MareGraph
Towards an Interoperable Marine Knowledge Graph

Project Title	Towards an Interoperable MARinE Knowledge GRAPH
Project Acronym	MareGraph
Grant Agreement No.	101100771
Start Date of Project	23/01/2022
Duration of Project	36 Months

D2.2 Guidelines on process and methodology for organisational interoperability (Version 5)

Work Package	WP 2, Process tuning, community building and dissemination
Lead Author (Org)	Eveline Vlassenroot (Digitaal Vlaanderen)
Contributing Author(s) (Org)	Lucas Cornette (Digitaal Vlaanderen), Lorenzo Vylders (Digitaal Vlaanderen), Liesbeth Van der Haegen (Digitaal Vlaanderen), Laurens Vercauteren (Digitaal Vlaanderen), Samuel Van Ackere (Digitaal Vlaanderen), Simon Claus (Digitaal Vlaanderen), Annelies De Craene (Digitaal Vlaanderen), Giorgia Lodi (CNR-ISTC), Marc Portier (VLIZ)
Due Date	30.06.2024
Date	29.01.2024
Version	V4.0

Dissemination Level

X	PU: Public
	PP: Restricted to other programme participants (including the Commission)
	RE: Restricted to a group specified by the consortium (including the Commission)
	CO: Confidential, only for members of the consortium (including the Commission)

Funded by the European Union

Versioning and contribution history

Version	Date	Author	Notes
0.4	1.11.2024	Eveline Vlassenroot (Digitaal Vlaanderen), Samuel Van Ackere (Digitaal Vlaanderen), Giorgia Lodi (CNR-ISTC)	Revisions on version 0.3
0.3	1.08.2024	Lorenzo Vylders (Digitaal Vlaanderen)	Revision after the workshop during the SEMIC pre-conference
0.2	28.03.2024	Lorenzo Vylders (Digitaal Vlaanderen), Lucas Cornette (Digitaal Vlaanderen)	Revision after General Assembly 22 and 23 February 2024
0.1	29.01.2024	Lorenzo Vylders (Digitaal Vlaanderen), Lucas Cornette (Digitaal Vlaanderen), Raf Buyle (Digitaal Vlaanderen)	Copy for D2.2 + adding vision and adjusting process and method

Funded by the European Union

1 Table of Contents

1 Table of Contents 3
2 Terminology 4
3 Executive Summary 6
4 Context and approach 6
4.1 What did we do in D2.1? 6
4.2 Identified challenges 7
4.3 Approach 9
5 Unifying principles 9
6 Methodology for developing semantic assets 13
6.1 Setting up a working group and its operational context 14
6.1.1 Setup of the working group 14
6.1.2 Declaration of intent 14
6.1.3 Stakeholder engagement workshop 15
6.1.4 Peer review and approval process 15
6.1.5 Working group meetings 16
6.1.6 Communication and engagement 16
6.2 Running a preliminary analysis of the domain 16
6.3 Creating drafts of semantic assets 17
6.3.1 Requirements elicitation 18
6.3.2 Design of the semantic assets 18
6.4 Quality control and feedback collection from communities 21
6.5 Finalising semantic assets 23
6.6 Publication of semantic assets 24
7 Change Management 25
8 Conclusion 25
9 Annex 27
9.1 Survey questions and feedback 27
9.2 Panel discussion 28

Funded by the European Union

2 Terminology

Terminology/Acronym	Description
Application profile	It is a specification built using one or more base standards/ontologies, adding more specificity by identifying mandatory, recommended and optional elements to be used for a particular application, as well as recommendations for controlled vocabularies to be used. Examples are the application profiles developed within the OSLO project or DCAT-AP for metadata exchange between (Open) data portals.
Controlled vocabulary	An umbrella term for code lists, taxonomies and thesauri. Controlled vocabularies are usually standardised lists of concepts denoted by terms of reference and codes. The list is useful for organising, describing, predefining, and indexing knowledge of a domain. An example of a controlled vocabulary is the Language Named Authority List of the Publications Office of the European Union (OP).
Declaration of Intent	A declaration of intent describes the domain and purpose of the ontology to be developed and is communicated to various relevant stakeholders at the start of the process.
Domain model	A domain model is a conceptual model of a certain domain. It is a formal representation that represents the knowledge of a domain in terms of entities and their relationships.
Graffoo	Graffoo is an open-source tool for drawing ontologies that are clear and easy to understand. It helps ontology designers and re-users better understand the ontologies, thus facilitating communication and collaboration among stakeholders in the ontology development process.
High-level domain model	A high-level domain model describes the relevant entities in a domain with a high level of abstraction. It results from a first step towards the formalisation of domain knowledge and the analysis of information needs.
JSON-LD	JavaScript Object Notation for Linked Data is a way to represent Linked Data in JSON.
Ontology	In computer science, ontology refers to a formal, shared and explicit specification of a representation (conceptualisation) of a knowledge domain, defined based on specific requirements. The representation consists of the definition, done typically through a collaborative process, of entities (or classes), entity attributes and relationships between entities. It is designed to describe a consistent and

Funded by the European Union

Terminology/Acronym	Description
	reproducible way of organising and categorising information, enabling data classification, retrieval, and interpretation within a specific domain or subject area.
Semantic assets	A semantic resource that is used to represent the semantics of data at different levels of granularity. Examples of semantic assets are ontologies but also code lists, taxonomies, thesauri.
SHACL	Shapes Constraint Language is a way to describe and validate data graphs (in RDF).
Specification document	It is a technical document that gives substance to an ontology by providing more information on the usage of the elements of the ontology. Specifications can be adjusted based on advancing insight without changing the corresponding ontology.
UML class diagram	A diagram expressed using the UML – Unified Modelling Language notation, describes a system's structure based on classes, attributes, relationships, and operations.

Funded by the European Union

3 Executive Summary

This executive summary presents an overview of Deliverable D2.2, which builds further upon Deliverable D2.1[1] on the process and methodology for semantic assets creation that can be used in different Member States. A defined process and methodology for organisational interoperability is crucial as it ensures seamless communication, data exchange, and collaboration between different departments, systems, and organisations, enhancing efficiency, reducing errors, and enabling effective decision-making.

This deliverable, based on the experience of the MareGraph project and feedback collected from a wider audience during the SEMIC Conference in June 2024, represents a flexible methodology that can accommodate different scenarios with the scope to scale at the European level and suggest common practices that can facilitate the reuse of various semantic assets.

The proposed methodology emphasises the importance of a collaborative approach, promoting interoperability between organisations. By establishing a form of governance, stakeholders from different EU Members states can actively participate in decision-making processes and contribute their expertise to ensure the effective development and maintenance of ontologies. Community building also plays a crucial role in successfully implementing the guidelines. This document highlights the need to engage with relevant stakeholders in developing and adopting ontologies, including domain experts, data custodians, and technology providers.

This document focuses on harmonising and aligning initiatives to develop semantic assets across borders in the EU. The guidelines we describe are striking a balance between strict governance—essential for ensuring quality and trust—and flexibility, which allows for customization, domain-specific considerations, scalability to scope and budget, optimization, and innovative approaches. This balanced approach fosters both reliability and adaptability, supporting a common framework that is both robust and responsive to diverse needs across Member States.

