DELIVERABLE D2.2 – REQUIREMENTS, BASELINES, KPI'S, ARCHITECTURE & SPECIFICATIONS

PROJECT ACRONYM:	DS2
PROJECT TITLE:	DataSpace, DataShare 2.0
GA NUMBER NO.	101135967
WEBSITE:	www.dataspace2.eu
DUE DATE OF DELIVERABLE:	30.9.2024
SUBMISSION DATE:	20.2.2025
LEAD BENEFICIARY:	ITC and ICE ES
LEAD AUTHORS:	Tomaz Zadravec (ITC), Tomaz Bokan (ITC), Stuart Campbell (ICE-ES), Noel Tomas (ICE-ES)
REVIEWERS:	Eliot Salant (IBM) and Elais Dakos (ATC)
TYPE:	R
DISSEMINATION LEVEL:	Public

DISCLAIMER

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or granting authority. Neither the European Union nor the granting authority can be held responsible for them.

STATEMENT OF ORIGINALITY

This Deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

DOCUMENT HISTORY

VERSION	DATE	DESCRIPTION	NAME	ORG
V0.1	12.4.2024	Table of content	T. Zadravec	ITC
V0.2	10.5.2024	Inputs from use case partners	A. Tanase, R. Tamas, S. Sukic, P. Tzimotoudis, G. Moutsinas, M. Bunderla, T. Bokan	ATIT, MOMS, ITC, UTH, CROWN
V0.3	30.6.2024	Finalization of use case scenarios and requirements	A. Tanase, R. Tamas, S. Sukic, P. Tzimotoudis, G. Moutsinas, M. Bunderla, T. Bokan	ATIT, MOMS, ITC, UTH, CROWN
V0.4	30.7.2024	Tech Part – Phase I Module definitions	S.Campbell + Multi tech partners per Module	ICE-ES + All partners
V0.4	30.8.2024	Updates and upgrades of the document, Tech Part – Phase II Module definitions	S. Campbell, P. Pussinen, JP. Soininen, T. Zadravec	ICE-ES, VTT, ITC, all partners
V0.5	18.9.2024	Linking requirements with DS2 modules, Tech Part – Phase III Module definitions	A. Tanase, R. Tamas, S. Sukic, P. Tzimotoudis, G. Moutsinas, M. Bunderla, T. Bokan	ATIT, MOMS, ITC, UTH, CROWN + All partners
V0.6	27.9.2024	Review of the the document	E. Salant, E. Dakos	IBM, ATC
V1.0	23.10.2024	Final version of the document	T. Zadravec, T. Bokan, S.Campbell, N.Tomas	ITC and ICE-ES
V1.1	20.2.2024	Revision of the document	T. Zadravec, T. Bokan, S.Campbell, N.Tomas, JP. Soininen, E. Salant, E. Dakos	ITC, ICE-ES, VTT, IBM, ATC

TABLE OF CONTENTS

EXECUTIVE SUMMARY

This deliverable, D2.2 – Requirements Baselines, KPIs, Architecture & Specifications, provides the foundational framework for the successful implementation of DS2. It consolidates key technical and functional specifications that will drive the development of the project's use cases, focusing on how data sharing can optimize decision-making and resource management.

The first part of the document provides an overview of use case descriptions, key performance indicators (KPIs), user requirements, and datasets integration. It highlights four use cases (City Scape, Green Deal, Precision Agriculture, and X-Intersector) to showcase data spaces' potential for cross-sector collaboration, defines KPIs to measure success, identifies user requirements to guide technical development, and explains how diverse datasets will be integrated to enable AI-driven insights and real-time analytics. This part of the document ends with the transition between the initial mapping of the DS2 modules and connections to the identified user requirements, while second part of the document concerns the technical architecture and approach of the DS2 environment.

The nature of DS2 is holistic, modular, pick-and-mix, and platform independent. As such each DS2 Module ("Module") has its own architecture and independent functionality but where 1+1=3 functionality increase can be obtained by linking features of Modules so offering increased exploitation potential. Overarching the Modules is a reference architecture containing additional system/foundation Modules which supports deployment, some common supporting features, and the DS2 trust environment focused on inter-dataspace sharing.

