UNDERPIN

DATA SPACE FOR MANUFACTURING

Pan-European Data Space for holistic asset management in critical manufacturing industries

D2.1 UNDERPIN Requirements and enablers adaptation mid-term report, review of relevant regulatory and standardization landscape, plan for implementation of pre-standardization activities

Co-funded by the European Union

This project has received funding from the Digital Europe Programme under grant agreement No 101123179

Document information	
Project name	Pan-European Data Space for holistic asset management in critical manufacturing industries
Project acronym	UNDERPIN
Grant Agreement No	101123179
Start / Duration	1/12/2023
Project Coordinator	MOTOR OIL (HELLAS) DIILISTIRIA KORINTHOU AE
Deliverable	D2.1 UNDERPIN Requirements and enablers adaptation mid-term report, review of relevant regulatory and standardization landscape, plan for implementation of pre-standardization activities
Work Package	WP2
Authors	SWC
Dissemination level (PU = Public; PP = Restricted to other program participants; RE = Restricted to a group specified by the consortium; CO = Confidential, only for members of the consortium)	PU - Public
Type	Document, Report
Due date (M)	M9 **Actual delivery date** 30.08. 2024

Document history

Version	Date	Author(s)	Comments / Description
V0.01	May 2024	Martin Kaltenboeck (SWC)	Table of Content
V0.02	July 2024	Robert David (SWC)	UNDERPIN Architecture added
V0.03	25 July 2024	Anil Turkmayali (IDSA)	Expanded section5
V0.04	30 July 2024	Amela Kurtić (SWC)	Expanded section 4
V0.05	31 July 2024	Jakob Quabek (sovity), Odysseas Kokkinos (INNOV)	Expanded section 4
V0.06	02 August 2024	Eleni Politi (HUA)	Expanded and refined section 5
V0.07	05 August 2024	Aristotelis Ntafalias (MOH),	Expanded section 4
V0.08	05 August 2024	Giorgos Chrysokentis (WM)	Expanded Section 3 and Section 4
V0.09	08 August 2024	Dimitrios Drakoulis (INNOV)	Expanded section 5
V0.10	9 August 2024	Martin Kaltenboeck (SWC)	Expanded Sections 2,3,4,5
V0.11	12 August 2024	Vladimir Alexiev (OT)	Expanded section 5
V0.12	12 August 2024	Martin Kaltenboeck (SWC)	Finalised Sections 5,6,7,8 and reviewed whole document
V0.13	16 August 2024	Vladimir Alexiev (OT)	Expanded section 5
V0.14	22 August 2024	Victoria Katsarou (SPH)	Review of D2.1
V0.15	25 August 2024	Odysseas Kokkinos (INNOV)	Review of D2.1
V0.16	29 August 2024	Martin Kaltenboeck (SWC)	Refinements based on review results, final formatting, handover for submission
V1.0	30 August 2024	Konstantinos Chatzifotis (MOH)	Submission to EC portal

List of Authors and Contributors

Name (sorted by last name)	Organisation
Vladimir Alexiev	OT
Giorgos Chrysokentis	WM
Dimitrios Drakoulis	INNOV
Robert David	SWC
Martin Kaltenboeck	SWC
Sotiris Karampatakis	SWC
Victoria Katsarou	SPH
Odysseas Kokkinos	INNOV
Amela Kurtić	SWC
Aristotelis Ntafalias	MOH
Jakob Quabek	Sovity
Anil Turkmayali	IDSA

UNDERPIN has received funding from the Digital European Programme under grant agreement No 101123179.

Table of Contents

This project has received funding from the Digital Europe Programme under grant agreement No 101123179

List of Figures

This project has received funding from the Digital Europe Programme under grant agreement No 101123179

List of Tables

This project has received funding from the Digital Europe Programme under grant agreement No 101123179

Acronyms and Abbreviations

API	Application Programming Interface
CaaS	Connectors as a Service
DAPS	Dynamic Attribute Provisioning Service
DSSC	Data Spaces Support Centre
EDC	Eclipse Data Space Components
EU	European Union
GDPR	General Data Protection Regulation
HTTPS	Hypertext Transfer Protocol Secure
IAM	Identity and Access Management
IDS	International Data Spaces
IDSA	International Data Space Association
ISO	International Organization for Standardization
JWT	JSON Web Tokens
MS	Member State
OAuth2	Open Authorization
OWASP	Open Worldwide Application Security Project
RfP	Request for Proposals
TLS	Transport Layer Security
TRL	Technology Readiness Level
UI	User Interface
UC	Use Case
W3C	World Wide Web Consortium
WP	Work Package

Executive summary

This Deliverable titled "D2.1 UNDERPIN requirements and enablers adaptation mid-term report, review of relevant regulatory and standardization landscape, plan for implementation of pre-standardization activities" provides the summary of the results of the work carried out in UNDERPIN by month 9 (M9) of the project, with the main focus on the work that has been carried out in Work Package2 / WP2 (Requirements and Enablers implementation). Furthermore, this deliverable summarises and refers to work results of the work packages: WP3 (Data Space Operational Infrastructure) and WP4 (Use cases for Data Space demonstration and validation), while it also provides special areas of WP5 (regulatory framework) and WP6 (Standardisation), that are relevant for requirements elicitation and engineering.

The content of this deliverable is thereby:

a. Requirements elicitation for use cases and the UNDERPIN Data Space: section 2, with main results available in section 2.2
b. Technical requirements specification including architecture and components: section 3
c. The status of Data Space component integration: section 3.4
d. An overview of the regulatory and standardisation landscape relevant for UNDERPIN, as well as the plan of concrete activities regarding standardisation of the project: section 4.

The methodology of requirements elicitation and specification in regards of the UNDERPIN use cases and the UNDERPIN Data Space system is explained (interviews, workshops, desk research) and the related results are provided in detail.

The following 3 main use cases have been identified:

- UC1.1: Monitoring and predictive maintenance in the refinery
- UC2.1: Predictive maintenance in the wind farms
- UC2.2: Wind turbine blade repair prediction

The technical requirements identified and specified show a clear path along the European wide (and beyond) developments of Data Spaces, namely following IDS architecture principles and EDC implementations. The UNDERPIN Data Space is built upon the Eclipse Data Space Components (EDC) and makes use of Sovity's Connectors-as-a-Service (CaaS) offering.

This basic Data Space system will be expanded by new features and services, that are based on UNDERPIN partners software products, namely OTs GdB, SWCs PoolParty Semantic Suite, and ARCs knowDE, and finally AITs Smart Contracting system. The interplay and integration approach is described in the UNDERPIN system architecture and components description in full detail. The main goal of this expansion is to provide a "Data Space Semantic Layer" that enables (semantic) interoperability in regard to metadata and data inside the UNDERPIN Data Space, as well as with other relevant European Data Spaces.

Finally, the document provides a comprehensive overview of the regulatory framework and standardization environment within which UNDERPIN operates, both in terms of use cases and the Data Space. The report explains the work already carried out regarding standardization and outlines future plans in this area. Semantic interoperability inside one and between several Data Spaces covers the biggest part of standardisation activities of UNDERPIN.

1 Introduction

This introduction section is split into two main areas: (i) a short introduction into the overall topic of Data Spaces and the UNDERPIN project as well as into the topic of this deliverable and (ii) the subsection "Relation to other work packages and deliverables" of the deliverable D2.1 on hand, to allow easy further reading of deliverables already in place as well as of future deliverables of UNDERPIN.

1.1 Data Spaces – UNDERPIN – Deliverable D2.1

The field of Data Spaces is continuously growing and emerging at the moment and more and more Data Spaces are available to be used to exchange data in a secure and trusted environment inside a domain or a set of use cases, as well as between domains and use cases, to exploit the full potential of data trading and –exchange in Europe and thereby foster the digital single market and the European data economy. For information about this emerging field of Data Spaces we recommend the Data Space Radar Report of the International Data Space Association, IDSA (https://internationaldataspaces.org/data-space-radar-report/). To best develop this growing and emerging market of Data Spaces, the continuous exchange of know-how and collaboration with similar international initiatives, and data sharing activities and technologies are also crucial among others, to gain best possible development and success for the European data economy.

Through the UNDERPIN project (UNDERPIN: The Pan-European Data Space for holistic asset management in critical manufacturing industries), European manufacturers, their value chain business ecosystems (such as machine tool manufacturers, integrators, machine vendors, maintenance service providers, remanufacturers, refurbishing, reuse, repairing and recycling companies), governmental and public, research and civil society stakeholders will be equipped with a coherent Data Space solution compliant to EU standards and GAIA-X guidelines (https://gaia-x.eu/, accessed 08/2024) encompassing secure structure and tools to accommodate data sharing among partners, and to enable data analysis for the benefits of the stakeholders in terms of improved company operations following the implemented scheme of industrial data sharing. The UNDERPIN Data Space will thereby provide a cross-organisational data sharing and -exchange solution that is secure and trusted and ensures data sovereignty, with a strong focus on the interplay of SMEs and large industry players to enable both to improve products and services. Furthermore, one of the UNDERPIN overall objectives is to incorporate the UNDERPIN outcomes within the global European Standardization landscape in the Manufacturing Data Spaces field in the area of Industrial Data Sharing. The aim is to develop pre-normative standards which can be later on adopted by the stakeholders as standards and / or can be taken up by the Technical Committees to be included in later standardisation efforts. An additional aim is to contribute to the formation of regulatory provisions forcing the use of those standards.

UNDERPIN addresses the vision of a European Digital Single Market, and thereby the idea of a European-wide Data Space. In more detail UNDERPIN will establish a European-wide Data Space for Manufacturing in the fields of dynamic asset management and predictive / prescriptive maintenance. This Data Space will – as a starting point - make use of the existing Motor Oil ecosystem (Motor Oil / MOH, Project Coordinator) in the field of dynamic asset management and predictive / prescriptive maintenance, that is an existing "embryonic Data Space" including data sharing technologies and - mechanisms as well as metadata- and data management tools and mechanisms for both: (i) internally at MOH between several divisions and groups, as well as (ii) between MOH and third parties like partners, suppliers and vendors.

This project has received funding from the Digital Europe Programme under grant agreement No 101123179

The UNDERPIN project will expand this existing MOH ecosystem to a full-blown European Data Space for Manufacturing through activities in the following fields of action:

A. Technology
B. Interoperability
C. Legal Framework
D. Demonstration and adoption
E. Scale out to become a European-wide Data Space
F. Communication, Dissemination and Scale Out

Whereby the Work Package 2 mainly focuses on A. and B. of the given UNDERPIN work streams above, and by this supporting all the other areas C. to F.

(A) technology-wise: bringing the UNDERPIN Data Space to a mature Data Space level, by making use of well-proven technologies and software components that are already following IDSA and GAIA-X guidelines, standards, and recommendations, as well as making use of W3C, ISO and industry-specific standards.

This technology upgrade and expansion will be managed by mainly making use of existing Data Space Technologies and components to be expanded along the needs / requirements of UNDERPIN and by new technology and software components. For this UNDERPIN technology partners (ARC, SWC, OT, AIT) bring ready-to-use technology and software to the UNDERPIN Data Space with a strong focus on semantic technologies, predictive maintenance, and analytics. All of them with Technology Readiness Level of at least TRL: 7, but many of them TRL:9 and all successfully used in production by hundreds of customers worldwide.

Finally, some additional data management and analytics tools will be integrated directly taken from the current Horizon Europe project DataBri-X (see: https://databri-x.eu/, accessed 08/2024), that started in 10/2022 and develops a data- metadata and AI toolbox especially made for Data Spaces (and beyond).

The UNDERPIN cloud environment will be provided and operated by the project coordinator and Data Space operator MOH as a trustworthy data Broker to continue operations beyond the end of the project.

With the described technology approach, UNDERPIN (B) enables interoperability for both: inside the Data Space regarding metadata and data, as well as in respect to interoperability with other Data Spaces that are using the same (IDSA; GAIA-X etc. applying semantic interoperability and data interoperability methods) approach to support the idea of a European-wide Data Space. This interoperability approach is reflected in the development and integration of a Semantic Layer for Data Spaces, that is described in detail in the section 4.1.5 of this document.

1.2 Relation to other work packages and deliverables

This deliverable D2.1 refers to work already carried out and partly summarised in already submitted deliverables, as well as to work results carried out in the work packages: WP3 (Data Space Operational Infrastructure) and WP4 (Use cases for Data Space demonstration and validation). Moreover, it provides special areas of WP5 (Business plan and sustainability) and WP6 (Communication, Dissemination and Scale Out), that are relevant for requirements elicitation and –engineering.

The most important past deliverables to take into account in relation with D2.1 are:

This project has received funding from the Digital Europe Programme under grant agreement No 101123179

- D1.1 Project Handbook and Risk Register, Data Management Plan: here the section including the Data Management Plan (DMP), submitted in M3 (February 2024).
- D4.1 Use case planning report: full deliverable, submitted in M8 (July 2024): the full deliverable is of relevancy.
- D6.1 Integrated Communication, Exploitation and Dissemination Plan (initial design): here the section about Orchestration, Standardisation, and Certification (Task T6.4), submitted in M6 (May 2024).

The most important upcoming deliverables to take into account in relation with D2.1 are:

- D3.1 UNDERPIN Data Space infrastructure, mid-term deployment and integration report, to be submitted in M12 (November 2024)
- D1.2 Periodic Activity and Technical Report, to be submitted in M12 (November 2024).

2 Requirements Elicitation and Specification

This section explains the chosen and implemented methodology for requirements engineering that consists of (i) the methodology of the requirements elicitation and (ii) the results of the requirements specification of UNDERPIN, that took place in WP2, along the two tasks of the WP, namely: Use Case specification (Task 2.1) and Requirements Specification (Task 2.2). The Technical Requirements Specification is to be found in the next chapter.

On 25-26 January 2024 a two-days technical workshop with the relevant stakeholders (17 participants in total) took place in Vienna at the partner AITs premises. The workshop was prepared and facilitated by SWC as WP2 lead. On those two days the use cases and the related requirements were discussed, furthermore a basic understanding of the UNDERPIN Data Space was developed, Data Space basics were presented and discussed, and the base requirements for the UNDERPIN Data Space were also discussed, specified and formulated.

In parallel a WP 2-3-4 call series (all technical WPs) was set-up and organised by SWC with 3-weekly calls to report on the status of work, and plan next steps and align on the overall time plan and the interplay with other WPs.

With the participation of the partner OT the technical lead was handed over from SWC to OT and the call series was switched to WP specific monthly calls for single WPs WP2, WP3, WP4.

The definition of the aforementioned WP2 related tasks, as well as the description of work carried out and the related results, are as follows:

Task 2.1

Definition and detailed analysis of vertical use cases and applications, baseline and corresponding target KVIs and KPIs, review of the regulatory and standardisation landscape, and service requirements

In this task we will provide a detailed definition of the Pilot Use Cases which will be implemented and demonstrated in WP4. This includes a thorough description of the scenarios, the context in which they take place, a list of pre-requisites and assumptions (technical and non-technical), and a description of all needed scenario-specific infrastructure. Positioning with respect to the project technical objectives and enumeration of demonstrated functionalities through the use cases will also be provided, as well as target KPIs, KVIs and expected service requirements for the use cases.

Remark: the main and full results of T2.1 are available in deliverable D4.1 that has been submitted in M8 (July 2024) of the UNDERPIN project. The summarised results are provided below in this document.

A review of the relevant Regulatory and Standardization European Landscape on industrial data sharing – in close alignment with T.6.5 Orchestration, Standardization, and Certification and T5.1 Legal Framework – will be carried out. The task also includes assessment and analysis of all applicable aspects, laws, regulatory legislation and available standards from a Worldwide perspective related to devices, technologies, processes and methodologies related to industrial data sharing.