4 Context and approach

4.1 What did we do in D2.1?

Deliverable D2.1[2] introduced the OSLO methodology. This governance framework enables effective collaborative decision-making and encourages the development of a stakeholder community integral to the success and broad adoption and reuse of semantic assets. This deliverable was characterised by describing the OSLO Process and Method[3] also to be used by other partners. This established a

[1] https://zenodo.org/records/8167337
[2] https://doi.org/10.5281/zenodo.8167337
[3] In Dutch: https://data.vlaanderen.be/cms/Proces_en_methode_voor_de_erkenning_van_datastandaarden_v1.0.pdf In English: https://github.com/belgif/review/blob/master/Process/201906-ICEG%20-%20process%20and%20method.docx

Funded by the European Union

preliminary set of guidelines on the creation of semantic assets promoting semantic interoperability between organisations and Member States.

Extensive documentation was prepared to promote transparency and ensure clarity in our approach. It outlines methodologies and processes that were partially adopted during Maregraph and the related technical and organisational challenges identified. These findings from D2.1 are now shaping the direction of the subsequent D2.2 phase as we attempt to outline and address the challenges identified. Specifically, key challenges for cross-border organisational interoperability were pinpointed, such as technical constraints, organisational hurdles, and the need to elevate awareness of the benefits of (semantic) interoperability in data exchange and reuse.

Building upon the groundwork established in D2.1, the MareGraph project has developed a high flexible methodology that can incorporate various steps from OSLO used in Flanders and eXtreme Design, a methodology available in the scientific literature and adopted by Italy in most of their semantic assets. The methodology consists of steps to undertake and principles that must be followed to construct semantic assets that can be shared according to open and FAIR principles. For each step, a set of recommendations for possible different implementations in a variety of scenarios are proposed.

4.2 Identified challenges

The collaborative effort between Belgium and Italy within the MareGraph project aims to establish a methodology for creating and managing semantic assets. This initiative centres on defining processes and guidelines to enable organizational interoperability in semantic asset creation. In MareGraph, we (1) initiated our work based on the OSLO guidebook, (2) applied these principles to biological taxonomies, specifically WoRMS[4] (World Register of Marine Species), and (3) coordinated efforts through CNR (Italy), leveraging their distinct background in practical methodologies and toolsets.

During this process of integrating diverse approaches, several challenges emerged. The table below outlines some of these challenges:

Challenge	**Description**
Endorsement group	A designated body should ratify the endorsed semantic assets in the member state from which the contracting party originates. However, the challenge lies in verifying the existence of a relevant party or body within each member state that can fulfil this endorsement role. It is important to note that, to the best of our knowledge, there is currently no equivalent European-level body responsible for endorsing ontologies.

[4] https://www.marinespecies.org/aphia.php

Funded by the European Union

Discoverability	Without a central registry or streamlined pathway to locate relevant "endorsement groups" across Member States, it becomes difficult to identify and discover domain-specific groups or standards in other regions, hindering efficient collaboration and alignment. This lack of visibility into existing frameworks and endorsed practices across borders creates silos, reducing opportunities for cross-border knowledge sharing and reuse of assets.
Governance on data standards	Establishing a governance working group on data standards is crucial in the central coordination and oversight of information standardisation efforts. There is still a need to find a European-level solution for this body, which ideally comprises experts from various member states.
Maintenance of recognised ontologies	Maintenance of recognised ontologies is essential once they have been adopted. Establishing the right responsibilities for conducting regular reviews, engaging stakeholders when needed, documenting changes, and fostering user feedback are key to ensuring the ongoing relevance and effectiveness of recognised ontologies.
Multilingualism	Multilingualism plays a crucial role in ensuring that semantic assets are comprehensible, usable, and accessible across various languages.
OSLO process	While developing ontologies, OSLO aims to reach a point where the ontology can be registered in its own standards registry. However, determining the appropriate timing to initiate this inclusion and addressing any discrepancies that may arise, such as variations in the development process and tooling or the required use of UML models for OSLO, become crucial considerations.
Scalability	Scalability is a key objective in developing a process and methodology for ontologies that other member states can effectively reuse. Addressing this challenge is closely intertwined with the issues surrounding the working group on data standards and the endorsement group. The goal is to establish a document that is as generic as possible, allowing for easy adoption by other member states.
Publication of the ontology and documentation	The challenge of where to publish the ontology and the accompanying documentation is closely linked to the 'OSLO Process' and 'Scalability' topics mentioned above. To overcome this challenge, it is crucial to establish a clear process for

Funded by the European Union

	determining the appropriate platform or repository for publication. Avoiding scattered repositories and ensuring easy accessibility and discoverability of the models and documentation are key objectives. By defining clear guidelines and identifying a centralised and easily accessible platform for publication (e.g., the international Linked Open Vocabulary - LOV catalogue, national semantic assets catalogues, and the possible European semantic assets catalogue), stakeholders can locate and access the different standards and ontologies and their documentation more efficiently.

4.3 Approach

This document outlines the findings and insights from discussions held within the MareGraph project, especially during its general assembly meetings in February and June 2024 and co-creation workshops in September 2024. Additionally, it draws from a workshop conducted during the SEMIC 2024 pre-conference, which focused on cross-border semantic interoperability. This workshop covered the guidelines for achieving organisational interoperability, presenting both the vision and lessons learned from the MareGraph project.

The workshop offered participants a platform to engage in productive discussions, share expertise, and exchange best practices on organisational interoperability. A panel discussion followed, featuring statements and questions from the audience (see Chapter 9 Annex). The workshop was recorded, and the recording is available via YouTube[5].

During SEMIC 2024, a survey was also conducted to gather insights into participants' current practices related to semantic interoperability and the creation and publication of semantic assets. The survey explored local governance structures, processes for defining semantic assets, tools in use, and the reuse of assets from other Member States. Participants were also invited to share their perspectives on EU collaboration, potential challenges, recommendations for a collaborative process, and ideas for a European repository of semantic assets (see Chapter 9 Annex).

5 Unifying principles

Local, regional, inter-federal and European governments frequently collaborate to enhance their services. In practice, a lot of data is exchanged between the various administrations. This data comes from different systems, may not be available in the same technical format, and does not necessarily follow the same semantics. Without agreements, achieving high-quality data exchange becomes exceedingly difficult. These agreements must be anchored as broadly as possible and, where relevant,

5 https://www.youtube.com/watch?v=3TXLDbpkRzQ&list=PLTqrUzsdFBKZmeePo_C8WnKrRGg6cG5K_&index=2

Funded by the European Union

lead to an ontology with a voluntary, 'comply or explain' or mandatory nature to avoid unnecessary data exchange costs.

When governments develop ontologies, the goals of the various stakeholders must be aligned, as well as inside the organisation's hierarchy. All parties involved must know the benefits of effective and efficient use of the ontologies. The stakeholders must be convinced of their usefulness and whether it benefits them directly. The development process in this document is based on international standards and ontologies, guarantees sufficient support among stakeholders, and provides for coordination with experts within their organisation and from the professional field.

The process and methodology outlined in the following chapters are grounded in fundamental principles for developing semantic assets, drawing from the **standards development guidelines of OpenStand**[6]. These principles are recognised as best practices and endorsed by organisations such as W3C, IEEE, IETF, IAB, and the Internet Society.

