lamboro, Institut Polytechnique de Paris and Trialog
Svein Olsen, Statnett TSO
Carlos Pereira, INESC-TEC

Danilo Pillitteri, Parco Scientifico e Tecnologico della Sicilia (PSTS)
María Poveda-Villalón, Universidad Politécnica de Madrid
Dave Raggett, W3C / ERCIM
Hossein Rimaz, Metaphacts
Martin Serrano, Insight SFI Research Centre
Claudio Sesto, RSE S.p.A.
Ken Wenzel, Fraunhofer IWU

Acknowledgements

The workshop chairs wish to thank the program committee and all of the participants for their help in making a successful event. We also want to thank INRIA for the meeting facilities, and the Nephele and SmartEdge projects for their support.

About AIOTI

AIOTI is the multi-stakeholder platform for stimulating AI, IoT and Edge Continuum Innovation in Europe, bringing together small and large companies, academia, researchers, policy makers, end-users and representatives of society in an end-to-end approach. We strive to leverage, share and promote best practices in the AI, IoT and Edge Continuum ecosystems, be a one-stop point of information to our members while proactively addressing key issues and roadblocks for economic growth, acceptance and adoption of the AI, IoT and Edge Continuum Innovation in society. AIOTI contributions goes beyond technology and addresses horizontal elements across application domains, such as matchmaking and stimulating cooperation by creating joint research roadmaps, defining policies and driving convergence of standards and interoperability.