Remark: the results are available in Section 4 of this document.

The work performed within the context of Task 2.1 primarily focused on the definition and preparation of the use cases that will be trialled within the duration of the UNDERPIN project. The use cases offer a means of identifying and analysing potential requirements of the involved parties, while also acting as a showcase for the implementation of UNDERPIN. On a higher level, successful use cases are key to generating value inside the Data Space, and, subsequently, provide the Data Space with tangible results that can be utilized to attract new stakeholders.

The methodology for the definition of the use cases and the relevant requirements used the following steps:

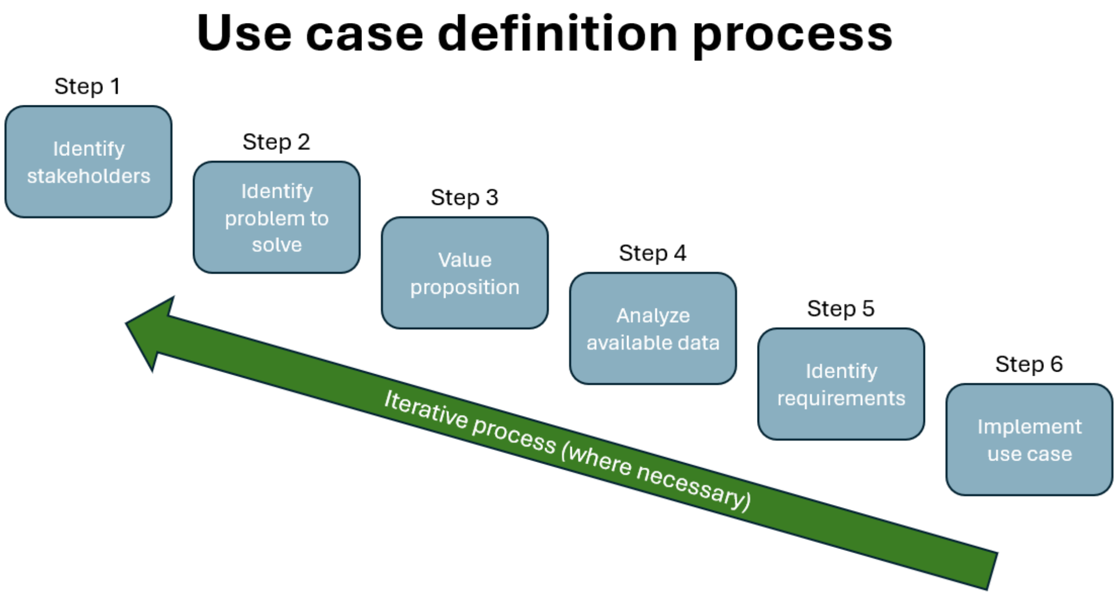

Figure 1: Use case definition process

1) **Stakeholder identification**: The first step of the process related to identifying the stakeholders that will participate in the use case(s) and the specific actors within them, as well as external stakeholders that are be interested in the results and thereby can be onboarded at a later stage.
2) **Problem to solve**: After identifying the stakeholders, the next step pertains to coordinating with them in order to identify potential problems that could be solved through the Data Space, analyse the current state of play and assess possible solutions.
3) **Value proposition**: Once the problem and potential solutions have been identified, examine the value the use case can offer to both the involved stakeholders as well as the Data Space itself and determine the expected outcomes.
4) **Data analysis**: A high-level analysis of the relevant data is performed, examining aspects such as availability and quality, so as to assess whether it is sufficient in order to achieve the expected outcomes outlined in the previous step.
5) **Requirements identification**: Once a clearer picture of the use cases has been created through steps 1-4, identify the actions that need to be performed and the related requirements that emerge throughout the process.

Co-funded by the European Union
This project has received funding from the Digital Europe Programme under grant agreement No 101123179

6) **Use case implementation**: While not strictly a part of the use case identification, the last step pertained to the implementation of the use case. This allows for evaluation of the results from the previous steps and highlight any necessary potential changes.

It must be noted that for the use cases to achieve the state goals and offer as much value to the Data Space and the stakeholders as possible, the process cannot be rigid and monolithic. To that end, feedback can be derived at any point of the process and integrated into previous steps to further refine the use case.

Task 2.2
Requirements, analysis and specifications of the end-to-end UNDERPIN Data Space

The aim of this task is to lay the foundations for subsequent development work in WP2 and WP3 by comprehensive requirements specifications to realise the evolution of the MOH embryonic Data Space to a technology-wise full blown, scalable and standards-compliant Data Space solution. The specifications are based on the results of T2.1 as well as they include functional and non-functional requirements, the final architecture, and a validation and deployment plan which will be executed in WP3 (and WP4).

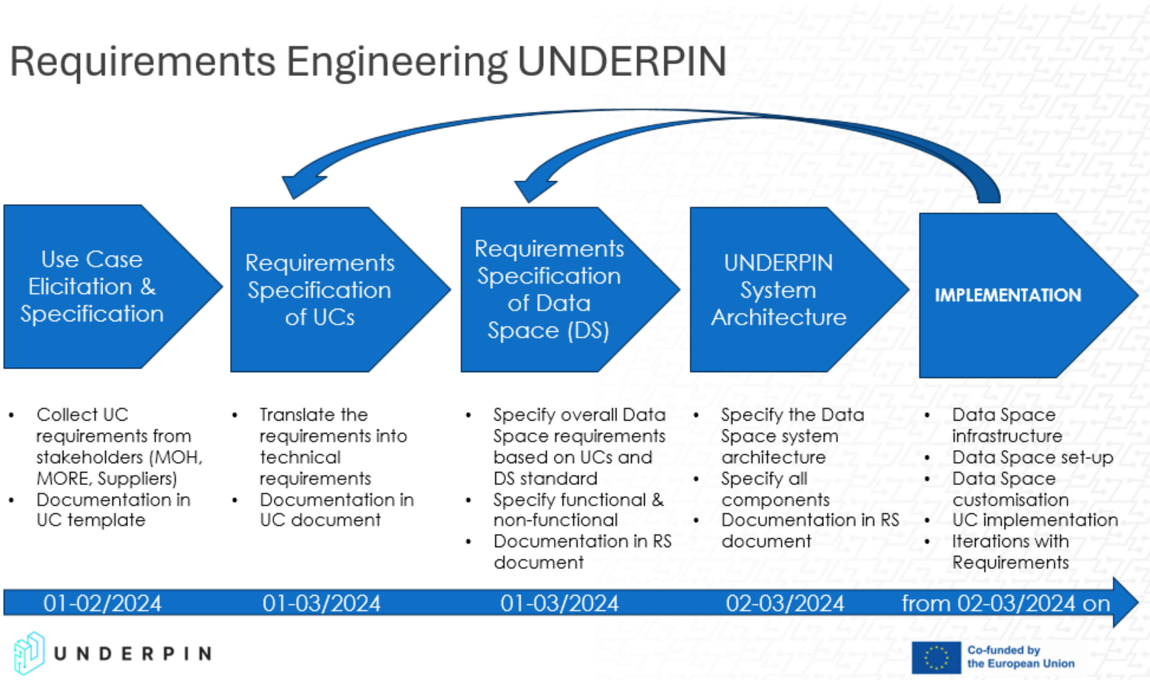

Figure 2: Requirements Engineering in UNDERPIN

In this task also the selection of the UNDERPIN technical Data Space implementation partner (and subcontractor) has been managed in the form of a **Request for Proposals (RfP)** that has been sent out to 4 candidates (namely: Sovity (DE), Nexyo (AT), I2CAT (ES), and NTT Data (ES team), that all are well known for their technical implementation work on Data Spaces and their knowledge in the field. The RfP process took place in 01-03/2024.

From the feedback of the identified candidates two of these have been short-listed: Sovity and nexyo and were invited to send a full proposal (RfP response) and both entities have been invited to present their approach and proposal. Please see **Annex B** of this document for the RfP document (requesting an RfP response). The RfP presentation and Q&A calls took place on 22 February 2024 (nexyo) and 23 February 2024 (Sovity) in the form of online meetings including presentations, demos and question & answering session.

Face to face meetings with both short-listed parties took place in Darmstadt at the Data Space Symposium 2024 (that took place 12-14 March 2024, see: https://www.data-spaces-symposium.eu/).

After this the UNDERPIN consortium aligned internally in the form of several calls to discuss the topic, and the consortium clarified additional and new upcoming questions with the two candidates (via emails and Q&A documents) and finally decided for Sovity.

The final reasons for this decision were as follows

- **Sovity** provides a full-blown Data Space edition (Sovity community edition) based on EDC components, and Connectors as a Service (CaaS) that are maintained and updated continuously along the respective developments in EDC connectors. Furthermore, Sovity has in depth experience in the set-up and maintenance and expansion of Data Spaces already in production (e g the German Mobility DS) and finally Sovity plays a central role in the working groups and standardisation work of the Data Space ecosystem and thereby can also support UNDERPIN in regard to important Data Space networks.
- **nexyo** has more focus on customisation and the nexyo data hub system is not in full production yet (although also based on EDC) and nexyo acts with a clear focus on Austrian DS ecosystem, whereby UNDERPIN has a European focus.

Thereby the contracts for A) architecture and technical consulting work and B) Connector as a Service have been negotiated and concluded and Sovity is the official subcontractor of UNDERPIN, already providing consulting and technology to the project.

2.1 Preparation and setup of requirements elicitation

2.1.1 Stakeholder and topics identification

At a high level, the key stakeholders to be involved in the use cases as well as the main topic were already identified during the inception of the UNDERPIN project. More specifically, MOH's refinery and MORE's wind farms were identified as the key stakeholders, with predictive maintenance in the manufacturing sector emerging as the key topic. AIT was also identified as a key partner for the use cases, based on their expertise in predictive maintenance algorithms.

An initial set of use cases was identified for the refinery and the wind farms during the technical workshop that took place in M2 of the project in Vienna, with the participation of the key stakeholders. These were primarily focused on the topics of predictive maintenance and dynamic asset management, taking into account the current state of play at the refinery and the wind farms and identifying the specific needs that could be addressed through the UNDERPIN Data Space. A first look at the nature and characteristics of the available data was also included in the process.

2.1.2 Stakeholder involvement

At a subsequent step, several specialized meetings were held with the participation of subject experts from the refinery and the wind farms for a more detailed evaluation of the proposed use cases. The scope of the use cases was refined through an iterative process, while certain use cases were excluded, for one or more of the following reasons:

- Insufficient data to support the scope of the use case
- Lack of substantial value to the users or the Data Space
- High level of complexity leading to difficulty of implementation within the scope of the project
- Legal implications as of the included data

2.1.3 Use case identification

Following the use case identification process described previously, the following three use cases have been identified:

Table 1: UNDERPIN use cases

Identifier	Title	Pilot
UC1.1	Monitoring and predictive maintenance in the refinery	Refinery
UC2.1	Predictive maintenance in wind farms	Wind farms
UC2.2	Wind turbine blade repair prediction	Wind farms

An overview of each individual use case is provided in the **Annex A** of this document. The reader should also take into account the information provided in the deliverable D4.1: *Use case planning report* for a detailed description of each use case and its processes. To avoid duplication of information we have omitted D4.1 information and results here.

2.2 Use Case Requirements and Specification

After the use cases were defined, as outlined in the previous section, a set of requirements was established for each use case, split into two categories:

- Business
- Functional

Due to the similar nature of the use cases and the shared organizational and operational procedures of the key stakeholders, a lot of the requirements are near identical, however they are included in each use case for the sake of completeness.

It should be noted that the requirements presented in this section are a preliminary set based on the specific needs of the involved stakeholders at the current state of the use cases. As the implementation of the use cases proceeds, new requirements or alteration to the existing ones could emerge. Additionally, should more stakeholders participate in the Data Space, potentially leading to an expansion of the scope of the use cases, these requirements could be enriched and/or additional requirements could emerge. Any such changes will be documented in the updated version of this report, in deliverable D2.2: *UNDERPIN requirements and enablers adaptation final report,* due on M18.

2.2.1 UC1.1: Monitoring and predictive maintenance in the refinery

From the perspective of the refinery, the following business requirements emerge from this use case:

- Predictive maintenance: Develop a predictive maintenance procedure that will enable the streamlining of maintenance operations, minimizing downtime and costs.

This project has received funding from the Digital Europe Programme under grant agreement No 101123179

- Equipment monitoring: Near real-time monitoring of equipment performance versus normal operating conditions to detect abnormal behaviour and apply corrective actions without interrupting operation.

In terms of functional requirements, the following key requirements have been identified:

- Data collection automation: Currently the data is manually downloaded from the DCS systems before entering the Data Space, which is not sufficient for the efficient implementation of the use case and needs to be automated. While this issue will most likely have to be solved within the user’s environment, it is included as a requirement as it highlights a potential issue that future stakeholders might also face.
- Streaming of data: The data from the refinery equipment will need to be streamed in hourly batches with a granularity of 5 mins to the predictive maintenance tool.
- Data pre-processing: A data cleaning procedure will need to be performed in order to identify and address errors in the datasets, such as inaccurate or incomplete data.
- Visualization: In order for the results of the predictive maintenance tool to turn into actionable information, they will need to be presented in a user interface/dashboard that can be accessed from the relevant parties (e.g., maintenance department) through relevant credentials.

An additional requirement for data anonymization has been identified, however, since anonymization needs sector-specific input and cannot easily be generalized, the intent is for it to be performed within the environment of the user before the data enters the Data Space.

2.2.2 UC2.1: Predictive maintenance in the wind farms

While this use case shares a lot of similarities with UC1.1, maintenance in the wind farms is not performed in-house, but instead through a contractor, which alters the business requirements from a user perspective:

- Maintenance benchmarking: Develop a predictive maintenance procedure that will allow the user to benchmark the quality of the maintenance services offered by the contractor.

In terms of functional requirements, the following key requirements have been identified:

- Data collection automation: Currently the data is manually downloaded from the SCADA systems before entering the Data Space, which is not sufficient for the efficient implementation of the use case and needs to be automated. While this issue will most likely have to be solved within the user’s environment, it is included as a requirement as it highlights a potential issue that future stakeholders might also face.
- Streaming of data: The data from the wind farms will need to be streamed in daily batches with a granularity of 10 mins to the predictive maintenance tool. The frequency of the batches is subject to change, depending on the outcome of the previously described requirement.
- Data pre-processing: A data cleaning procedure will need to be performed in order to identify and address errors in the datasets, such as inaccurate or incomplete data.
- Visualization: In order for the results of the predictive maintenance tool to turn into actionable information, they will need to be presented in a user interface/dashboard that can be accessed from the relevant parties (e.g., maintenance department) through relevant credentials.

Co-funded by the European Union

This project has received funding from the Digital Europe Programme under grant agreement No 101123179

An additional requirement for data anonymization has been identified, however, since anonymization needs sector-specific input and cannot easily be generalized, the intent is for it to be performed within the environment of the user before the data enters the Data Space.

2.2.3 UC2.2: Wind turbine blade repair prediction

This use case acts as a subset of UC2.2 and largely shares the same requirements. The core, however, business requirement is somewhat different:

- Blade damage identification: Monitoring of the state of wind turbine blades and timely identification of potential blade damage through the use of a predictive maintenance model, which will be used as feedback for ad hoc visual inspections.

The functional requirements are identical to those of UC2.1.