1. The ontology is developed in **collaboration with relevant stakeholders** and respects everyone's autonomy, integrity, processes, and intellectual property. Moreover, participation is free for all interested and informed parties.
2. The process aims to reach a broad consensus. Decisions are made fairly and transparently. Mechanisms are provided for appealing against decisions and for periodic assessment of the ontologies. Furthermore, all decisions and relevant documentation are made **publicly available**.
3. The developed ontologies strive for technical merit, interoperability, and scalability.
4. Ontologies and their relevant documentation are **made available for implementation** by all parties. Specifications are being developed that allow reasonable implementation.

Additionally, the FAIR principles—guidelines designed to enhance the Findability, Accessibility, Interoperability, and Reusability of digital assets— can also serve as overarching principles in this context. A key focus of these principles is for **machines to automatically** find, access, interoperate, and reuse data with minimal or no human intervention. In addition to the technical and structural principles outlined, **multilingualism** plays a crucial role in this process. Given the cross-border nature of data exchange, particularly in the European context, semantic assets must support multiple languages to facilitate seamless interaction and collaboration between different linguistic and cultural communities. Ensuring multilingual support is essential for maintaining the inclusivity and accessibility of semantic assets, allowing them to serve diverse user groups across regions and administrations effectively.

It is recommended that this **methodology for creating and developing semantic assets (chapter 6)** be based on principles mentioned above such as openness and transparency, the stimulation of high involvement, and the provision of necessary guarantees regarding stability, quality, and applicability. Moreover, semantic assets exist in a changing environment, so there must be room for managing changes and maintaining agreements and standards. The aim is a scalable process and method for developing and modifying ontologies and managing their life cycle.

[6] https://open-stand.org/

Funded by the European Union

Our proposed approach for improving interoperability incorporates achieving technical and semantic consensus and an end-to-end method that follows the Linked (Open) Data principles. This method ensures the maintenance of semantic assets within a functional public sector context. When this approach is applied across the different member states within the EU to develop semantic assets, collaboration and reuse are more effective as semantic assets undergo a process that includes the necessary alignments with relevant stakeholders and sufficient harmonisation with other standards and ontologies. The application of this approach can be broken down into four high-level steps:

1. Establish local governance: The standardisation process should be anchored in an existing governance structure or initiate a new one. This step is essential as it builds trust among various stakeholders and influences the adoption of data standards. This document describes this governance structure as the peer review community; it could consist of experts and stakeholders tasked with reviewing and providing constructive feedback on developing semantic assets.

2. Create a straightforward process for achieving semantic and technical agreements: The process should define the roles of different actors, timelines, and standard technical rules for creating high-quality semantic assets.

3. Implement an end-to-end method based on Linked (Open) Data principles: All decision records, discussions, and semantic assets should be publicly accessible, with the latter documented using a formal language based on Semantic Web standards (e.g., OWL, RDF). The method should include an implementation framework that ensures the traceability and alignment of semantic assets to suit various stakeholders such as policymakers, domain experts, analysts, and developers.

4. Co-create data standards: Agreements about the content of semantic assets should be achieved in open thematic working groups that may include domain experts from the public sector, private sector, and academia. These groups should follow the process and method within a governance framework.

Funded by the European Union

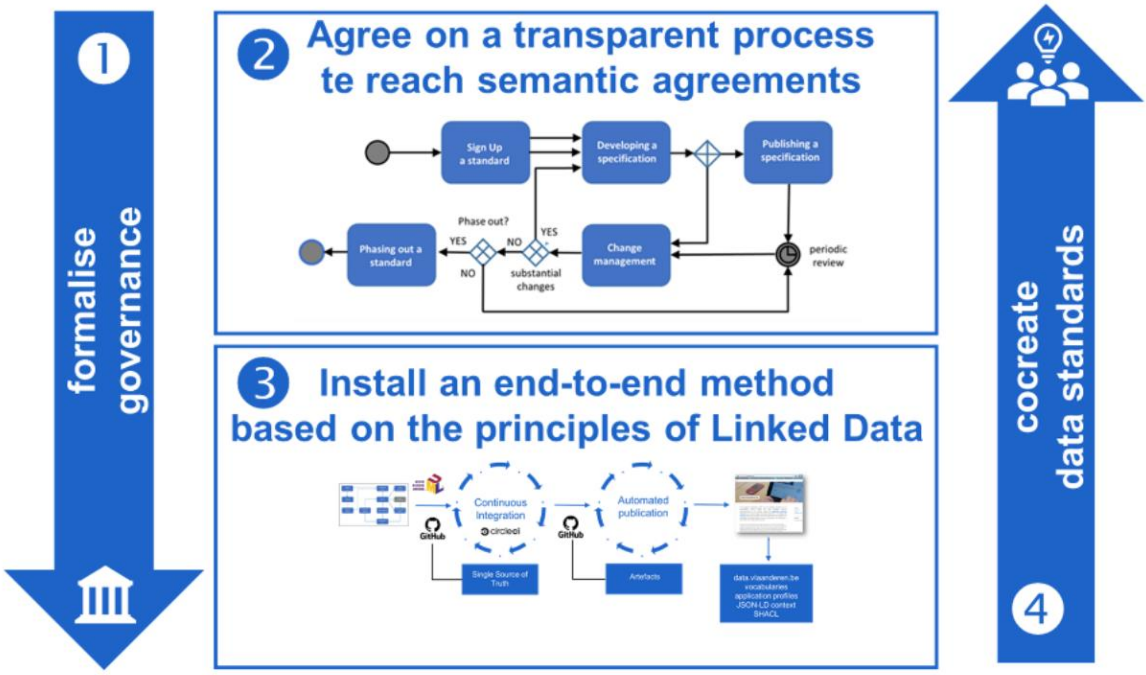

Figure 1: Raising interoperability in the public sector[7]

It is recommended to consider multiple aspects and phases when developing and maintaining semantic assets; e.g., the various actors' responsibilities, the change management, and the phasing out.

[7] Buyle, R. (2021). *Raising semantic and technical interoperability in the public sector*. Ghent University. Faculty of Engineering and Architecture, Ghent, Belgium.
Retrieved via https://biblio.ugent.be/publication/8712631

Funded by the European Union

6 Methodology for developing semantic assets

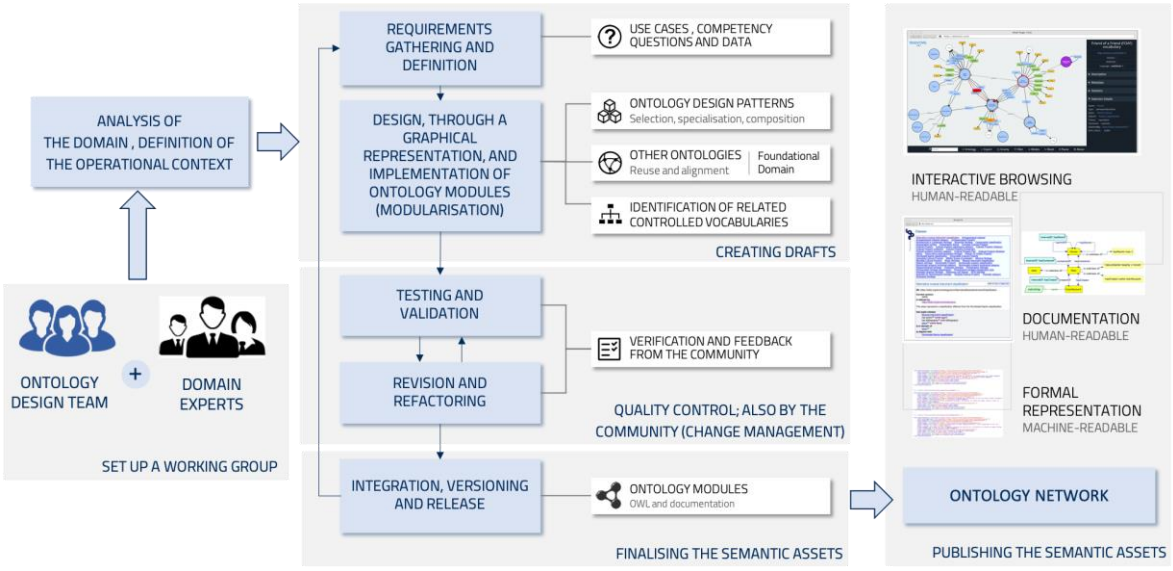

Figure 2: The proposed process for semantic assets creation and publication

Figure 2 shows the overall process we propose for the development of semantic assets; that is, for their definition, creation and publication.

The process consists of macro steps, each of which additionally broken down in tasks to be accomplished. The process can be viewed as an iterative approach, where some steps can be repeated in a loop based on the use and feedback coming from communities adopting the produced semantic assets for their data.

The macro steps we identified are listed in the following:

1. **Setting up a working group** and **setting up the operational context:** to work effectively, the group should be formed by both domain experts on one side, usually coming from communities of interest for the domain being considered, and, on the other side, ontology designers, with a specific expertise in developing semantic assets;
2. **Running a preliminary analysis of the domain** of reference;
3. **Creating drafts** of the semantic assets;
4. **Running quality control** (change management) operations, also by involving external communities;
5. **Finalising the semantic assets** so that to be ready to make them publicly available;
6. **Publishing** on the Web t**he semantic assets.**

The following sections will describe in detail the various steps. In each step we highlight the more important tasks to be carried out, some pre-requisites that are necessary to consider phasing that step of the methodology, some post conditions and possible recommendations.
All the steps of the proposed methodology have been designed also considering the answers to a questionnaire we have submitted during the SEMIC conference held in Brussels last June 2024.

Funded by the European Union

6.1 Setting up a working group and its operational context

PRE-REQUISITES

- Clear objectives for the project and mandate to proceed
- Identification of key stakeholders
- Establish supportive structure, which may include project management, dedicated facilitators, communication tools and documentation repositories (e.g. GitHub, SharePoint, ..)

This framework provides a structured, adaptable approach for forming and guiding a working group to develop an ontology. By defining clear objectives and fostering stakeholder collaboration, the group can achieve meaningful and standardized outcomes, flexible enough to meet diverse regional requirements.

6.1.1 Setup of the working group

The process begins by setting foundational objectives and defining the group's scope. It is crucial to articulate the mandate, objectives, and scope of the ontology project in a way that highlights the added value of the project and its relevance to all stakeholders involved. Additionally, it is important to identify the stakeholders who will contribute to and benefit from the ontology. This phase also includes organizing the necessary infrastructure to support the working group's activities. This infrastructure may include project management tools, dedicated facilitators, communication platforms, and documentation repositories, such as GitHub or SharePoint, depending on the group's specific needs.

6.1.2 Declaration of intent

The next step involves developing a declaration of intent, which serves as a guiding document for the project. This declaration should clarify the ontology's purpose and anticipated benefits, outline any existing relevant standards or ontologies, and identify the key stakeholders whose involvement is necessary to the project's success. The declaration may vary in formality depending on the country or organization's practices; in some contexts, it may be formally documented, while in others, it may be shared as an informal guiding reference that remains flexible and adaptable.

In the context of Maregraph this declaration of intent lead towards the development of a Working Group Charter[8]. This Working Group Charter, modelled on the W3C Standardization process, formalizes expectations for the working group's deliverables, covering:

- **Objective and Scope:** Define the thematic working group's purpose (e.g., ontology development for Domain X).

[8] https://www.maregraph.eu/files/Charter_Maregraph_OSLO.pdf

Funded by the European Union

- **Evaluation Criteria:** Set benchmarks for implementation, such as the number and nature (proof-of-concept or production) required for ontology approval.
- **Timeline:** Determine the duration of the working group (e.g., six months).
- **Deliverables:** Specify output types (e.g., specification documents or software components).
- **Milestones:** Identify expected completion dates, where known.
- **Approval Processes:** Define the internal criteria for deliverable approvals (e.g., unanimity or near-unanimity).
- **Dependencies:** Outline dependencies with other thematic working groups.
- **Meeting Logistics:** Clarify meeting locations and frequencies, as well as the date of the first in-person meeting, if applicable.
- **Communication Channels:** Establish tools (e.g., GitHub, mailing list, or Google Drive) for ongoing communication.
- **Intellectual Property:** Detail intellectual property and licensing terms.
- **Change Management:** Outline processes for post-publication updates and new releases, managing change and release cycles.

The Working Group Charter guides deliverable expectations, enabling the Peer Review Community to assess the ontology's relevance and application.

6.1.3 Stakeholder engagement workshop

Following the declaration of intent, it is beneficial to hold an initial workshop with relevant stakeholders, including both business and technical representatives. This workshop provides a platform for discussing the primary goals of the ontology, sharing relevant use cases, and identifying specific process-oriented needs. From this workshop, the group can create an action plan that serves as a roadmap for the project. This plan may include objectives, timelines, deliverables, approval processes, and dependencies. While some contexts may formalize this plan into a structured document, others may keep it as a flexible reference to accommodate various regional or organizational requirements.

6.1.4 Peer review and approval process

Once the action plan or project overview is developed, it should be presented to a peer review community or advisory board. This step allows for feedback and potential endorsement, which helps solidify the group's direction and provides valuable input from outside perspectives. During this stage, it is also essential to work with the peer review community to define compliance expectations for the ontology. Depending on national or organizational context, the ontology may be designated as voluntary, subject to a "comply or explain" requirement, or mandatory.

In the Maregraph project the Working Group Charter was then submitted to the Peer Review Community for review and approval. Upon endorsement from the peer review and public working groups, the ontology can be registered as "under development" in a relevant registry (e.g. OSLO Standaardenregister in Flanders[9]).

[9] https://data.vlaanderen.be/standaarden/

Funded by the European Union

6.1.5 Working group meetings

With the preliminary stages complete, the working group is ready to conduct regular meetings in phases, each with a specific focus.
Our survey results reveal variations in the structure of these working group meetings across member states. For example, Spain operates without specific steps, while Sweden follows a more structured approach with tabular annotations and specification documents. These examples highlight the importance of establishing adaptable governance structures to support diverse practices across regions. In Flanders a typical development process will require go through three phases, structured in at least 3 working group meetings. This approach was also followed within the Maregraph Project.
In the first phase, group members are introduced to the project structure and any existing use cases, followed by brainstorming sessions to explore other potential use cases and address identified information needs. The second phase is dedicated to in-depth discussions of the domain, where participants review draft specifications, address any unresolved issues, and compile a list of action points. In the final phase, the group works to finalize the specification, validate it against established use cases, and resolve any outstanding discussion points. After each meeting, editors or designated facilitators provide participants with a report, including updates on action items and an invitation to continue discussions on designated online platforms, such as GitHub.