2.3 Data Space Requirements and Specification

This chapter provides the information about the UNDERPIN Data Space (core system) requirements identified via the different methodologies described in section 2.1.

UNDERPIN hereby fully follows the trends, recommendations and standards in place. These are dynamically developing in an emerging technology field. Namely the recommendations of:

- the International Data Space Association and the International Data Space recommendations (IDSA / IDA, https://docs.internationaldataspaces.org/knowledge-base, accessed 07/2024),
- GAIA-X (https://gaia-x.eu/), accessed 08/2024), and
- the Eclipse Data Space Components (EDC, EDC – Eclipse Data Space Components, https://projects.eclipse.org/projects/technology.edc, accessed 08/2024), as well as
- existing ISO Standards (https://www.iso.org/, accessed 08/2024), and
- W3C Recommendations (World Wide Web Consortium, https://internationaldataspaces.org/, accessed 08/2024 | for example ODRL for digital rights management, and others).

Figure 3 Infographic is depicting the elements and components of an IDS-compliant Data Space. It represents a conceptual framework by the International Data Spaces Association for data exchange in a "Data Space." It outlines the structure and flow of data among various stakeholders, and the related technical components:

Data Owner: The initial source of data, responsible for managing data and defining usage policies.

Data Provider: Facilitates the provision of data to the Data Space, integrating with an IDS Connector to ensure secure and policy-compliant data sharing.

IDS Connector: A critical component that acts as a secure interface between data providers and data consumers. It enforces data usage policies and ensures data integrity and security during transmission.

Data Space: The central area where data is exchanged and processed. It supports the entire data value chain, enabling secure interactions among participants.

Co-funded by the European Union

This project has received funding from the Digital Europe Programme under grant agreement No 101123179

Data Consumer: The recipient of the data, which can be used according to the pre-defined usage policies. This entity also interacts through an IDS Connector to maintain security and compliance.

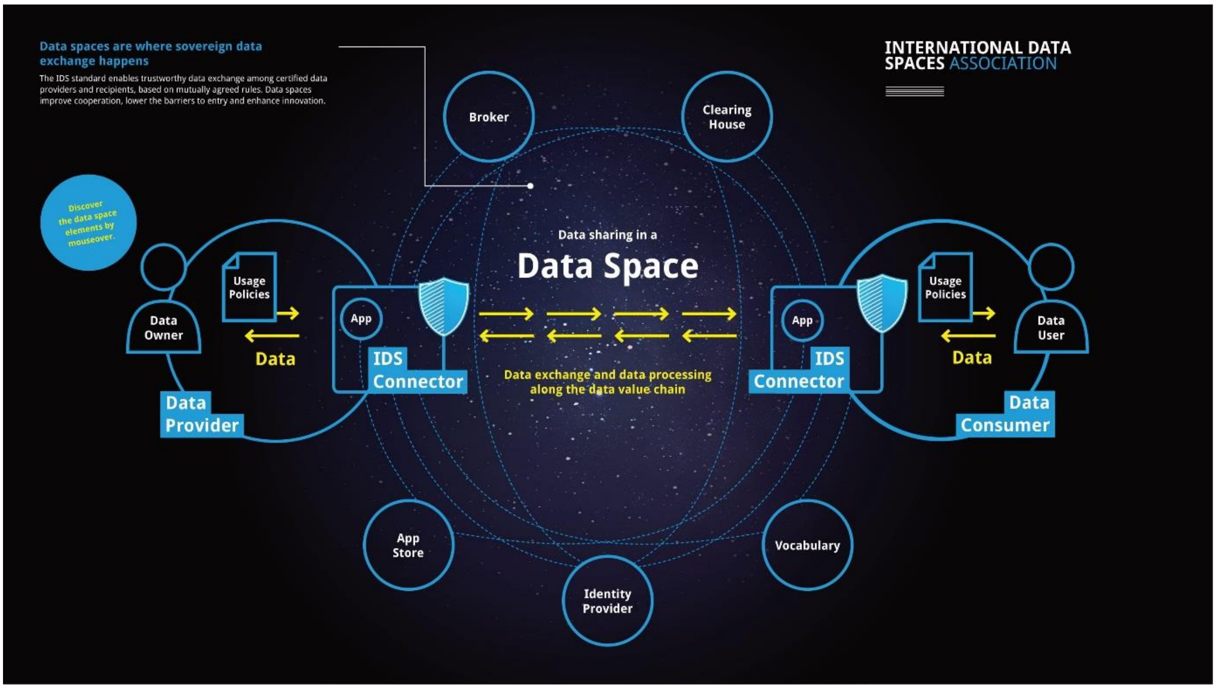

Figure 3: IDS typical Data Space Architecture (Source: International data Space Association)

Brokers and Clearing Houses:

- **Broker**: Helps in discovering and linking data providers with potential data consumers, ensuring the right data reaches the right user.
- **Clearing House**: Handles transaction accountability, ensuring transparency and compliance in data exchanges.

Supporting Infrastructure:

- **App Store**: Provides tools and applications that facilitate data analysis and utilization.
- **Identity Provider**: Ensures secure authentication and authorization within the Data Space.
- **Vocabulary**: Establishes common standards and terminologies to ensure consistent and interoperable data exchange.

This framework highlights the emphasis on secure, standardized, and policy-compliant data sharing, ensuring that data can be exchanged and utilized efficiently across different entities while maintaining trust and data sovereignty.

The UNDERPIN system architecture (see chapter 8.1.3 in this document) follows this approach and recommendation, to be interoperable with other Data Spaces and develop along de-facto standards in the field.

Co-funded by the European Union
This project has received funding from the Digital Europe Programme under grant agreement No 101123179

3 Technical Requirements Specification

This chapter outlines the technical requirements for integrating and deploying various components for the **UNDERPIN** project. As outlined above we follow the IDS architecture recommendations and make use of the Sovity community edition based on EDC software components.

These components include a Connector with its User Interface (UI), DAPS (Dynamic Attribute Provisioning Service) and an Authority Portal. These components are sourced from the following GitHub repositories:

- Connector: https://github.com/sovity/edc-ce
- Connector-UI and Broker-UI: https://github.com/sovity/edc-ui
- DAPS: https://github.com/sovity/sovity-daps
- Authority Portal & Data Catalogue (Broker): https://github.com/sovity/authority-portal

The primary purpose of these components is to facilitate secure and efficient data exchange between various systems within the project and its Data Space. The intended users include developers integrating these components, system administrators managing the deployment, and end-users interacting with the UI components.

The desired system architecture integrates these components in a modular fashion, allowing for flexible deployment and scaling. Each component communicates with the others using standard protocols, ensuring interoperability and security.

The technical requirements for the UNDERPIN manufacturing Data Space focus on creating an environment where data sharing and processing (mainly machine learning based services for predictive maintenance) can take place for shared datasets with sensor data from refineries and wind farms. As described, these datasets contain structured data from sensors, alarms and historical maintenance data. Generally, we process mainly numeric data of time series. The technical requirements for the Data Space are therefore to provide a detailed and semantic description of the actual data (structure and semantics) and to expose services in the Data Space to easily make use of this data regarding training the predictive maintenance and similar machine learning based approaches. For this, we determine the technical components for functional requirements and non-functional requirements. We design the system architecture considering the Data Space components and Machine Learning services, describe the components of the architecture and enable semantic interoperability via a Semantic Layer for data integration.

3.1 Technical Requirements Specification

We identify key requirements and discuss how these are addressed in the Data Space architecture, including component descriptions and infrastructure requirements. Furthermore, we describe requirements for the Semantic Layer for Data Spaces, which will address the semantic interoperability for data and services, but also enables improved Data Space metadata management (by richer metadata) and thereby enables improved search and recommendation in the Broker (catalogue) component.

3.1.1 Technical Key Requirements

We identify the following technical key requirements for the manufacturing Data Space:

Co-funded by the European Union

This project has received funding from the Digital Europe Programme under grant agreement No 101123179

- **Semantic description** of structured use case data: a major factor for (re)usability of data regarding refineries and wind farms, specifically structured sensor data, is to provide a clear semantic description of the data.
- **Consolidation services**: to be able to (re)use different data sets for training predictive maintenance we provide services for data alignment based on semantic technologies.
- **Management of time series data**: the data for predictive maintenance is sensor data and alarm logs, including historical data of incidents with refineries and wind farms. We provide the technological components for using this data efficiently to train the predictive maintenance, e.g. select a specific time window out of a large data set.
- **Semantic Layer:** we provide semantic services based on and extending IDS components, like the Vocabulary Hub (see: https://docs.internationaldataspaces.org/ids-knowledgebase/v/ids-ram-4/layers-of-the-reference-architecture-model/3-layers-of-the-reference-architecture-model/3_5_0_system_layer/3_5_6_vocabulary_hub, accessed 08/2024), to implement semantic interoperability in the Data Space. These services harmonize semantic descriptions and use data consolidation to provide a unified Semantic Layer as a basis for the predictive maintenance services.
    - **A Semantic Vocabulary Hub** will furthermore be implemented on the Data Space metadata level by making use of PoolParty Semantic Suite (of SWC). This allows the UNDERPIN system to semi-automatically semantically annotate the metadata of all datasets in the Data Space by pre-defined controlled vocabularies (like eg EuroVoc, see: https://eur-lex.europa.eu/browse/eurovoc.html?locale=en, accessed 08/2024), and thereby allow (i) better search, (ii) better recommendation, and (iii) cross-Data Space searches via the Broker for the future (if agreement with other Data Spaces on certain vocabularies can be made).
    - A **semantic repository** (database) will be deployed, using Ontotext GraphDB. It will store semantic metadata and provide SPARQL querying and faceted search, facilitating richer discovery services.

We note that all requirements consider IDS standards and will be implemented in alignment with IDS components.

3.1.2 Data Space Infrastructure

Deployment and Integration

Each UNDERPIN component should be containerized using Docker (see: https://www.docker.com/, accessed 08/2024). This approach will facilitate scalable and flexible deployment in cloud environments. Communication between components must be secured using HTTPS, with appropriate authentication mechanisms such as OAuth2.0 (see: https://en.wikipedia.org/wiki/OAuth, accessed 08/2024) and JSON Web Tokens (JWT, https://jwt.io/introduction, accessed 08/2024). Furthermore, each service must implement comprehensive logging and monitoring capabilities to ensure high availability and rapid issue resolution.

Compliance and Standards

All components must comply with relevant data protection regulations, including GDPR (see: https://eur-lex.europa.eu/eli/reg/2016/679/oj, accessed 08/2024), and adhere to industry best practices for secure software development and deployment. The software should be designed with scalability, maintainability, and security in mind.

Co-funded by the European Union This project has received funding from the Digital Europe Programme under grant agreement No 101123179

Sovity Infrastructure

The components of the UNDERPIN project are currently hosted on the Sovity infrastructure, which is built on Microsoft Azure (https://azure.microsoft.com/, accessed 08/2024).

This cloud-based setup provides robust scalability, flexibility, and high availability, essential for the efficient operation of the project's Data Space infrastructure. For identity and access management, Sovity utilizes Keycloak (see: https://www.keycloak.org/, accessed 08/2024), an open-source Identity and Access Management (IAM) solution that offers comprehensive security features. The Sovity infrastructure integrates all components seamlessly, ensuring secure communication and data exchange. The infrastructure's security measures include:

- HTTPS and TLS: All communications are encrypted using HTTPS and TLS to protect data in transit
- OAuth2 and JWT: These authentication protocols are used to secure API communications and ensure that only authorized entities can access sensitive data and services
- Monitoring and Logging: Comprehensive monitoring and logging are implemented using Grafana (see: https://grafana.com/, accessed 08/2024) and Azure Monitoring, providing insights into system performance and security events

This infrastructure setup ensures that the UNDERPIN project components are deployed in a secure, scalable, and efficient manner, leveraging the capabilities of Microsoft Azure and Keycloak to meet the project's operational and security requirements.

Estimated infrastructure utilization:

DAPS: Based on currently available information, the deployment of a DAPS will utilize only a single-digit percentile of one CPU core during its usage phase. A peak load is expected during the initial startup of the DAPS. In addition, the DAPS will require approximately 1 GB of RAM under average load scenarios.

Authority Portal/Broker: The Authority Portal, including the Broker, is estimated to use approximately 5% of a CPU core in idle mode and approximately 650MB of RAM. Please note that actual usage may vary as components are deployed in the UNDERPIN infrastructure.

Co-funded by the European Union

This project has received funding from the Digital Europe Programme under grant agreement No 101123179

3.1.3 UNDERPIN System Architecture

The UNDERPIN system architecture needs to consider the requirements above and provide a technological basis for the further development of use cases and components. We note that the IDS components to form the IDS system layer architecture are seamlessly integrated into the overall system architecture. However, there are services which are not involved in the IDS-based data sharing, and which are therefore not part of the IDS system layer.

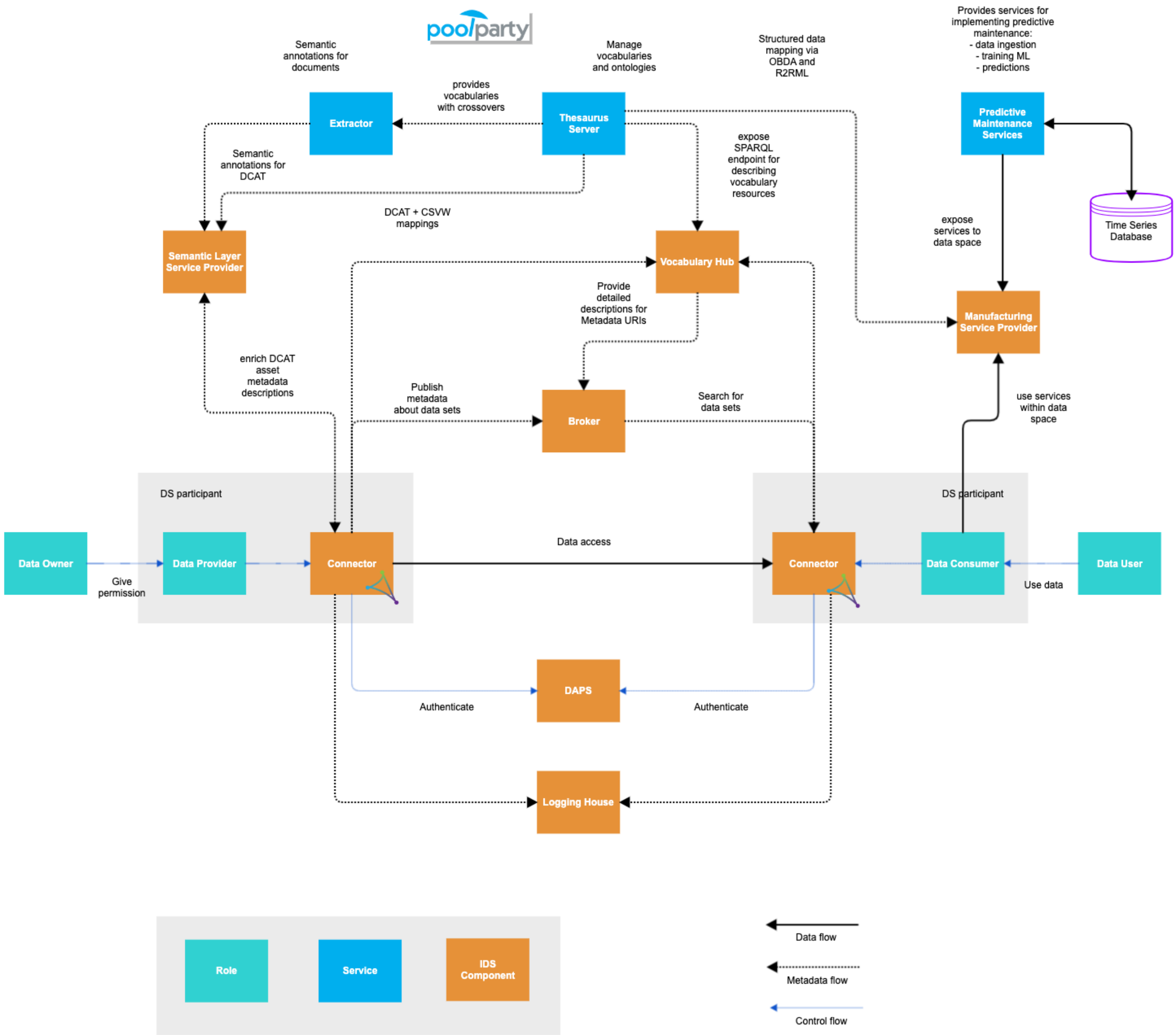

Figure 4: UNDERPIN System Architecture

3.1.4 Description of UNDERPIN Components

We describe the components and services shown in the system architecture grouped by i) IDS components (shown in orange colour in the architecture diagram), ii) UNDERPIN Data Space services (shown in blue colour in the architecture diagram) and iii) the time series database (shown as a database icon in the right top of the architecture diagram).