6.1.6 Communication and engagement

Throughout this process, it is important to encourage active co-creation within the working group. Setting clear expectations around participation and contributions fosters a collaborative environment. Communication methods should also be adapted to participant preferences and regional requirements, ensuring that group members can engage easily and effectively. Transparency in decision-making is another critical component, as it helps establish trust and clarity within the group, allowing members to feel involved and informed about all significant developments.

RECOMMENDATIONS

- We recommend encouraging engagement by incorporating active co-creation practices within working groups
- We recommend setting clear expectations around attendance, engagement and contributions of participants
- We recommend adapting the communication strategies to preferences of the participants
- We recommend promoting transparency in decision-making processes

6.2 Running a preliminary analysis of the domain

PRE-REQUISITES

Funded by the European Union

- Access to domain documentation (ensure availability and accessibility of relevant documentation, reports and resources that describe the domain's informational landscape and data structures)
- Inventory of existing related semantic assets
- Domain expertise
- Analytical tools and resources

In parallel with establishing the working group, a preliminary analysis of the domain is conducted to gain a comprehensive understanding of the information needs and requirements. This process involves reviewing available documentation and analysing existing semantic assets relevant to the domain. By synthesising this information, a structured overview of the key informational needs is created. This overview provides insights into the current landscape of semantic assets, highlights gaps, and identifies alignment opportunities. The results from this analysis serve as a foundational reference, guiding further discussions and decisions within the working group to ensure alignment with established best practices and domain-specific requirements.

RECOMMENDATIONS

- We recommend to document gaps and opportunities along the way.
- We recommend engaging domain experts for insights (e.g. schedule brief interviews with domain experts to gather insights efficiently).
- We recommend considering mapping frameworks (such as SKOS[10]) to simplify the process of aligning overlapping or complementary semantic assets.
- We recommend using this analysis as a dynamic reference.
- We recommend preparing to present this analysis to the working group, with a special attention point to make this analysis relatable by incorporating real-world examples to illustrate abstract points

6.3 Creating drafts of semantic assets

The aim of this phase of the proposed methodology is to define and create the first drafts of the semantic assets.

PRE-REQUISITES

- It is recommended to set up a working group where domain experts and data modellers work together to achieve the goal of producing semantic assets for the representation of the domain of interest.
- It is recommended to define clear objectives and a plan for the releases.
- It is recommended to run a first preliminary high-level analysis of the domain of interest.

[10] https://www.w3.org/2004/02/skos/

Funded by the European Union

- It is recommended to choose since the beginning open platforms where the semantic assets and the overall modelling work can be shared transparently with communities. Examples include: Zenodo, GitHub or GitLab, these latter also for managing versioning and interact with users according to the foundational principles of the open-source development.

6.3.1 Requirements elicitation

In order to create semantic assets, through the use of Semantic Web standards like OWL and/or RDF, the first task to accomplish is to specify requirements; that is, collect functional requirements of the domain of interest that can be considered for effectively representing the entities, and relationships among those entities, of the domain of interest.
The elicitation of requirements that guide the modelling process can be done in different ways, according to the specific methodology being used.
In general, the first step common to consolidated methodologies available at the state of the art (e.g., OSLO, eXtreme Design, Linked Open Terms) is to list user stories or use cases with the collaboration of domain experts who know the insight of the data. This can also be done by using directly the data, which has to be analysed to understand its characteristics.
After producing the list of use cases/user stories, some of these methodologies diverge in the next step to carry out; however, most of them, and within the MAREGRAPH project, use cases/user stories are transformed into so-called competency questions. With the term competency questions (CQs) we mean a natural language representation of the semantic assets' commitments that drive their development. Specifically, CQs are a set of questions written in natural language that can be afterwards concretely transformed into queries on the data and that semantic assets, and in particular ontologies, should be able to answer effectively.

RECOMMENDATIONS

- We recommend to write the collected requirements in some specifications (e.g., in the form of a deliverable, slides to be further presented to the communities, or collaborative boards such as Mural, Figma, Miro to cite a few) to be publicly available: for the transparency principle toward communities, explaining the requirements can be useful to understand the rationales behind the proposed modelling.

6.3.2 Design of the semantic assets

Considering the specific type of semantic assets of the ontology, the next task to be carried out is its design.

Funded by the European Union

PRE-REQUISITES

- It is recommended to identify requirements for starting the modelling. They can be elicited in different ways, also simply looking at the data at disposition.

Based on the requirements, this step aims at developing the semantic assets in terms of ontologies and, if necessary, controlled vocabularies that can be used in conjunction to control the values of specific individuals of some entities represented in the ontologies.
To this end, there are four fundamental elements to be considered when producing the semantic assets:

1. The **analysis of existing ontologies** for the domain of interest that can be directly re-used, or indirectly re-used by defining semantic links or alignments to them;
2. The **analysis of existing so-called ontology design patterns (ODPs).** ODPs are reusable modelling solutions to solve recurrent ontological modelling problems that can be useful to reduce the arbitrariness of the ontology design. From the scientific literature, it has been proven that the use of ODPs when modelling can: i) reduce modelling errors; ii) help in detecting requirements that are not so evident; iii) help in improving the overall ontology quality[11]; iv) help in representing data that is then sounder[12]. In general, the re-use of ODPs across ontologies augments the possible semantic interoperability achievable across semantic assets;
3. The **application of the modularisation principle**, according to the different types of data that can be treated in a specific domain. When producing semantic assets, modularisation can be an alley to better maintain the semantic assets over time, also according to the different life cycles of the various data. Modularisation means that, in contrast to construct a big monolithic ontology, different ontological modules can be developed, possibly semantically linked together by re-using directly entities and relationships defined across the modules or by aligning all the modules to an upper-level or foundational ontology (e.g., DOLCE, DOLCE Ultra-light, etc.). This latter is an ontology that usually defines abstract and general entities that could be common to different domains or different types of data. In this sense, foundational ontologies can be used to relate specific entities in ontology modules back to more general abstract concepts, thus allowing for validation of the consistency of semantics between modules.
4. The **definition of a clear URI policy** to be adopted. URIs should be neutral to guarantee persistence over time and should not mix different languages. The use of the English lingua franca may help in the presence of "spoken URIs"; that is, URIs that are not only codes but "explain" the data being uniquely referred to; this lets them be understood by a wider plethora of possible users.