IDS Components

- **Connector**: The IDS connectors implement the data exchange via the exposed data endpoints. They form the actual Data Space with a decentralized approach for data sharing and enable and enforce data sovereignty by means of contractual agreements between a data provider and a data consumer.

Co-funded by the European Union

This project has received funding from the Digital Europe Programme under grant agreement No 101123179

UNDERPIN will use the EDC-based Sovity connector implementation and extend it as needed by the project requirements. In addition, UNDERPIN will integrate smart contracts into the connectors to enable logging capabilities for documenting events and actions associated with data exchange. This is further explained below at Logging House.

- **Identity Provider**: identity and access management are a mandatory part of the Data Space to implement access control for data sharing.
  UNDERRPIN will use a KeyCloak-based solution which covers the necessary project requirements regarding authentication.

- **Metadata Broker**: the metadata Broker, which is a connector itself, provides endpoints for registration, publication, maintenance, and query of asset metadata, called Self-Descriptions. The Broker is a central service on the IDS system layer, but contains metadata only, while data sharing is done mutually between two Connectors. UNDERPIN will base the metadata Broker on the Sovity Connector and develop graph-based metadata management and search based on Semantic Web standards.

- **Vocabulary Hub**: a central service which provides vocabularies and ontologies as a common basis for interoperability. UNDERPIN will establish PoolParty (see: https://www.poolparty.biz, accessed 08/2024) as a Vocabulary Hub, which covers the necessary functionality with a rich set of vocabulary and ontology management features and exposes the descriptions via SPARQL endpoint.

- **Service Provider**: UNDERPIN introduces several services (described below) for implementing the use cases. To make these services available as part of the trusted IDS Data Space, we introduce service provider Connectors, which expose the service functionality via Web APIs. We identify two services to be exposed.
  - **Predictive Maintenance**: these are services for training and applying predictive maintenance. We need to establish trusted access for Data Space participants, who send shared training data to these services.
  - **Semantic Layer**: the Semantic Layer is a set of unified services to enable semantic interoperability in Data Spaces. It is based on the Vocabulary Hub component to provide data descriptions, consolidation and integration to make it easier for participants to make best use of the data available in a Data Space.

- **Logging House**: the Logging House is a central logging facility to be able to track data sharing actions within the Data Space and to provide billing information.
  UNDERPIN intends to introduce a Smart Contract implementation as a component to cover the Logging House functionality. By integrating Smart Contracts into the connectors, we can automatically log actions, like the establishment of a contractual agreement or the exchange of data, to the central Logging House.

<u>The core IDS components are described in more details as follows:</u>

Connector and its UI

The connectors are responsible for facilitating the actual data exchange and managing connections between different connectors, the data-sources and data-sinks within the UNDERPIN Data Space. The connector UI (connector User Interface) provides a graphical interface for users to interact with and manage the connector easily. In addition, it can also be controlled via APIs and applications can be built around it that access these APIs.

DAPS (Dynamic Attribute Provisioning Service)

The DAPS handles identity management and attribute provisioning within the UNDERPIN Data Space. It ensures that identities like connectors are authenticated, and appropriate attributes are provisioned securely to facilitate trusted data exchange.

Authority Portal and Data Catalogue (Broker)

The Authority Portal is a central management system for authorities and Data Space members, providing oversight of the platform, participant management, and ensuring compliance with Data Space policies and regulations. The Data Catalogue (Broker) is integrated within the Authority Portal, allowing users to utilize both the Authority Portal and Data Catalogue seamlessly in one location. The Data Catalogue is exclusively available to Data Space members, thereby enhancing the benefits of their membership. Access to the Data Catalogue requires registered membership through the Authority Portal.

For the Data Catalogue, we aim to provide extended semantic search capabilities beyond the IDS metadata descriptions. For this we will integrate the k*nowDE* (knowledge-based recommendations) tool from the project partner ARC as a semantic search assistant into the Broker search component.

The integration of both components is already in place, so that the Data Catalogue “knows” the user information (access rights, authentication) of a user accessing the Data Catagolue component.

UNDERPIN specific Data Space Services

- Predictive maintenance services
- Data Space Semantic Layer: is implemented as an IDS Vocabulary Hub using the PoolParty Thesaurus Server for vocabulary and ontology management and publishing and as a set of interoperability services exposed via service provider connectors. A detailed description is provided under 3.1.5.

Time Series Database

- A time series database (TSDB) will be deployed to provide efficient processing and querying of sensor data, thus enabling scalability and facilitating predictive maintenance analysis.
- We have tentatively selected the leading open source TSDB: InfluxDB 2.x.
- Influx has two query languages: Flux for functional-style processing and querying, and InfluxQL for declarative querying in the style of SQL. Flux offers over 400 functions (https://docs.influxdata.com/flux/v0/stdlib/all-functions/, accessed 14 Aug 2024) for all kinds of data manipulation, including moving averages, windowing, derivatives, integration with machine learning, etc.
- Influx comes with an ecosystem of tools (https://docs.influxdata.com/influxdb/v2/tools/ and https://www.influxdata.com/time-series-platform/, accessed 14 Aug 2024): Telegraf for ingestion of various data sources, Chronograph for managing the database including visualization and dashboarding, Kapacitor for real-time data processing of Influx data

This project has received funding from the Digital Europe Programme under grant agreement No 101123179

including machine learning, InfluxQL shell for making queries, Influx CLI (command line interface) for automation, etc.

We plan to integrate InfluxDB in the Data Space as follows:

- Use Data Space users and permissions (signed data contracts) to provision access control in Influx.
- Use semantic descriptions of sensor data (e.g. each column described using the CSV on the Web standard by the W3C) to provision Influx data ingestion processing.
- Implement a service to ingest described columns of a Data Space dataset to Influx and enable querying through a Data Space endpoint.

3.1.5 Description of UNDERPIN Semantic Layer

The Semantic Layer for the manufacturing Data Space is the key enabler for clear and transparent semantic interoperability and services which provide easy access for participants to implement this interoperability based on Semantic Web standards. The Semantic Layer includes:

- **Central provisioning of standard vocabularies and ontologies** via Vocabulary Hub component implemented using the PoolParty Thesaurus Server. Standards-based access is provided via SPARQL query endpoint.
- **Services for applying these vocabularies and ontologies** for data descriptions:
    - Document annotation services to provide semantic annotations based on vocabularies.
    - Structured data descriptions via ontologies.
- **Services to consolidate different data sets** based on the descriptions:
    - Interlinking of vocabularies using semantic crossover relations.
    - Structured data consolidation based on standards like OBDA and R2RML (as needed).
- **Smart services for discovering semantic elements** within data sets:
    - Content analysis to retrieve dataset metadata.
- **All services are backed by scalable data storage** and standards-based access via GraphDB (see: https://www.ontotext.com/products/graphdb/, accessed 08/2024) SPARQL.

3.2 UNDERPIN Data Space identified required accompanying services

The UNDERPIN Data Space will provide secure access to valuable data resources, promoting data sharing and collaboration. Essential accompanying services include a smooth onboarding process with clear eligibility criteria and robust identity verification. Comprehensive user support features a documentation portal, FAQ, and ticket system for technical issues. These services ensure data quality, privacy, and security while they could also be considered as additional revenue streams through fees and value-added services.

3.2.1 Onboarding & Authentication

The UNDERPIN DS aims to provide secure and controlled access to valuable data resources, facilitating data sharing and collaboration among participants. Establishing a robust onboarding and authentication process is crucial to maintaining the integrity and value of the DS. This process addresses key business and administrative aspects to ensure a smooth experience for new participants.

Key components of this process include defining eligibility criteria based on industry, data relevance, and compliance requirements. Prospective participants must demonstrate a clear value proposition, such as contributing valuable data or engaging in collaborative projects. Additionally, data requirements for registration will necessitate details about the data types to be provided or used, formats, quality, and ownership, along with data sharing agreements and privacy policies. To ensure the authenticity of participants, robust identity verification processes will be implemented. This may involve multi-factor authentication or other advanced security measures.

The onboarding process may also include review periods to assess eligibility and data quality, potentially leading to waiting times before full access is granted. While it has not yet been decided if onboarding fees will be applicable to cover administrative costs, this will be carefully considered to avoid deterring potential participants. Pricing models, that will be detailed in D5.1 (due M12), could include standardized fee structures based on data volume, usage, or value-added services.

The onboarding experience will be designed to be guided and user-friendly, offering clear instructions and support throughout the process. Data provided by participants will undergo validation and curation to ensure quality and consistency. Participants will be informed about data protection measures and how their data will be used and shared. Additionally, access to knowledge and training resources will be provided to help participants fully understand the DS and its benefits.

From a business model perspective, onboarding fees, data access charges, and value-added services can generate revenue. Efficient onboarding processes and automation will help minimize administrative costs, while effective marketing and outreach strategies will attract and onboard new participants. Clear data governance policies and procedures will be crucial for managing data quality, privacy, and security.

3.2.2 User Support & Guidelines

While the onboarding process will be a crucial first step in the integration of new participants in the Data Space, it is also important to offer continuous user support, establishing procedures that will enhance the experience of the Data Space users, address any technical issues and difficulties that may arise and offer communication channels for suggestions and potentially even feature requests. Although the exact features of the user support depend on and are heavily interlinked with the business model that the Data Space will operate under, an initial outline of the planned features is presented below:

- **Documentation portal**: A portal containing information on the structure and the individual components of the Data Space, the functions and the features that it offers as well as guidelines and best practices for its use. It will be an evergreen resource that will be periodically updated depending on any relevant developments within UNDERPIN's ecosystem.

- **Frequently Asked Questions**: The FAQ will address common questions from the participants, by offering concise answers and linking to relevant resources (i.e. documentation) where appropriate. It will also indirectly act as feedback in order to identify information that could potentially be included in the documentation portal and is currently missing.

- **Ticket system**: For any technical issues and difficulties that may arise, a ticket system will be available, offering a centralized platform for users to submit support requests and the support team to manage and resolve them.

Additional support features may be included depending on needs that are identified during the operation of the Data Space. Examples of such features would be: a) a set of video tutorials that will act as visual aid for key information from the documentation portal regarding the use of UNDERPIN's functions and features, and b) a community forum where technical questions, suggestions and potentially requests for new features could be discussed among the community. The latter may also be conducive to further fostering an atmosphere of cooperation and collaboration within the UNDERPIN ecosystem.

3.3 Security & Trust mechanisms in the UNDERPIN Data Space

The UNDERPIN Data Space leverages Sovity's infrastructure and security frameworks, hosted on Microsoft Azure, to provide a secure and scalable environment for data exchange. Keycloak handles Identity and Access Management, while the DAPS manages participant identities with secure token-based authentication. The Data Space adheres all regulations, ensuring data protection through encryption and anonymization. The infrastructure includes comprehensive security measures and real-time monitoring to safeguard against threats.

3.3.1 Security & Trust in Data Spaces

Data spaces provide a collaborative framework for sharing data among multiple stakeholders, driving innovation and efficiency. However, the effectiveness and success of Data Spaces are heavily dependent on two foundational elements: security and trust. This section explores the significance of security and trust in Data Spaces, highlighting their value and the advantages they bring to data sharing, with specific reference to the IDS RAM (International Data Spaces Reference Architecture Model).

Security in Data Spaces is mainly about protecting sensitive information from unauthorized access, breaches, misuse and other potential threats. The IDS RAM outlines comprehensive security measures across multiple layers to ensure data is safeguarded throughout its lifecycle.

Here are key reasons why security is essential in Data Spaces:

Protection of Sensitive Data

Confidentiality: The IDS RAM emphasizes the importance of data confidentiality. This is achieved through encryption techniques that protect data both at rest and in transit. Secure communication protocols such as TLS are implemented to prevent unauthorized access during data exchanges.

Compliance: Organizations must adhere to regulatory requirements such as GDPR, HIPAA, PIPEDA and other data protection laws. IDS RAM provides guidelines to ensure compliance through robust security policies and mechanisms, helping organizations avoid legal and financial repercussions.

Ensuring Data Integrity and Availability

Data Integrity: The IDS RAM specifies the use of techniques such as digital signatures and hash functions to ensure that data remains unaltered during transmission. This guarantees that the data received is exactly as it was sent, preserving its accuracy and reliability.

Availability: Security measures outlined in the IDS RAM include redundancy and failover strategies to ensure data availability even in the event of system failures or cyberattacks. This ensures that critical data is always accessible to authorized users when needed.

Access Control and Authorization

Access Management: The IDS RAM employs advanced access control mechanisms such as Attribute-Based Access Control (ABAC) to enforce policies that determine who can access specific data resources and under what conditions. This minimizes the risk of unauthorized access and data breaches.

Usage Control: Beyond simple access, the IDS RAM supports usage control, which governs how data can be used once accessed. This includes setting conditions and obligations that must be met during data usage, ensuring that data is handled according to agreed-upon policies.

Monitoring and Incident Response

Continuous Monitoring: IDS RAM promotes continuous monitoring of data transactions and system activities to detect and respond to security incidents in real-time. This involves maintaining logs of all access and usage events for audit and forensic purposes.

Incident Management: An incident response framework is outlined in the IDS RAM, enabling organizations to quickly detect, respond to, and mitigate the impact of security breaches. This includes predefined procedures and roles for effective incident handling.

Trust is another foundational element for the effective operation of Data Spaces. Without trust, stakeholders would be hesitant to share their data, fearing misuse or breaches. The IDS RAM addresses trust management through several key mechanisms, ensuring that participants can confidently engage in data exchanges.

Identity and Trust Management

Unique Identification: The IDS RAM ensures that every participant in the Data Space is uniquely identified using a reliable identity management system. This prevents impersonation and unauthorized access, establishing a secure foundation for trust.

Authentication: Robust authentication methods such as OAuth2 and OpenID Connect are used to verify the identity of participants before granting access to data and services. This ensures that only legitimate entities can participate in the Data Space.

Certification and Accreditation

Participant Certification: The IDS RAM mandates that participants undergo a certification process to verify their adherence to security and operational standards. Trusted third parties conduct these certifications, ensuring that all participants meet the required trust levels.

Component Certification: Individual components within the Data Space, such as connectors and data Brokers, are also certified to ensure they meet security and interoperability standards. This adds an additional layer of trust to the overall system.

Co-funded by the European Union

This project has received funding from the Digital Europe Programme under grant agreement No 101123179

Trust Framework and Governance

Trust Anchors: The IDS RAM establishes trust anchors, such as certification authorities and governance bodies, that provide a foundation for trust. These anchors issue digital certificates and accreditations, verifying the trustworthiness of participants and components.

Trust Level Agreements: Participants agree on trust levels based on security and compliance requirements. These agreements ensure that all parties adhere to a common set of standards, fostering a trustworthy environment for data sharing.

Transparency and Accountability

Transparent Operations: The IDS RAM promotes transparency by ensuring that participants have visibility into how their data is being used. This includes access to usage logs and audit reports, enabling participants to verify that their data is being handled according to agreed-upon policies.

Accountability Mechanisms: Mechanisms are in place to hold participants accountable for their actions. This includes legal recourse in cases of non-compliance or misuse of data, ensuring that participants can trust that their data will be used responsibly.

Values and Advantages of Security and Trust in Data Spaces

Implementing security and trust mechanisms in Data Spaces brings numerous values and advantages, including:

- **Enhanced Data Utilization**: Secure and trusted environments enable organizations to confidently share and utilize data, driving insights and innovation.
- **Risk Mitigation**: Strong security and trust mechanisms reduce the risk of data breaches, financial losses, and reputational damage.
- **Compliance and Legal Assurance**: Organizations can ensure compliance with regulatory requirements, avoiding legal issues and penalties.
- **Operational Efficiency**: Streamlined data sharing processes and reduced need for redundant checks increase operational efficiency and speed.
- **Market Competitiveness**: Being part of a secure and trusted Data Space enhances an organization's market position, attracting more partnerships and opportunities.

Security and trust are indispensable elements of effective Data Spaces. They protect data from various threats and foster a collaborative environment where participants can confidently share and utilize data. By implementing robust security measures and establishing a trust framework, as outlined in the IDS RAM, Data Spaces enable secure and trusted data exchange, driving innovation, efficiency, and competitive advantage.

3.3.2 Security & Trust Mechanisms in the UNDERPIN Solution

The UNDERPIN Data Space leverages Sovity's robust infrastructure and security frameworks to ensure a secure and trusted environment for data exchange and collaboration. The core components are hosted on Microsoft Azure, providing a reliable and scalable platform. Keycloak is utilized for Identity and Access Management (IAM), adding an additional layer of security and control. Below are the primary security and trust mechanisms employed within the UNDERPIN Data Space:

Co-funded by the European Union

This project has received funding from the Digital Europe Programme under grant agreement No 101123179

A. Identity and Access Management (IAM)

Keycloak serves as the central IAM solution for users and applications, offering comprehensive user authentication and authorization features for accessing the different components and their APIs. The DAPS additionally serves as an issuer of DAT tokens so that the connectors can authenticate each other. The DAPS plays a critical role in managing and verifying the identities and attributes of Data Space participants. The DAPS issues short-lived tokens that include verified attributes and credentials. These tokens are used to authenticate and authorize participants during interactions, ensuring that trust is established dynamically and securely. All messages are digitally signed with a DAT. Receivers can verify these DAT and its signatures using the sender's public key, ensuring that the data has not been tampered with and confirming the sender's identity.

The trust mechanisms are based on widely accepted standards, such as X.509 for digital certificates and OAuth2 for authorization. This ensures that participants from different organizations and systems can interact seamlessly. They collectively ensure a secure and reliable environment within the UNDERPIN Data Space, enabling participants to engage confidently and securely in data exchange activities. By leveraging DAPS and standardized trust frameworks, the Data Space not only maintains high levels of security but also fosters a transparent and trustworthy ecosystem.

B. Data Protection and Privacy

All components within the Data Space adhere to strict data protection regulations, including GDPR:

- Data Encryption: Data in transit is encrypted using HTTPS and TLS, ensuring secure communication channels. Data at rest is also protected using strong encryption standards.
- Data Minimization and Anonymization: Personal data is minimized and anonymized where possible, reducing exposure to potential data breaches.
- Decentralized Data Space: The Data Space is designed to be decentralized, meaning that the actual data remains with the companies that own it, as the data-sources and data-sinks of the Connectors stay at the companies. This is ensuring that the actual data stays within the control of the respective companies using access- and contract-policies through Connectors when transferring the data to other Data Space participants.

C. Infrastructure Security

The infrastructure is designed with multiple layers of security to protect against various threats:

- Network Security: Azure's security tools, such as Cloud Defender and its firewalls, protect the infrastructure from unauthorized access and cyber threats.
- Monitoring and Incident Response: Comprehensive monitoring using Grafana and Azure Monitoring ensures real-time detection of anomalies and incidents. Incident response plans are in place for quick remediation.

Co-funded by the European Union
This project has received funding from the Digital Europe Programme under grant agreement No 101123179

3.4 Status of implementation of the UNDERPIN Data Space

Based on the existing embryonic Data Space at MOH that was already in place before the UNDERPIN project, and the requirements engineering carried out in the requirements specification process managed, the process and the related results are described in detail in this document (elicitation, specification, technical specification). Based on this the provision and development of the UNDERPIN Data Space enablers has been started and is in full progress at the moment of submission of this deliverable (M9, August 2024).

The MOH Embryonic Data Space is based on SAP HANA and provides some basic data and processes for predictive maintenance in the area of refineries at MOH.

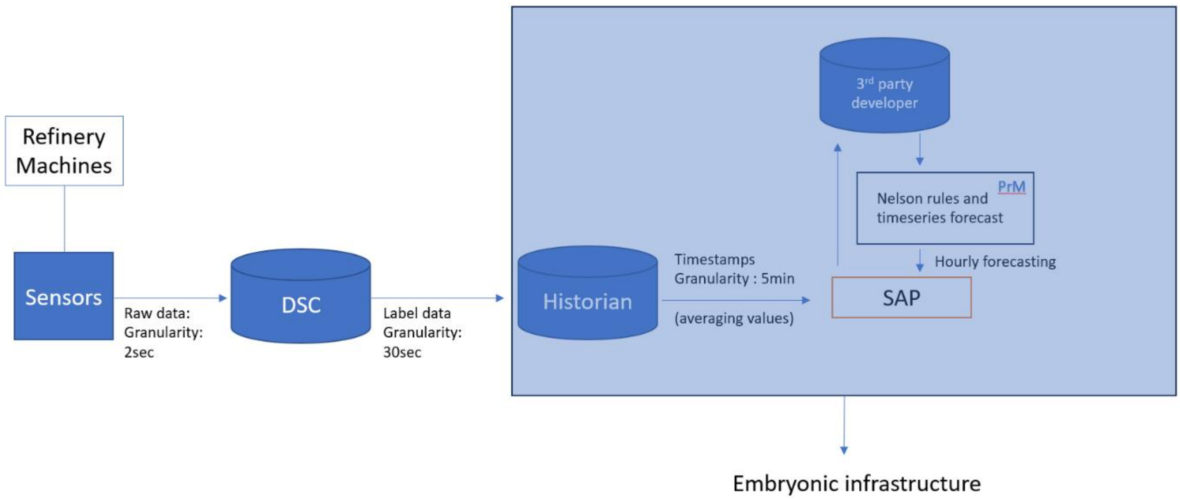

Figure 5: Architecture of the MOH Embryonic Data Space

The users of this system are MOH and some external partners that have been selected and invited specifically. The data exchanged / provided is raw data from all sensors. Annotated data from specific sensors are sent to a specific database (Historian) with different granularity. 5 minutes timestamps are sent from Historian to SAP HANA.

Some selected external partner retrieves data from SAP and performs timeseries forecasting. The outcomes of the forecasting model (predictions) are sent back to SAP and MOH uses them.
The SAP system is a cloud environment that consists of three main parts, the database (HANA), the orchestration module (scripts running) (WebIDE) and the visualization platform (SAC). There are 85 Tags (sensors' readings), and they generate a vast amount of data. Approx. 1GB /year is transferred from DSC to Historian. DSC and Historian cannot be accessed from externals. SAP can be accessed upon request and user validation.
Thereby the MOH embryonic Data Space is a closed ecosystem with specifically selected participants and a limited scope of one central use case. By shifting this embryonic Data Space to an IDS based "standard" Data Space environment of UNDERPIN this Data Space can create additional value for MOH, for its near ecosystem and for additional participants in a secure and trusted way.
In the UNDERPIN Data Space the Connectors (Connectors as a Service) by Sovity are in place. At the moment 3 connectors are up and running for tests as well as for further implementation of the UNDERPIN Data Space Services. A full-blown Data Space system of Sovity (based on Sovity

community edition) is up and running and is used for tests and for developments of UNDERPIN specific Data Space Components and Services for the overall Data Space (e.g. the Semantic Layer) and for the Use Cases (e.g. the described Time Series Database). The Data Space base system (Sovity community edition) will be migrated / set-up again in the UNDERPIN own MS Azure infrastructure in the coming weeks (after delivery of this D2.1 deliverable) and this system will become the production environment or UNDERIN.

On 26-28 June, an UNDERPIN DataThon (development sprint) was organised in Vienna by SWC at AIT premises to specify data flows of the UNDERPIN Data Space in more detail and draw the final architecture and specify all required software components of the UNDERPIN Data Space. Furthermore, the technical partners (all of them participated in this DataThon) have started implementation and integration of software components already at this event and used the DataThon as a kind of an integration kick off that is ongoing since then. More DataThons are planned for a) specific Use Case integration work and b) for integration work of Data Space system and Use Case mechanisms from September 2024 to November 2024 to have a full blown and fully tested first release version of UNDERPIN in place (including all components described in this document D2.1) by end of year 2024 the latest.

The Semantic Layer including partners' products GdB (OT) and PoolParty Semantic Suite (SWC) have been set-up for UNDERPIN and test integrations have already been made e.g. at the UNDERPIN DataThon as well as in the projects continuous development process and work.

The integration of all components is a continuous process (following an agile continuous integration approach), where the Data Space basic system (Sovity community edition), the Connector as a Service ecosystem (CaaS), the UNDERPIN specific Data Space components (Semantic Layer) and the Use Case components (Time Series Database) are integrated and tested and provided for production and use in release cycles (iterations).

4 Review of the Regulatory and Standardisation Landscape on industrial data sharing

This section provides a review of the relevant Regulatory and Standardization European Landscape on industrial data sharing – in close alignment with T.6.5 Orchestration, Standardization, and Certification and T5.1 Legal Framework – that has been carried out. The task also includes assessment and analysis of all applicable aspects, laws, regulatory legislation and available standards from a worldwide perspective related to devices, technologies, processes and methodologies related to industrial data sharing.

4.1 Review on the Regulatory Landscape

The regulatory landscape for Data Spaces is undergoing rapid development, reflecting the crucial need for understanding the particular legal requirements that apply to data use for the different types of involved stakeholders. The increasing complexity of the regulatory framework, as well as the unclear interrelation with existing legislation, such as data protection laws, competition laws, or regulations on intellectual property, adds numerous degrees of uncertainty to the particular legal requirements that apply to participants of the Data space. On the other hand, regulatory compliance with existing regulations is crucial for fostering responsible data sharing, as well as ensuring transparency, sovereignty and trust through the entire lifecycle of data sharing ecosystems. This trust is invaluable, as it fosters customer loyalty and encourages data sharing, which is fundamental for innovation and the development of new services and products.
In an evolving legal environment, navigating through the specific legal regulations that apply to data use is a complex task, which includes a range of activities among diverse data types and multiple stakeholders. These activities range from embedding various design considerations of the Data Space within regulatory parameters, to identifying participants' roles and responsibilities, and formulating internal policies to ensure compliance. Compliance with these regulations is essential for ensuring data privacy and sovereignty, which are critical components of any sustainable Data Space. Figure 6 provides an overview of the primary European cross-cutting legislation and an outlining of regulatory requirements relevant to Data Spaces.
The figure 6 bellow is based on a mapping done by KU Leuven (see:

https://www.law.kuleuven.be/citip/blog/from-data-silos-to-data-ecosystems-the-role-of-data-spaces-support-centre-in-addressing-the-legal-challenges-of-the-future-single-market-for-data/, accessed 08/2024.). In the light of the aforementioned matters, in recent years there has been an introduction of multiple significant initiatives at multiple governance levels, aiming to maximise the social and economic benefits of data by improving the ways in which they are shared, processed and utilised. For instance, the EU's General Data Protection Regulation (GDPR), which sets stringent requirements for (personal) data handling and user consent. Adhering to GDPR helps and obliges businesses avoid substantial fines and legal repercussions, which can be financially crippling and damage their reputation. Moreover, compliance builds trust with customers and partners, who are increasingly concerned about how their data is being used and protected.

.

Co-funded by the European Union

This project has received funding from the Digital Europe Programme under grant agreement No 101123179

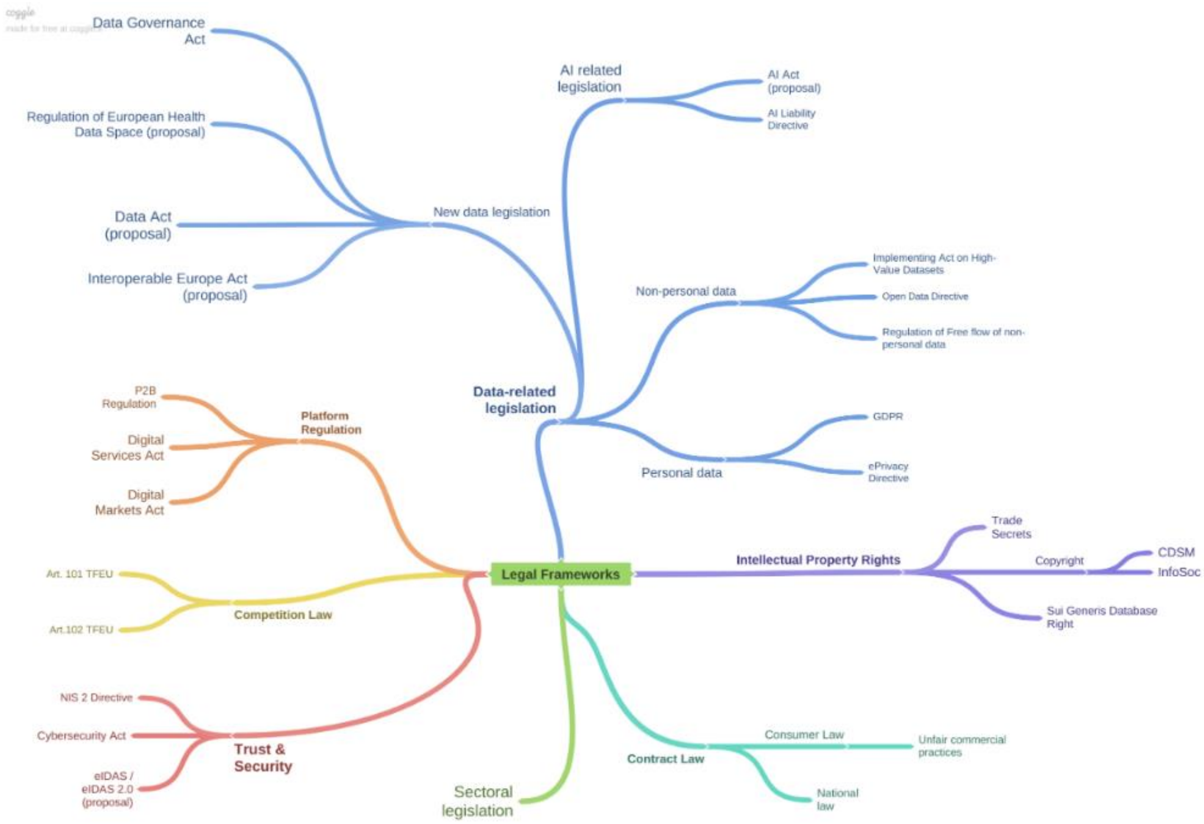

Figure 6: Legal initial mapping of the primary European cross-cutting legislation for Data Spaces

Another noteworthy initiative, the GAIA-X is a collaborative ecosystem where businesses, researchers, and governments can share and access data seamlessly while maintaining control over their data assets. The GAIA-X facilitates a federated and secure data infrastructure that promotes data sovereignty, playing a crucial role in shaping Data within Data Spaces. The goal is to establish a framework for empowering users to make informed decisions across various domains, accommodating specific needs of different industries, and laying the groundwork for automated compliance processes. The Gaia-X Association covers three principal pillars (Figure 7):

- Compliance: This pillar establishes a common digital governance based on European values. Decentralized services enable objective and measurable trust, ensuring that data can be stored, processed, and transferred efficiently across different platforms while maintaining high security and performance standards.
- Federation Services: This pillar provides the necessary interoperable & portable (Cross-) Sector data-sets ontology and logic rules to facilitate interoperability across federated ecosystems. Federation services ensure that different participants in the GAIA-X ecosystem can work together, maintaining a high level of trust and regulatory compliance, and thus promoting the European values of transparency, openness, self-determination, privacy, and interoperability.
- Data Exchange: The Data Exchange pillar aims to establish a framework for secure and sovereign data exchange and anchor data contract negotiation results into the infrastructure. This includes setting standards for data formats, access controls, and usage policies to ensure data integrity and privacy.