The development process for semantic assets consists of two important steps; namely, the design and production of the source code of the assets according to Semantic Web standards. In the Linked Open

[11] *Blomqvist E., Gangemi A., Presutti V. Experiments in Pattern-based Ontology Design, Proceedings of KCAP09, Los Angeles, ACM Press, 2009*

[12] *Paulheim, H. and Gangemi, A. Serving DBpedia with DOLCE – More than Just Adding a Cherry on Top. Proceedings of ISWC2015, the Thirteenth International Semantic Web Conference, LNCS, Springer, 2015*

Funded by the European Union

Terms methodology adopted in Spain, these two steps are also called Ontology Conceptualisation and Ontology Developing/Encoding, respectively.
Typically, the design or conceptualization starts using some graphical notation specification to draw the main or all elements of the semantic assets. The usage of a graphical representation can be convenient to convey to all possible users in an easy and intuitive way the semantics of the domain of interest, based on the collected requirements. For such a purpose, there are different tools and graphical specifications that are worth mentioning in the context of this work.
In the OSLO methodology, UML is the only notation adopted: the overall process of producing, and even publishing, semantic resources originate from it, although the expressiveness offered by a modeling language like UML is not comparable to that provided by standards such as OWL and RDF. Tools such as Enterprise Architect are then used to generate the UML diagram of the ontology.
Italy implements a different practice. There are several graphical specifications used by different ontology modelers such as Graffoo[13] and Graphol[14] from which the source code of an ontology can be derived automatically. However, this is not compulsory, and no-one prevents one from creating a drawing of the asset to give only the main intuition of the data representation, and then using specific ontology editors such as Protégé or Topbraid to actually create the overall ontology then released in the final version in the later stages of the methodology. Tools like draw.io or Eddy can be used to draw the ontologies in the context of the two specifications, respectively.
In Spain, Chowlk[15] is a visual notation specification used to draw ontology analogously to Graffoo and Graphol previously mentioned.
Finally, in Sweden, the Entryscape Models software component is used to generate semantic resources. No other unified suite or toolchain is mentioned. According to the answers provided in the questionnaire as reported in Chapter 9 Annex, the mechanism seems based on the use of the profile vocabulary PROF currently published as a working group note by W3C[16][1], a meta-layer ontology that was created to describe relations between specifications and profiles and the profile resources that define and implement application profiles.

Finally, when generating source code for semantic resources, the principles of multilingualism should be considered, generating labels and comments possibly in more than one language, always including English as the technical lingua franca.

The drafts of the semantic assets can be published at this stage in open platforms to potentially apply an agile approach and engage with communities since the beginning of the development.

RECOMMENDATIONS

- We recommend using as much as possible open graphical representation specifications and tools for designing semantic assets and apply a flexible approach when choosing such a specification and related tool, to accommodate different needs and practices.

[13] Graffoo specifications - https://essepuntato.it/graffoo/specification/
[14] https://www.mdpi.com/1999-5903/14/3/78
[15] https://chowlk.linkeddata.es/notation.html
[16] https://www.w3.org/TR/dx-prof/

Funded by the European Union

- We recommend to clearly specify a URI policy since the beginning of the development of the assets, where URIs are as neutral as possible, speaking about the data rather than a specific system or organisation. This provides stronger guarantees of the persistence of the URIs over time. The European Commission document "10 rules for persistent URIs"[17] can be used as basis for such a purpose.
- We recommend using foundational ontologies[18] to help verifying the consistency of the overall modelling and better support modularisation of the semantic assets according to different types of data to be represented.
- We recommend taking care of the multilingualism principle during the development of the semantic assets.

6.4 Quality control and feedback collection from communities

PRE-REQUISITES

- It is recommended to identify requirements for validating the semantic assets against them.
- It is recommended to have a draft of the semantic assets in RDF and/or OWL.
- It is recommended to publish the drafts produced in previous steps of the methodology in open platforms like GitHub or GitLab to potentially interact with interested communities.

When ontology modules are designed and implemented using the semantic web standards (e.g., OWL), the working group may proceed with testing and validation activities. This happens in the Italian and Spanish practices. Firstly, the modelers verify whether the ontology modules do not show syntactic, modelling or semantic errors and then that the modules fulfil all the requirements previously collected and listed. In this latter case, one technique can be to (i) convert the identified CQs into SPARQL queries and (ii) execute those queries on a data sample that is represented according to the target ontology modules.
There are different tools that can be utilized for achieving these objectives. One tool can be Oops! a web application that automatically detects 33 types of different pitfalls in OWL ontologies spanning from semantic and structural checks. This is used in Italy and Spain.

As far as the fulfilment of the identified requirements is concerned, a tool used in Italy is TESTaLOD[19], a web application that uses the TestCase OWL meta model[20], a reference schema for representing unit tests, with analogies with the unit tests in software development, a means for validating ontology as

[17] https://interoperable-europe.ec.europa.eu/collection/semic-support-centre/document/10-rules-persistent-uris
[18] Cassia Troiahn, "Is Your Data 6-Star?", *In the proceedings of ISWC (Demo/industry),* **2020,** https://ceur-ws.org/Vol-2721/paper573.pdf
[19] https://github.com/TESTaLOD/TESTaLOD
[20] http://www.ontologydesignpatterns.org/schemas/testannotationschema.owl

Funded by the European Union

well as data commitments. In such a schema, a unit test is modelled as a competency question expressed in natural language and associated with a corresponding SPARQL query.
In Spain, Themis[21] is an online tool that, similarly to TESTaLOD, can be used to execute test cases on semantic assets to verify whether requirements are satisfied or not.
In Flanders, the testing process extends beyond the UML model[22] to include checks on HTML, SHAQL, and RDF. Integrated into the OSLO toolchain[23], this testing runs automatically as part of the publication process, providing a comprehensive overview of publication quality. The results are publicly accessible, allowing the community to review outcomes and identify areas where semantic assets may need improvement.

The validation activity produces results. If the validation is not successful, as depicted in Figure 2, a revision and refactoring of the semantic assets is necessary, which may trigger in turn additional testing and validation sessions.

Parallel to these activities, drafts of semantic assets available in open platforms can be exposed to communities for feedback and comments, and for collecting potential additional requirements.
The interactions with communities are of utmost importance to guarantee a wider acceptance of semantic assets among the various stakeholders operating in the domain of reference.
Comments from the communities can be considered for revision and refactoring of the draft assets produced and for validation of the changes applied, as shown in Figure 2.
The collection of feedback from communities may happen according to different practices. The use of platforms such as GitHub or GitLab not only allow for semantic assets transparent versioning but also for enabling interactions with interest users.
Specific public consultations, as it happens in Italy and Belgium may also be a way to ask for feedback from stakeholders and could be exploited, as it has been done in the MAREGRAPH project experience.

RECOMMENDATIONS

- We recommend testing and validate produced semantic assets, also by producing data represented with those assets. This can also allow modelers to understand whether all the possible available data can be fully modelled with the assets being produced.
- We recommend opening the results of the previous phases of the methodology to external communities to collect additional feedback, comments and requirements. The form of engagement can be different spanning from structured public consultations to simply publishing the assets in open platforms that could also be properly configured to accept structured comments and feedback (e.g., issue tracking in GitHub).

[21] https://themis.linkeddata.es/
[22] In Flanders UML is used as a graphical representation of the semantic assets.
[23] https://github.com/Informatievlaanderen/data.vlaanderen.be2-generated/tree/dev4.0/report4

Funded by the European Union

6.5 Finalising semantic assets

PRE-REQUISITES

- Both validation and comments from communities have been considered and run.
- External existing semantic assets available in the Web of Data are identified.