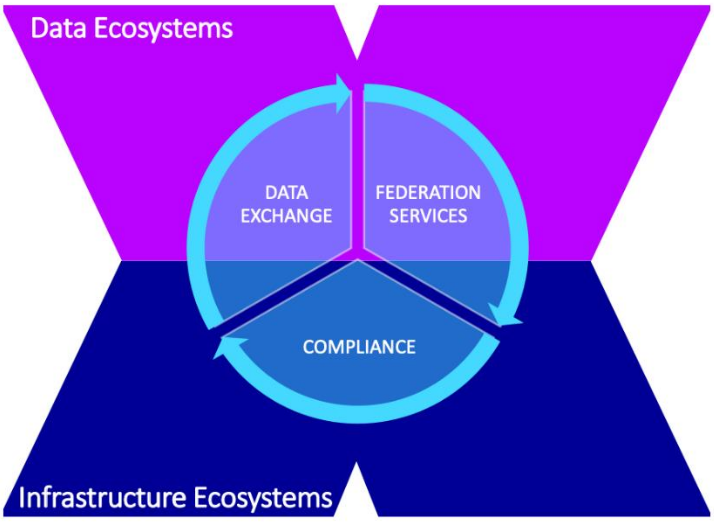

Figure 7: The 3 pillars of the GAIA-X initiative

Recently, the European Commission introduced multiple regulations to standardize digital transformation laying the foundation for the creation of Common European Data Spaces. Launched in 2020, the European Strategy for Data aims to establish a single data market promoting Europe's global competitiveness and data sovereignty. This will be reinforced through a set of policy measures, unique in terms of their primary focus and extent, including the Data Act (DA-E) and the Data Governance Act.

- The DA-E sets a harmonised set of rules on fair access to and use of data and clarifies who can use what data and under which conditions, particularly for products connected to the Internet of Things (IoT). This ensures fairness in the allocation of the value of data amongst companies, citizens and public administrations in the data economy and facilitates secure access to personal data in various key economic sectors.
- The DGA is focused on increasing trust in data sharing processes across sectors and EU countries. It aims to leverage the potential of data for the benefit of European citizens and businesses by fostering mechanisms to increase data availability and overcome technical obstacles to the reuse of data. The DGA will also support the setup and development of Common European Data Spaces in strategic domains, involving both private and public players, in sectors such as such as health, environment, energy, agriculture, mobility, finance, manufacturing, public administration and skills.

4.2 Review on the Standardisation Landscape

This section provides an overview of the relevant standardisation and standards section of UNDERPIN. Standardisation and standards are crucial for the proper development of Data Spaces as of interoperability (especially as of data interoperability as well as of semantic interoperability) to ensure interoperability is provided (a) inside a Data Space = inter Data Space (to ensure metadata and mainly data can be easily exchanged, integrated and used along envisaged use cases to create value) and (b) between different Data Spaces = intra Data Spaces. This section is split into three sub-sections, namely (i) Review of the Standardisation Landscape of UNDERPIN Use Cases, (ii) Review on the Standardisation Landscape: Data Spaces and (iii) the

Current State of Standardisation Activities of the UNDERPIN project as well as an outlook on further plans in this area.

4.2.1 Review on the Standardisation Landscape: Use Cases

The exploration of the standardization landscape in the specific use cases addressed by the project, has produced results concerning both formal standards (i.e. standards from the international and European standardization organizations, ISO and CEN, respectively standards) as well as industrial (informal) standards from industry associations active in the areas of process industry and specifically from Oil and Gas associations.

ISO 14224:2016 titled “Petroleum, petrochemical and natural gas industries — Collection and exchange of reliability and maintenance data for equipment” is the most relevant standard to the specific use cases addressed by UNDERPIN, as it applies to the collection of reliability and maintenance (RM) data in a standard format for equipment. It is a confirmed International Standard developed by Technical Committee ISO/TC 67. Its scope includes facilities and operations within the petroleum, natural gas and petrochemical industries. It describes data collection principles and associated terms and definitions that constitute a "reliability language" that can be useful for communicating operational experience. The applications where collected data may be used include: 1) failure events and failure mechanisms; 2) equipment availability, system availability, plant production availability; 3) maintenance, e.g. corrective and preventive maintenance, maintenance plan, maintenance supportability; 4) equipment failures with consequences for safety and/or environment. The end-goal is the facilitation of a "reliability thesaurus" useful for evaluating equipment in different domains.

ISO 10303:2024 "Automation systems and integration — Product data representation and exchange", is an ISO (formal) standard for the machine interpretable representation and exchange of product manufacturing information. The standard has significant relevance to the Digital Product Passport task. It is known informally as "STEP", which stands for "Standard for the Exchange of Product model data". The international standard is developed and maintained by the ISO/TC 184, "Automation systems and integration", and specifically by the sub-committee SC 4, "Industrial data". The information generated about a product includes different phases in its lifecycle, such as the design, manufacture, usage, maintenance, and disposal. Many information systems may be involved in the different stakeholders, from different organizations, associated with the product. Organizations must be able to convey their product information in a standard computer-interpretable format that is expected to be full and consistent when shared across various information systems in order to support such purposes.

ISO 15926:2018 "Industrial automation systems and integration—Integration of life-cycle data for process plants including oil and gas production facilities", is a standard for data integration, sharing, exchange, and hand-over between computer systems. It is the result of the work of the Technical Committee ISO/TC 184, Sub-committee SC 4. It is informally referred to as "the O&G ontology". The scope of the data model covers the entire lifecycle of a facility (and especially oil refineries, thus the relevance with the case of UNDERPIN) and its components (e.g. pipes, pumps and their parts, etc.) which have numerous properties, each property only applicable to some of the components. Thus, any specific combination of the properties has limited scope and cannot facilitate the description of all components. To this end a dynamic representation (an "ontology") is needed that describes the physical object itself (eg. "pump"), the specific property type ("time inactive"), the base property ("time") and the scale ("seconds").

CFIHOS ("Capital Facilities Information Handover Specification") is an industry standard published by International Association of Oil & Gas Producers (IOGP), thus it is not a formal standard . Nevertheless, it is a very comprehensive standard for the exchange of design data for process facilities from designer/constructor to owner/operator. Its origins may be traced back to 2012, when Shell decided to open up its process and equipment specifications standard and make it an industry wide one. The standard specifically addresses the process and energy industry domains. CFIHOS includes a data model used for structuring data and documents about assets and facilities. It could be used for a deeper semantic description of refinery assets, which can provide richer context for the interpretation of sensor data.

IEC/ISO 81346 "Industrial systems, installations and equipment and industrial products – structuring principles and reference designations" also called Reference Designation Systems (RDS) is used to capture systematically a complex plant or facility (e.g. oil platform, refinery or wind farm) and all its constituent assets, machines and parts. It works along 3 hierarchies: functional (function of equipment), product (the actual products installed to fulfil these functions) and site/location. In particular, IEC 81346-10 Reference Designation System for Power Supply Systems (RDS-PS) can be used to describe power plants and wind farms. Consider this example from the site https://www.81346.com (accessed 08/2024):

Overall structure of the farm:

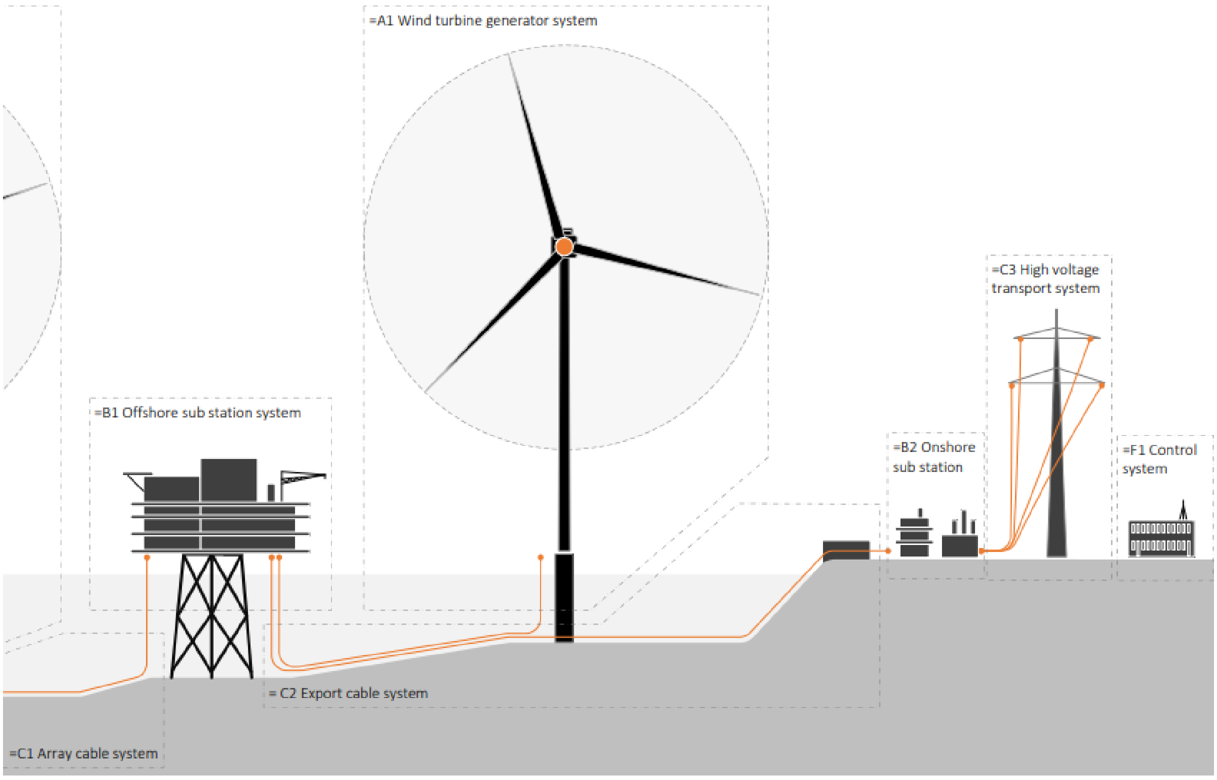

Figure 8: Overall Structure of the Farm

Technical Systems in a Wind Turbine Generator:

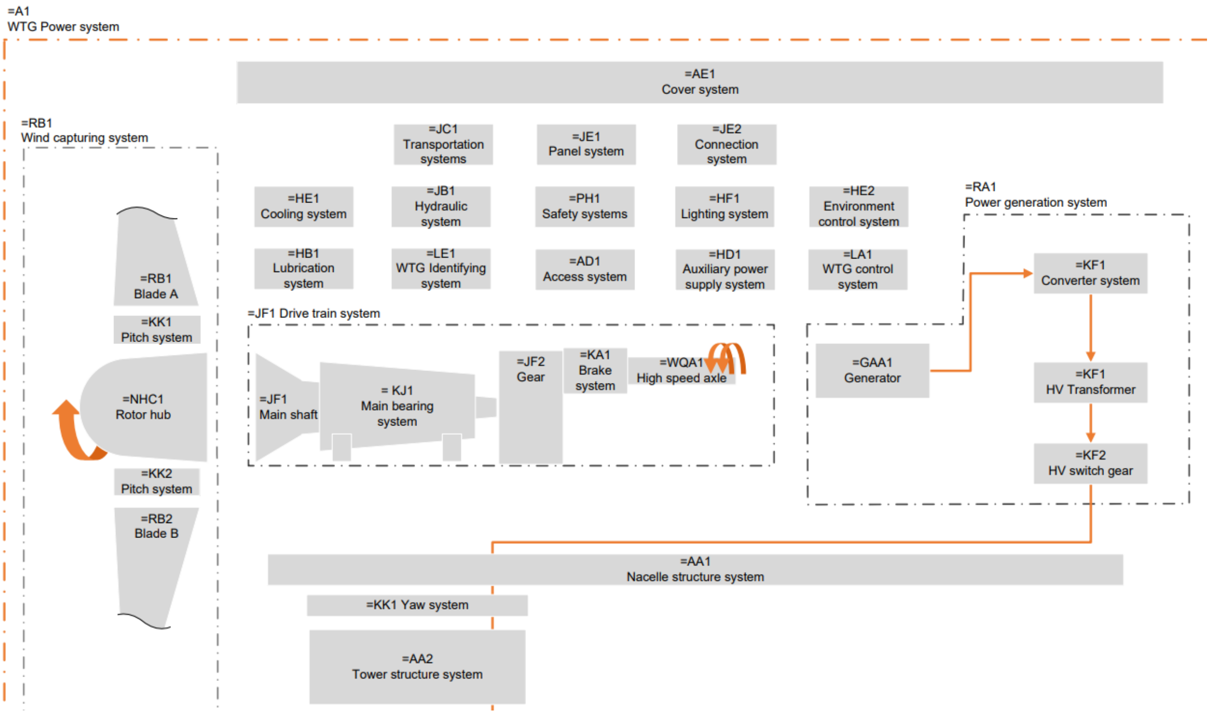

Figure 9: Technical System in a Wind Turbine Generator

All codes are picked from a predefined hierarchy, and the numeric suffixes designate individual components of the listed kind.