In this step of the methodology, the semantic assets are finally built and integrated with other, in order to be ready to be published in production on the Web.
Before so, to guarantee that FAIR principles are met, semantic assets should be:

- Possibly aligned to external existing ontologies of the Web of Data, both standards and reference for the domain of interest. This guarantees a certain level of interoperability across semantic assets (interoperable principle).
- Provided of the necessary metadata that allow them to be discoverable in both national and international data catalogues (findable, accessible and reusable principles). There are a variety of meta-layer ontologies to describe metadata for semantic assets. In Italy, a specific application profile based on the European profile ADMS-AP, which in turn uses a W3C work named ADMS – Assets Description Metadata Schema[24] and DCAT, is used so that to enable harvesting mechanisms employed by the Italian national catalogues schema.gov.it. National assets catalogues are also available in Belgium with the purl.eu initiative, and in Sweden through the national data catalogue and the use of the W3C PROF vocabulary.
  At the time of this writing, the European Commission, in the context of the SEMIC community, is launching piloting activities with some Member States, offering national semantic assets catalogues, with the aim to create a pan-European catalogue and thus to promote cross-border semantic interoperability.

Through metadata specifications, it is also possible to convey the status of the development of the semantic assets. This can be very useful to inform communities about the sustainability over time of the produced semantic asset and be the discriminant to decide the type of re-use of that asset that one adopt.
For instance, in Italy, the used metadata profile includes a property named *status* whose goal is to advice whether ontologies (not controlled vocabularies) are stable or unstable. When an ontology is unstable it means that it can be under development/refinement or under revision of the governance structure that manages the national registry of semantic resources.
In Belgium, the status that determines the lifecycle of semantic assets is more articulated according to the W3C Recommendation Track, where an asset can be in development (a draft), in treatment (candidate to be a national standard), in use, in revision and phased out stages.
It is worth pointing out that this type of metadata can be very useful to be specified also in previous steps of the methodology, when semantic assets are firstly created and published in open platforms, keeping then track of their overall lifecycle since the beginning and not only when the assets are actually made available in official registries.

[24] https://www.w3.org/TR/vocab-adms/

Funded by the European Union

RECOMMENDATIONS

- We recommend linking or semantically align produced assets to existing ones, according to different approaches, directly or indirectly[25]. This contributes to provide interoperability across semantic assets.
- We recommend providing information about the stage in which semantic assets are during their lifecycle so that possible interested stakeholders can better understand the degree of sustainability over time of the assets when potentially re-using it.
- We recommend enriching semantic assets with metadata that describes them and their content. This contributes to better communicate the availability of the semantic assets to communities.
- We recommend first making sure that the semantic assets are correctly documented in official national semantic assets, also provided within national data catalogues, when available. However, to augment the possible plethora of re-users, international catalogues should be also considered. Examples are LOV – Linked Open Vocabulary[26], in the context of the MAREGRAPH reference domain, Bioportal[27] and Ecoportal[28].

6.6 Publication of semantic assets

PRE-REQUISITES

- Semantic assets should be ready to be published into production with both an associated graphical notation and/or their source code.

Publication in production is the last step that completes the overall methodology, although it does not exhaust the possible activities that can be done on the semantic assets as from this moment change management mechanisms must be considered to ensure their continuous maintenance.

Publishing semantic assets in the Web means being sure that they can be easily accessed through the Web so that anyone is able to access to its elements by using browsers or by technically querying them.
To this end there exist different tools that can be used. For instance, in Belgium through the UML diagram produced in early stages of the methodology, the HTML version of the semantic assets is produced to enable a web navigation of its elements.

[25] Presutti, V., Lodi, G., Nuzzolese, A., Gangemi, A., Peroni, S., Asprino, L. (2016). The Role of Ontology Design Patterns in Linked Data Projects. In: Comyn-Wattiau, I., Tanaka, K., Song, IY., Yamamoto, S., Saeki, M. (eds) Conceptual Modeling. ER 2016. Lecture Notes in Computer Science, vol 9974. Springer, Cham. DOI: 10.1007/978-3-319-46397-1_9

[26] https://lov.linkeddata.es/dataset/lov

[27] https://bioportal.bioontology.org/

[28] https://ecoportal.lifewatch.eu/

Funded by the European Union

In Italy, semantic assets and their single elements, are accessible through the Web via tools like LODE and WebVOWL for ontologies and LODView for single elements of the ontologies and controlled vocabularies. These tools use the source code of the semantic assets to work and not the graphical representation of it as in the case of Belgium.
In Spain, Widoco is used that integrates LODE and WebVOWL in one unique solution based on the source code of the semantic assets. In Sweden an HTML version is produced using EntryScape Model software component.
In MAREGRAPH, pyLODE, a python version of LODE is currently used to visualise the ontologies being published on the Web.

RECOMMENDATIONS

- We recommend enabling the necessary mechanisms for URI dereferentiation. This can be done by using a variety of tools available at the state of the art (e.g., LODE, Widoco, pyLODE). This guarantees an easier access online of the semantic assets for a more general public, not only technical (accessibility principle).
- We recommend when publishing into production to enable content negotiation, the mechanism that is used for serving different representations of a resource to the same URI to help the user agent specifying which representation is best suited for the user's needs.
- We recommend providing technical access to the sematic resources via a SPARQL endpoint, thus allowing developers to query their content.

7 Change Management

At any stage of its lifecycle, an ontology is open to feedback and necessary updates. Feedback is encouraged openly and is managed through a transparent, controlled process to ensure changes are traceable and minimize impact on current implementations. Feedback received during the "in development" or "in treatment" phases is promptly evaluated. Minor editorial changes can be addressed immediately, while more significant adjustments, whether minor or major, ideally follow the established process for semantic asset creation. For ontologies in any phase, substantial changes may require initiating a new working group (see 6.1) to properly manage and implement these updates.

8 Conclusion

The insights from this document focus on the vision for European collaboration and the guidelines on process and methodology for achieving organisational interoperability.

Funded by the European Union

This approach seeks to foster consensus among public administrations and promote semantic, syntactic, and technical interoperability. The organisation of this process is supported by a methodology that ensures clear communication and thorough documentation, enabling all stakeholders—project managers, business analysts, developers, and others—to implement the semantic assets effectively. The described process and method serve as a foundation for developing new semantic assets, adopting, or modifying existing ones, and potentially phasing them out when necessary.

Overall, the governance and guidelines outlined in this document seem to be needed in most contexts. However, our conversations have made it clear that not every EU member state can operate under a strict governance model like Flanders. Also feedback from SEMIC 2024 participants emphasized the value of a flexible, unified governance model for EU-wide collaboration. While a standardized governance model is beneficial, it should also allow for lightweight structures that accommodate both simple and complex semantic requirements. Therefore, flexibility is needed in governance to ensure its broad applicability. This document's guidelines and best practices are undoubtedly valuable and leave sufficient room for flexibility. Still, it will be unfeasible for many EU member states to follow them completely. To maintain the right level of flexibility, it is beneficial for member states to select the practices out of the guidelines that work for them and combine them according to their own needs without following the whole process and method. To support this approach, it is also essential to be inclusive towards a wide range of tools for ontology creation.

The importance of adopting a market-driven approach that balances costs and value when creating ontologies cannot be overstated. Unlike government attempts to standardise all data, a market-driven environment allows specifications to develop organically to meet specific needs. This approach enables natural selection, where the most effective and practical specifications gain traction and are adopted by others.

Perceptions of reusing ontologies revealed significant advantages and notable challenges regarding this topic. On the one hand, the reuse of ontologies is strongly advocated for its potential to substantially enhance interoperability. By leveraging existing ontologies, data is more easily shared and understood across different systems and borders. However, valid concerns and challenges are associated with reusing ontologies, particularly from a government funding standpoint. Developing new, customised ontologies often attracts specific funding and resources, whereas reusing ontologies does not offer these opportunities.

In summary, the MareGraph project has provided valuable insights into what a European collaboration towards achieving cross-border semantic interoperability could look like. The emphasis on inclusivity and flexibility in methodologies, along with a market-driven approach to semantic standards, highlights the need for adaptable and practical solutions related to ontology creation and adoption. Balancing these diverse needs and perspectives will be crucial for advancing interoperability across Europe.

Funded by the European Union

9 Annex

The following annex presents detailed findings from the SEMIC 2024 Pre-Conference survey and panel discussion, capturing the diverse practices, governance structures, and challenges faced by EU member states in semantic interoperability. These insights have been integrated into the preceding sections of this document, providing practical examples and supporting adaptable strategies for European collaboration on semantic assets.

9.1 Survey questions and feedback

The survey and participant feedback cover ten key topics, highlighting a range of approaches and experiences from member states in achieving semantic interoperability. The specific questions and feedback are listed below.

Question 1:
"Do you have any current practices related to semantic interoperability and semantic assets creation and publication? If so, please describe briefly."

To describe specifications, Spain uses Linked Open Terms Methodology[29]. Sweden uses publishing and Referencing Ontology for Formal Specifications (PROF), which can include application profiles, RDFS, UML diagrams, and human-readable documents.

Question 2:
"Do you have a local formal governance structure? If so, could you please briefly describe it?"

Governance structures seem to vary. Answers indicated that most countries' production of specifications is not strictly governed. For instance, Spain did not mention specific steps, and Sweden mentioned a diverse range of practices, including specification documents and tabular annotations.

Question 3:
"Could you briefly explain the main steps of the process you follow for defining semantic assets in your country, if any?"

Among the respondents, Spain and Sweden, there seem not to be any formal steps in place currently.

Question 4:
"Do you use specific tools in the process? If so, could you please list them?"

The tools used also vary. Spain uses a suite of tools like Chowlk[30], OOPS![31], and Ontoology[32], while in Sweden, tools are not unified, although EntryScape Models[33] is mentioned.

Question 5:
"Do you reuse existing assets from other Member States?"

It is indicated that existing assets are often being reused. In Sweden, it is mentioned that they declare reuse.

[29] https://lot.linkeddata.es/
[30] https://chowlk.linkeddata.es/
[31] https://oops.linkeddata.es/
[32] https://ontoology.linkeddata.es/
[33] https://entryscape.com/en/

Funded by the European Union

Question 6:
"Where do you publish the semantic assets you have produced? Do you have national catalogues? Do you use international catalogues (e.g., LOV, Bioportal, others)?"

Semantic asset publication occurs in national and international catalogues, with Spain contributing to the LOV catalogue and Sweden having a national catalogue accessible via their dataportal[34].

Question 7:
"What are your initial thoughts on the vision for EU collaboration we've presented?"

Participants from Sweden expressed support for the approach where catalogues are first harvested into national portals and then later to Europe's data portal. They also indicate that they believe the same should happen with specifications. They also believe extending the specifications from the European level or other countries to your own country is a challenge because it requires either harvesting back certain specifications or establishing a lookup mechanism.

Question 8:
"What are the biggest pitfalls to be addressed when moving towards European collaboration for semantic assets creation and sharing?"

It is crucial to have stable identifiers for specifications to enable their referencing from datasets. Additionally, non-semantic specifications must be accepted, which necessitates using the PROF vocabulary to express specifications in a broader sense beyond just RDFS, OWL, or application profiles.

Question 9:
"Do you have any suggestions regarding a European collaboration process for creating and sharing semantic assets?"

Sweden suggests that focusing on this is not realistic. Instead, they recommend concentrating on the technical requirements for expressing specifications across various levels of smartness and semantics. Emphasising the importance of stable URIs for specifications is crucial. Additionally, they advocate for a "live and let die" approach, allowing different specifications to compete and letting the best one prevail.

Question 10:
"What are your ideas regarding a European repository of semantic assets? What shall it include?"

Sweden has indicated that specifications should be expressed using PROF or, potentially a PROF-EU application profile. This profile could include additional properties to facilitate the search and identification of specifications within the European context. The emphasis should be on how application profiles and vocabularies can be articulated and integrated as parts of the specifications, particularly for those aiming to develop genuine "semantic assets."

9.2 Panel discussion

During the panel discussion at SEMIC 2024 Pre-Conference, several statements were presented for participants to indicate their agreement or disagreement, followed by discussions. It is important to

[34] https://www.dataportal.se/en/specifications

Funded by the European Union

note that only the statements and discussions relevant to the MareGraph project and this deliverable are presented here. The statements presented and the discussions were as follows:

Statement 1:
"Implementing a scalable, EU-wide system for creating and reusing semantic assets necessitates a unified (governance) and collaborative method (collaboration environment)."

Most participants agree on the necessity of a unified governance and collaborative method for managing semantic assets across the EU. However, one opposing opinion highlighted the diversity in complexity among different semantic asset creators. Heavy governance might be excessive for those with more straightforward needs; a more flexible approach focusing on broader specifications would be more appropriate.

It was emphasised that governance is essential, drawing attention to the MareGraph project as an example. However, a lightweight governance model is often more effective compared to Belgium's more robust governance, as public administrations in other member states, such as Italy tend to operate more independently. Differences in funding across countries also affect governance implementation, as some nations lack dedicated agencies to manage this responsibility.

The overall conclusion is that there is a need for a unified yet flexible governance model that accommodates both simple and complex semantic asset requirements. The effectiveness of governance models varies by country and is influenced by local administrative practices and funding capabilities.

Statement 2:

"Local, regional, national and international catalogues are crucial for maximising the sharing and reuse of semantic assets and should be integrated into the overall system".

Most participants agree on this; however, the statement was challenged by focusing on the term "catalogues." The overemphasis on catalogues could diminish the importance of self-publishing and the responsibility of maintaining and publishing semantic assets in one's own domain. Catalogues are helpful, but they are not essential for the sharing and reuse of semantic assets.

Copying data from a local catalogue to a European catalogue does not always make sense. Local solutions often contain nuanced ideas or metadata that might get lost when integrated into a higher-level catalogue. There is a risk of losing valuable context and details in the integration process.

Statement 3:

"Cross-border Data Spaces need a data co-existence approach."

The opinions on the need for a data co-existence approach in cross-border data spaces are more mixed, with a general lean towards agreement but with significant distribution in viewpoints.
The original presentation from which this statement was derived lacked a discussion on the importance of standards in this context. While standards would be beneficial, what percentage of data can realistically be standardised and what it means for data to be standardised in the first place should also be considered. The term "standard" carries much weight and might be better replaced with "specification." In a market-driven approach toward ontology creation, specifications are targeted to meet certain needs and, if successful, get adopted by others. This approach allows for a more flexible and adaptive framework where specifications can be advertised and adopted based on their value to the users.

Funded by the European Union

Statement 4:

"The technological solutions/components are (mostly) there, but ontology reuse is still tedious."

Most participants agree with the statement that while technological solutions and components are largely available, the process of reusing ontologies remains cumbersome. There are still numerous opportunities for leveraging new technologies. For example, Artificial Intelligence offers many use cases, such as desk research. This could potentially help address some challenges related to ontology reuse. However, the human factor will always play a crucial role in the process of ontology creation.

Funded by
the European Union