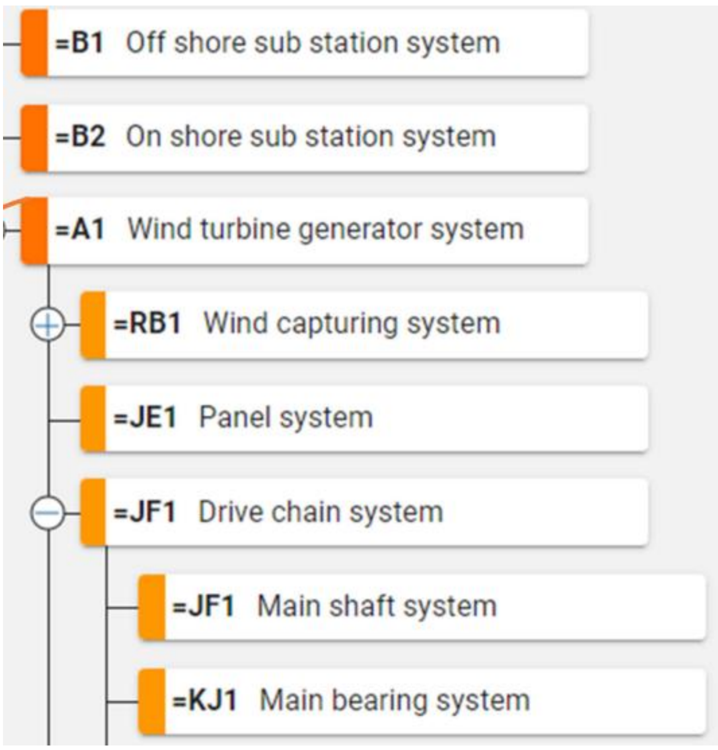

Figure 10: List of Codes

This facilitates interoperability and can be the basis for defining both semantic classes, and semantic URLs of assets and their parts.

Co-funded by the European Union
This project has received funding from the Digital Europe Programme under grant agreement No 101123179

IEC 63278-1 Asset Administration Shell for industrial applications (AAS) is the newest Industrial IoT and Digital Twin standard, and is widely adopted in Germany by companies like Siemens, Bosch, Festo, Sick, ABB, Grundfos, Phoenix Contact, Lenze, Schneider Electric etc. It's also widely promoted by ZVEI (German Association of Electrical and Electronics industry), VDMA (Association of Machine Manufacturers) and the German Ministry of Economy and Climate Change. AAS is the concrete realization of Reference Architecture Model for Industry 4.0 (RAMI 4.0). While initiated by the German Industry 4.0 initiative, it is now adopted in other countries and is standardized as an IEC standard.

AAS consists of "submodels" that describe various product and asset aspects and use cases (e.g. Digital Nameplate, Technical Data, Digital Product Passport, Interfaces/APIs, etc). For example, one can describe the static structure of a robot, then its kinematics and dynamics, PLC design, etc. AAS incorporates ideas from and "subsumes" prior Digital Twin specifications: COLLADA, OPC UA, AutomationML. It can also capture payload from other standards, e.g. DEXPI Piping and Instrumentation Diagrams (P&ID) used in the process industries. Thus, AAS enables a complete description of a complex plant or engineering product.

4.2.2 Review on the Standardisation Landscape: Data Spaces

Standardization brings several critical benefits to Data Spaces, enhancing their functionality and reliability. Standardized protocols and interfaces ensure seamless integration and data exchange across different systems and sectors, which is crucial for enabling secure and sovereign data sharing and collaboration. Moreover, standardization provides a baseline for implementing robust security measures and trust mechanisms, ensuring data protection from unauthorized access and misuse, thus fostering confidence among participants. By adhering to standardized guidelines, organizations can comply with regulatory requirements such as GDPR, reducing the risk of legal issues and enhancing the trustworthiness of Data Spaces. Standardized processes and technologies optimize the implementation and management of Data Spaces, reducing complexity and operational costs, which allows organizations to focus on leveraging data for innovation and value creation. Additionally, international standards facilitate cross-border collaboration, enabling global data sharing and cooperation.

This landscape of standardization activities for Data Spaces involves contributions from multiple organizations and initiatives. Key entities in this space include the International Data Spaces Association, the European Commission's Data Spaces Support Centre (DSSC), and international standardization bodies such as ISO and IEC. These organizations work collaboratively to establish comprehensive frameworks and protocols that support the creation, implementation, and maintenance of Data Spaces. The following section reviews the ongoing standardization activities and the significance of these efforts in fostering a robust and trusted data economy.

4.3 Key Standardization Efforts and Assets in Data Spaces

International Data Spaces Association (IDSA)

IDSA continues to refine and expand its standards and protocols, ensuring they remain relevant and effective in a rapidly evolving digital environment. This includes updating the IDS-RAM and the Dataspace Protocol based on feedback from the community and technological advancements.

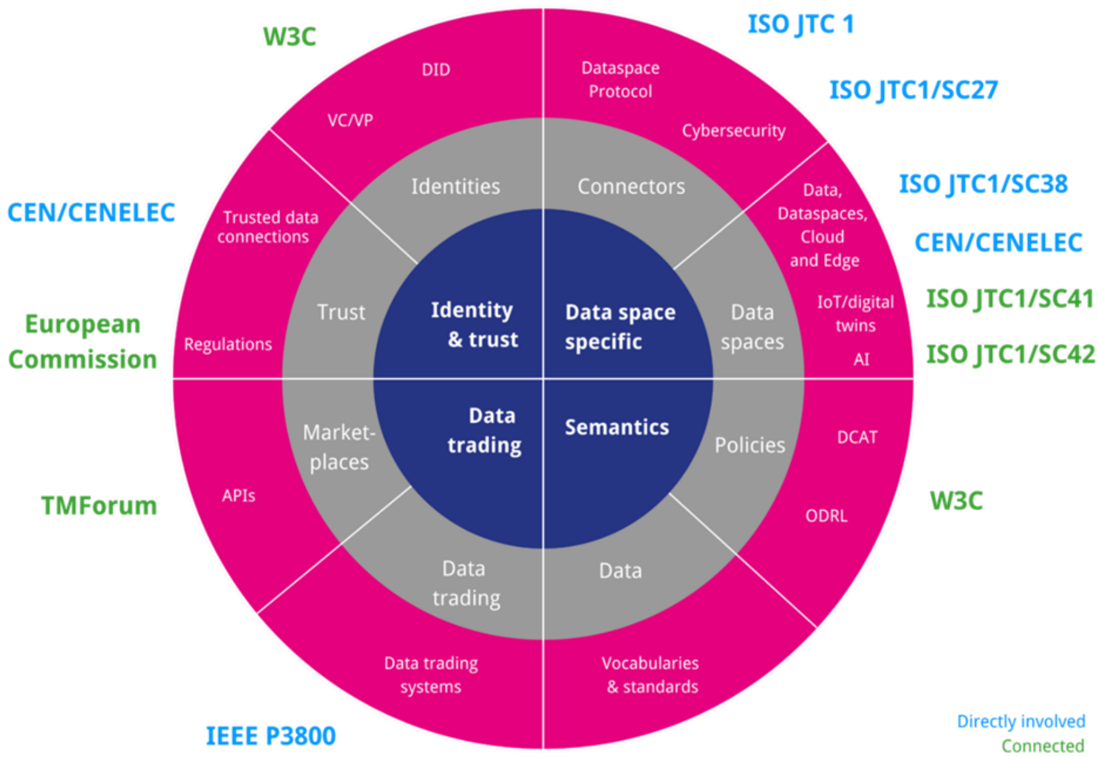

Figure 11: Overview of IDSA's Standardization Efforts

- **Dataspace Protocol**: The Dataspace Protocol is a critical framework developed by the IDSA to facilitate secure and standardized data sharing. It orchestrates steps such as requesting data catalogs, negotiating contracts, and managing data transfers. This protocol ensures that data exchanges are conducted securely and interoperable across different systems.
- **Reference Architecture Model (IDS-RAM)**: The IDS-RAM provides detailed guidelines for building and operating secure and interoperable Data Spaces. It covers aspects like data sovereignty, trust management, and secure data exchange, serving as a foundational framework for the IDS ecosystem.
- **IDSA Rulebook**: This document outlines operational and governance principles for Data Spaces, supporting the implementation of standardized and trusted data-sharing practices.

European Commission Initiatives

- **Multi-Stakeholder Platform on ICT Standardization**: This platform aligns EU policy priorities with standardization efforts, integrating contributions from the IDSA and other stakeholders to create a cohesive standardization landscape that supports the European digital strategy.
- **Data Spaces Support Centre (DSSC)**: The DSSC provides a comprehensive collection of standards and technologies to support Data Space implementation (https://dssc.eu/space/SE1/185794561/Collection+of+Standards+and+Technologies+landscape+%7C+Version+1.0+%7C+October+2023, accessed 08/2024). It

aims to foster interoperability and data sovereignty, which are essential for the success of the European digital strategy.

- **European Commission DIGIT Unit B2 Department "Interoperability",** formerly known as ISA and ISA2 programmes (and as SEMIC team) works on standards and recommendations in regard to interoperability and semantic interoperability with a strong focus on public administration, including the area of Data Spaces.
- **High-Level Forum on European Standardization,** is an expert group of the European Commission. It was set up in January 2023 as requested by the EU Strategy on Standardisation. The purpose of the forum is to identify standardization priorities in support of EU policies and legislation, and to discuss horizontal issues such as international leadership and education and skills, in a multi-stakeholder setting.

ISO/IEC Standardization Efforts

- **ISO/IEC JTC 1**: The Joint Technical Committee of ISO and IEC is actively developing international standards for Data Spaces, focusing on ensuring global interoperability and security. These standards provide a solid foundation for Data Space implementations worldwide.
- **ISO 20151**: This upcoming standard outlines the concepts and characteristics of Data Spaces, providing guidelines for their development and operation. It aims to facilitate international collaboration and data exchange by establishing universally accepted standards for Data Spaces.

Other Standardization Efforts in Place

- **Big Data Value Association / BDVA:** BDVA's position paper on "Data Sharing Spaces and Interoperability", see: https://bdva.eu/news/bdvas-position-paper-on-data-sharing-spaces-and-interoperability/, accessed 08/2024.
- **World Wide Web Consortium / W3C:** with several recommendations in place that are used in Data Spaces and a just founded Community Group: https://www.w3.org/community/Data Spaces/, accessed 08/2024
- **Deutsches Institut für Normung / DIN**: Normung von Datenräumen mitgestalten, see: https://www.din.de/de/din-und-seine-partner/termine/normung-von-datenraeumen-mitgestalten-1099316, accessed on 08/2024.

To further the standardization of Data Space components, tools, and frameworks, and to align specification work with open-source software development, a working group has been established under the Eclipse Foundation, that is called Eclipse Data Spaces Working Group (EDWG, https://www.eclipse.org/org/workinggroups/Data Space-charter.php, accessed 08/2024) This group brings together leading initiatives and organizations to steer the standardization and specification efforts in the area of Data Spaces.

4.4 Current State of Standardisation Activities and Plans: UNDERPIN

Currently UNDERPIN partners are in the early phases of understanding the business environment, review existing standards in the areas of harmonization of data collection and exchange, and understanding the existing gaps in standardization that (may) require interventions by the UNDERPIN stakeholders. In terms of understanding the market dynamics that formulate the business environment we have focused on identifying the stakeholders involved in the industry, the stakeholders' respective motivation in sharing information and the previous and ongoing harmonization activities.

In the recent months and especially following the definition of the use cases, our focus has been to explore the standards that focus on the harmonization of the information collected to be shared with other stakeholders and facilitate the emergence of new algorithms, services and applications, results which are among the major outcomes expected of this project. We have also understood that among the major obstacles that hinder the potential for reuse is:

- the lack of an unambiguous and consistent representation of the monitored assets, not only across the same industry but also among the stakeholders. This includes the equipment classification (currently industry is using diverse naming conventions), its properties and its position in the process and documents,
- the lack of facility information and more specific a plant structure, along with the distribution of functions within the plant or facility and the hierarchy of the components together with their interconnections,
- the non-uniform representation of data measurements.

Thus, the approach we will follow in the next months is to identify the areas where harmonization may be required to overcome these obstacles, e.g. by proposing data formats and semantic (metadata) structures (e.g. ontologies), to establish data and meta-data interoperability.

The UNDERPIN project (and in particular Ontotext) has been active or plans to participate in the following standardisation semantic interoperability initiatives:

- AAS and ECLASS semantic representation: while both standards have a semantic rendition, there is much to be desired in terms of ontological faithfulness, reusing semantic web standards, using Linked Data principles. See the following "holder" issue at the AAS github: https://github.com/admin-shell-io/aas-specs/issues/384 (accessed 14 Aug 2024). The International Digital Twin Association (IDTA) has a Semantics WG that considers solutions to these issues. Resolving them will enable easier capturing of AAS data in a semantic database, powerful querying using SPARQL, and the ability to share submodels and definitions of product characteristics, rather than having to copy them.
- A W3C Dataspaces community group started in May 2024 (https://www.w3.org/groups/cg/Data Spaces, accessed 14 Aug 2024), and is still picking up speed.

If the existing standards (presented before, indicatively ISO 14224:2016, ISO 10303:2024 and ISO 15926:2018), are found to not suffice for the needs of this approach, a new standard may be required. Given the innovative character and the very short timeframe of the project, it may be possible that the solutions have not reached a sufficient level of stability in time for the process required. Thus, a formal standard may be a less suitable solution considering the process in place which requires unanimous support by EU MS which can only be achieved through a lengthy time- and resource-intensive process. To this end the methodology we intend to follow, in case a new standard is required, is the one described in the CEN-CENELEC Guide 29:2024, which aims to the elaboration of a CEN/CENELEC Workshop Agreement. The informal pre-standard nature of the

CWA, does not limit its value, given that when the CWA is finalized, it may then be proposed for conversion into a European Standard to a Technical Committee.

Activities in Semantic Interoperability in Data Spaces by UNDERPIN

The UNDERPIN project organises and supports and pro-actively participates in the Workshop: Semantic Interoperability in Data Spaces III, taking place on 01 October 2024 in Budapest back2back to the European Big Data Value Forum 2024 (EBDVF2024, see: https://european-big-data-value-forum.eu/, accessed 08/2024).

The overall objective of this workshop series is to push standards and recommendations in the field of Semantic Interoperability in Data Spaces, raise awareness for the topic, showcase and discuss the state of the art in this field, and bring people together to work on standards and implementations in this field.

As follows the official outline of the workshop:

Workshop: Interoperability in Data Spaces III
Use Cases, Tooling, Governance
Workshop back2back with EBDVF 2024
Date: 01 October 2024
Time: 09.00-16.00
Venue: Danubius Hotel Arena - Hungary, Budapest Ifjúság útja 1-3. 1148
All Information & Registration: https://semantic.internationalData Spaces.org/workshop-2024/

Workshop Description
The continuously increasing number of Data Spaces and data markets in Europe, as well as the related funding lines and regulations by the European Commission (for instance the Data Governance Act and several programme lines in Horizon Europe and Digital Europe) regarding the topics of (secure) data sharing of industrial and personal related data as well as research data clearly shows the increasing need to discuss, specify and realise (syntactic and semantic) data interoperability between such Data Spaces and beyond.
In June 2022 the 1st workshop: Data Spaces & Semantic Interoperability was accomplished by the TRUSTS project, University of Economics Vienna, IDSA, W3C, SWC, FhG FIT and DIO identifying the bottlenecks and gaps in Data Space interoperability.
In September 2023 the 2nd workshop: Use Cases, Tooling, Governance took place exploring and discussing the current tool & best practice landscape regarding Syntactic and Semantic Interoperability in Data Spaces organised by IDSA, ERCIM, SWC, Insight, Glaciation, and DataBri-X at the SEMANTiCS 23 conference.
All information from the workshop series is available via: https://semantic.internationalData Spaces.org/
In October 2024 the 3rd workshop will take place back2back with the European Big Data Value Forum 2024 (EBDVF2024) on 01 October 2024 in Budapest, Hungary. The focus of the 3rd workshop is two-fold: (i) to discuss different approaches for semantic interoperability in and between Data Spaces (e.g. ontologies and controlled vocabularies, SHACL shapes, JSON schemas, etc...), and (ii) to raise awareness for and discuss cross-domain (semantic) interoperability concepts (e.g. methodologies, vocabularies) AND to bridge the gap between such concepts, overall and particular in regards to standardisation.
This one-day workshop will be a mix of presentations, interactive parts and discussions and brings together researchers, decision makers and practitioners in the field of the development

This project has received funding from the Digital Europe Programme under grant agreement No 101123179

and operation of European (common) Data Spaces, data markets and other web-based data management systems that allow data sharing, trading and data collaboration, to discuss requirements, standards, tools, licences and more regarding (semantic data) interoperability in Data Spaces.

The output document will summarise all the inputs and discussions about mainly (a) different approaches of semantic interoperability inside and between Data Spaces, (b) current concrete applied methods of semantic interoperability in specific domains and thereby highlighting differences and overlaps. The outcome will be published freely available under an open licence / under open access.

Workshop Organisers (in alphabetic order)

- **BDVA** - Big Data Value Association, https://bdva.eu/
- **ERCIM** - the European Research Consortium for Informatics and Mathematics, https://www.ercim.eu/
- **IDSA** - International Data Space Association, https://internationalData Spaces.org/
- **SWC** - Semantic Web Company, https://www.semantic-web.com

Supporting Projects (in alphabetic order)

- **champi4.0ns project**: https://www.champi40ns.eu/
- **DataBri-X project** - Data Process & Technological Bricks for expanding digital value creation in European Data Spaces, https://databri-x.eu/
- **Glaciation project**, https://glaciation-project.eu/
- **UNDERPIN project**, https://underpinproject.eu/
- **VELES project**, https://veleshub.eu/

The results of this workshop will be presented and discussed at EBDVF on 02 October with (i) a short talk about the results in the auditorium and (ii) a panel discussion about the topic in a parallel session.

5 Conclusion

Deliverable D2.1 provides a comprehensive overview of the UNDERPIN Data Space's current status, including requirements elicitation, technical specifications, component integration, and regulatory landscape. The use case requirements and specification details are available in the previously submitted Deliverable D4.1.

The technical requirements identified and specified show a clear path along the European wide (and beyond) developments of Data Spaces, namely following IDS architecture principles and EDC implementations. The UNDERPIN Data Space is built upon the EDC / Sovity community edition and makes use of Sovity Connectors-as-a-Service, that ensure always up-to-date Data Space Connectors available for UNDERPIN in a dynamically developing environment.

This basic Data Space ecosystem – that is the basis to develop from the existing embryonic Data Space into a full fletched European Data Space for manufacturing – will be expanded by new features and services, that are based on the partners software products, namely OTs GdB, SWCs PoolParty Semantic Suite, and ARCs knowDE, and finally AITs Smart Contracting system.

The basis system is up and running, the additional components are in set-up process and integration work has started, and the launch date of the UNDERPIN Data Space is being planned at the moment of the submission of this deliverable for late 2024.

Moreover, the document provides a comprehensive overview of the regulatory framework and standardization environment within which UNDERPIN operates, both in terms of use cases and the Data Space. The report explains the work already carried out regarding standardization and outlines future plans in this area.

Deliverable D2.1 provides thereby a good status about the requirements and the implementation of UNDERPIN, and it will be complemented with D2.2 by the end of the project, that will offer the full pictures of requirements and enablers provision of UNDERPIN.

Co-funded by the European Union

This project has received funding from the Digital Europe Programme under grant agreement No 101123179

6 Bibliography / References

All references have been integrated directly into the text of this deliverable D2.1 to enable the reader to keep the flow of reading and thereby gaining information as uninterrupted as possible.

This project has received funding from the Digital Europe Programme under grant agreement No 101123179

7 Annexes

Annex A - Use case templates

Table 2: UC1.1 overview

Title	Monitoring and predictive maintenance in the refinery
Description	A predictive maintenance algorithm will be developed that will allow for predicting equipment failure as well as detecting abnormal behaviour trends
Use case owner	MOH (maintenance)
Involved partners	MOH, AIT, SPH, INNOV
Assets	5 compressor machine groups from the main process of the refinery
Expected outcomes	Asset owner will be able to appropriately schedule maintenance works for impeding failure, as well as apply corrective actions without interrupting operations based on detected abnormal behaviour
Datasets involved	Sensor data (temperature, pressure, vibration and axial displacement)
Data structures	Format: .xls/.csv Granularity: 5 minutes
Existing infrastructure	- Operations monitored through SAP and SCADA systems - Preexisting predictive maintenance model
Challenges	Insufficient failure data may lead to low accuracy of predictive algorithm

Table 3: UC2.1 overview

Title	Predictive maintenance in wind farms
Description	A predictive maintenance algorithm will be developed that will allow for predicting equipment failure as well as detecting abnormal behaviour trends
Use case owner	MORE (operations)
Involved partners	MORE, AIT, INNOV
Assets	Wind turbine generators from MORE's wind farm portfolio
Expected outcomes	Asset owner will be able to optimize operations based on detected abnormalities, as well as benchmark the maintenance works carried out by contractor
Datasets involved	Sensor data from multiple components within the wind turbine Fault alarms and warnings
Data structures	Format: .xls/.csv Granularity: Every 10 mins
Existing infrastructure	Operations monitored through proprietary SCADA systems offered by wind turbine manufacturers
Challenges	Loss of communication with wind turbines means alarms and errors may not always be detected

Co-funded by the European Union
This project has received funding from the Digital Europe Programme under grant agreement No 101123179

Table 4: UC2.2 overview

Title	Wind turbine blade repair prediction
Description	Statistical analysis of wind turbine blade damages from lightning strikes and prediction of necessary blade repairs based on relevant historical data
Use case owner	MORE (operations)
Involved partners	MORE, AIT
Assets	Wind farms from MORE's portfolio with focus on two wind farms equipped with lightning strike monitoring equipment
Expected outcomes	Asset owner will: - gain deeper understanding on the impact of lightning strikes on wind turbine blades - be able to prevent impending blade failures by carrying out blade repairs in a timely manner
Datasets involved	Sensor data from multiple components within the wind turbine Fault alarms and warnings Historical data for peak current, time of lightning strike, strike intensity Blade repair historical data
Data structures	Format: .xls/.csv /.txt Granularity: Every 10mins
Existing infrastructure	Data collected through proprietary SCADA systems offered by wind turbine manufacturers Additional specialized data collected through Lightning Key Data (LKD) systems currently installed in two wind farms
Challenges	Low availability of lightning data since only two wind farms have LKD systems installed

This project has received funding from the Digital Europe Programme under grant agreement No 101123179

Annex B – Request for Proposal (RfP) Document for UNDERPIN Technical Partner

Request for Proposals (RfP): Technical Support for IDS/EDC related specification and implementation work for UNDERPIN

Vienna, 18.01.2024

Introduction

UNDERPIN is an implementation project funded by the Digital Europe programme, starting 01 December 2023 with a total duration of 24 months.

UNDERPIN addresses the vision of the European Digital Single Market with the idea of a European-wide Data Space for Manufacturing for dynamic asset management and predictive/prescriptive maintenance. European manufacturers in the refinery and renewable energy domains, their value chain ecosystem of SMEs and governmental, research and civil society stakeholders will be equipped with sustainable and secure yet user friendly Data Space aiming at improving processes, cost structure and material management beyond the state-of-the-art data analysis. The solution is compliant to EU standards, Data acts, IDSA and GAIA-X guidelines encompassing enterprise-ready infrastructure and tools to accommodate secure and trusted data, improving company operations. UNDERPIN Data Space will provide a cross-organizational and cross-use-case data sharing and exchanging, a solution that ensures data sovereignty, with strong focus on the interplay of SMEs and large industry players to improve products and services. Real world demonstrators in the demanding and diverse oil refinery and wind farm domains will be implemented. UNDERPIN will ensure higher performance, better insight on asset critically, reduced overall downtimes and maintenance costs, extended machine usage periods, improved future machine designs, new service models, improve production line operations, company-internal processes, increased usability and enhance business opportunities for industrial data value added services, supporting the transition towards circular economy. UNDERPIN will make use of existing embryonic Data Space of the lead partner Motor Oil. This will be transferred to existing mature Data Space technology to ensure security, interoperability, scalability, high performance, increased usability, standards compliance. UNDERPIN develops a trust framework, a sustainable business model for the Data Space operator Motor Oil and implements a Europe-wide scale-out plan to establish a European Data Space for Manufacturing.

The UNDERPIN Consortium (as described below) is requesting a proposal for specification and implementation support as outlined in the Scope (of Work) section of this document.

This document provides general information regarding required services, supplier selection criteria, and response guidelines.

- UNDERPIN Consortium
- MOTOR OIL (MOH) - Co-ordinator
- ATHENA (ARC)
- SEMANTIC WEB COMPANY (SWC)
- WATERMELON (WM)

- AUSTRIAN INSTITUTE OF TECHNOLOGY (AIT)
- TIKO PRO (TP)
- INNOV-ACTS (INNOV)
- SPACE HELLAS (SPH)
- MOTOR OIL RENEWABLE ENERGY (MORE)
- Optional future partners: Nova ICT (NICT), Ontotext (OT)

UNDERPIN Work Package Structure

- WP1 Project management and coordination
- WP2 Requirements and Enablers implementation
- WP3 Data Space Operational Infrastructure
- WP4 Use cases for Data Space demonstration and validation
- WP5 Business plan and sustainability
- WP6 Communication, Dissemination and Scale Out

Scope of Work

The contract (subcontract to the UNDERPIN project concludes with the project partner Semantic Web Company GmbH, headquartered in Austria) for: Technical Support for IDS/EDC related specification and implementation work for UNDERPIN includes work in the following areas, whereby a percentage / % of work related to the whole contract volume is provided. Please indicate in your proposal the pricing according to this percentage per task.

If you think that a different percentage of work distribution would bring value to the project, please specify and explain in your RfP response accordingly.

Table 5: RfP Work Structure overview of UNDERPIN

Work Package No	**Subcontract Task No**	**Subcontractor Name** (subcontracted action tasks)	**Description** (including task number and BEN/AE to which it is linked)	**Estimated Costs** in % of total contract volume
WP2	S2.1	Subcontract via RfP	**T2.2: Requirements, analysis and specifications of the end-to-end UNDERPIN Data Space** To support the specification of the UNDERPIN Data Space, more concrete to support the project team to specify the use of IDSA / GAIA-X standards and guidelines in a sustainable way.	25%

This project has received funding from the Digital Europe Programme under grant agreement No 101123179

WP2	S3.1	Subcontract via RfP	**T3.1: Adaptation of the UNDERPIN infrastructure towards deploying a manufacturing Data Space** To support the set-up of the UNDERPIN Data Space infrastructure by providing best practises from previous work done, and also to ensure that IDSA/GAIA-X standards (connector, DAPS, Broker etc) are used and deployed in the correct way, as the subcontractor has been involved in the development of such standards / is a certified technical IDSA partner or has done extensive work in the field and thereby is expert in the field. Development / adaptations of Connectors and work on Data Usage Policies.	50%
WP3	S3.2*	Subcontract via RfP	**T 3.5: Dry-run testing and operational readiness:** to support the testing of the UNDERPIN Data Space, by supporting the set-up and maintenance of testbeds that are compliant with IDSA/GAIA-X standards and guidelines, with an existing testbed already in place for other projects.*	25%

* ...the Subcontract Task S3.2 can also be shifted into the Subcontract Task 3.1 for implementation, integration and employment work, if this can be argued in a proper and reasonable way and makes sense for the project work and results.

Reasonable travel costs will be covered by the contractor along contractor's travel policies. This means the contractor will book and pay the travel directly or the travel costs will be reimbursed according to agreed costs. The contract will be a fixed-price contract.

In your response to this RfP please describe your expertise in the requested / required field of knowledge and technological know-how and outline your experience along reference projects from the last five years. Please name the concrete persons with short CVs (golden paragraphs with education, experience and add. information you would like to provide) that will work on the contract. Named persons should not be replaced by the Vendor in the course of the contract duration. A replacement can only take place on mutual agreement with SWC and the costs of onboarding to be covered by the Vendor.

A VERY IMPORTANT prerequisite of contracting with the UNDERPIN project (that is a Digital Europe project) is a positive ownership control assessment by the European Commission. To shorten the process on this we recommend to recipients of this RfP to immediately get in touch with the UNDERPIN project officer [first name, last name] via email, refer to A) this subcontracting process of the UNDERPIN project (Project 101123179 — UNDERPIN) and the contact point for this RfP ([first name, last name] of Semantic Web Company, SWC) and ask if such an assessment is already in place or still required, and if it is required to start the process asap. No contract can be granted for this RfP w/o such a positive assessment. This assessment analyses if the ownership of your legal entity is >50% European - as an important remark: such an ownership assessment also opens additional doors for your business via the EC and beyond!!

In addition, an NDA (Non-Disclosure Agreement) needs to be executed with the UNDERPIN co-ordinator (Motor Oil Hellas, MOH).

Duration of Agreement

The agreement starts with the execution of the contract for work and services (as a subcontract to the UNDERPIN project concluded with Semantic Web Company GmbH) and will be specified until the end of the project duration that is fixed with 30.11. 2025. Nevertheless, services need to be provided in 2024 and are planned to be finalised until mid 2025 the latest.

Response Time Frame

Please see the below outline of key dates and deadlines related to this sourcing event. Dates are subject to change at SWC's / UNDERPINs sole discretion.

- **Event issued to respondent: 18.01.2024**
- **Ask Questions! Deadline**: **23.01.2024, 10pm cet**
- **Responses to Questions**: **25.01.2024**
- **Final Deadline for respondent to submit response: 31.01.2024, 10pm cet**

We plan to evaluate responses to this RfP fast and efficiently within a few working days (including a Q&A session if necessary) and thereby conclude the **subcontract by mid of February 2024 at the** latest.

Work on the task S2.1 needs to be provided from contract execution onwards (from 02/2024). Work on the other 2 contract tasks will take place later in the course of the project duration in 2024 and 2025 (the project end is 30.11.2025 but tasks to be delivered until mid 2025 as of the project plan).

General Instructions

You are asked to base your response on the information contained in this document. All responses must be submitted as emails and PDF documents (or similar formats). Responses in any other format, or through any other means, will not be accepted.

RfP Point of Contact

Please do not contact SWC / UNDERPIN project team personnel other than those named below without the express permission of those named below. Any such contact may disqualify your response.

Name: First name, last name
Title: Role
Email: email

Co-funded by the European Union

This project has received funding from the Digital Europe Programme under grant agreement No 101123179

Remark: we expect the recipient of this RfP to treat the RfP and all related information strictly confidential.

Evaluation of Proposals

SELECTION CRITERIA

Demonstrated competence to complete the full scope of work SWC is looking for in relation to the listed UNDERPIN tasks in this RfP (section: Scope)

Demonstration / description of expertise in the required fields and technologies of your organisation and the personnel carrying out the work (CVs, experience, reference projects)

Cultural fit and willingness to be a true collaborative partner

Responsiveness and timeliness of participation in the RFP process

Performance of response(value for money)

Deliverables

Based on the scoping details provided, you are asked to provide a proposal which includes the topics listed below. Please keep your proposal thoughtful and concise, while keeping it as brief as possible. **Please refrain from providing generic content.**

- Capabilities document: company description including expertise and experience in the requested field of expertise and technology, short CVs of personnel provided in case of contracting
- Pricing for the 3 tasks and total price (and the included person-days and services delivered per task)
- Description on how you can support / carry out the listed 3 tasks in the scope section of this RfP
- Contract Redlines (if no redlines are submitted we expect you to accept the contract template provided)