All Modules need to support both the holistic nature of DS2, common dataspace paradigms and expectations such as IDSA, challenging and advancing them, as necessary. The Modules can be used independently and also together.

This document is used as a reference and blueprint for all partners, both users and technicians, to understand the features and relative positioning of DS2 and primarily its Modules within the project and how each Module interacts both externally and with each other. Modules may either enable participants/data for data sharing ("In Dataspace"), across dataspaces ("Inter Dataspace"), or both. Most Modules logically operate OnPremise (although most would be useable by Service Intermediaries) with the remainer operating InCloud at a service intermediary and as part of the DS2 Platform. Many Modules represent the Key Exploitable Results of the project (KERs) with many more as natural outcomes which may become KERs.

Moreover, the document groups the Modules into four main tiers which they will primarily operate in: Tier 0: DS2 Support, Tier 1: DS Marketplace and Deployment; Tier 2: In Data Space Enablement; Tier 3: Inter Data Space Sharing. The nature of each Module which cross-cuts the tiers are: Foundation (Implementation of DS2 would not be possible without these), System (Modules/features used to deploy or support other data related Modules but which are not part of data sharing itself and whose use is conditional on the context or user needs), and Optional (Modules which are generally directly dealing with data).

The Modules are provided by developers to a web portal which provides a marketplace for user selection, and packaging into the Intersector Dataspace Toolkit (IDT), which is a containerised environment for the deployment of Modules to partners. The IDT enables data sharing between IDTs via an embedded connector for those Modules which may need it.

The D2.2 document thus represents an architectural reference for the "big picture architecture" of the DS2 environment and represents the technical vision based on current knowledge which in variably will evolve not least since dataspace technology is novel and technical partners will adapt it to any practical implementation scenario.

1 INTRODUCTION

1.1 Preamble

The motivation of this document is:

- To translate the needs of various stakeholders—ranging from municipalities and research institutions to agronomists and industry leaders—into concrete technical requirements.
- To enable a data sharing environment which facilitates inter-sector and inter-dataspace data sharing.

The purpose is:

- To establish clear use case requirements: Detailing the needs of stakeholders in smart cities, precision agriculture, and other sectors, this document outlines the baseline requirements for achieving the goals set forth by the DS2 project and individual use case.
- To define Key Performance Indicators (KPIs): The document identifies measurable indicators of success for each use case. These KPIs are designed to evaluate both the industry-specific and technology-specific achievements of the use cases.
- To provide the link between user requirements and technical modules: The deliverable outlines the technical modules and system architecture that will enable the seamless sharing and integration of data across different sectors. The initial selection of the modules is done, and linkages with user requirements are provided.
- To provide a modular, secure, trust-sensitive, platform neutral environment for the networked sharing of data.
- To provide an integrated and comprehensive approach, where Modules can seamlessly interact with each other or function independently, and optionally with other Modules, thereby enhancing their value. To identify synergies with other dataspace enabling technologies including partner background, standards etc.
- To specify a series of Modules and how they can enable a "Big Picture" practical and implementable data sharing approach which supports and is validated in the use cases.

The technical architecture is based on the vision of the project which is as "a modular, secure, trust-sensitive, platform neutral environment for the networked sharing of data", aspects further refined from a previous DS2 Deliverable D2.1 "DS2 Problem space definition", and partners continuing expansion of their knowledge in this domain. It relates to DOA Task T2.3: Intersector DataSpace Architecture and Functional/Technical Specifications The purpose is as an architectural reference for the technical work packages WP4, 5, 6 and to an extent WP3. This document breaks the environment down into a series of Modules and each of these is explained in detail on GitHub with links provided in an annex to this document and only an example of the Orchestration Module is embedded in the main body of the second part of the document.

The architecture is focused on gathering Modules from the tasks provided in the technical work pages and fitting them into a coherent holistic and interoperable one for Modules that can take advantage of each other in a 1+1=3 scenario. Beyond the contractual commitments of the DOA there is a natural intersection with Part A of this document which has discussed use cases and user requirements. However, it should be noted that as an RIA the technology is primarily research and innovation driven with the Use Cases providing validation, tuning etc vs a specifically for commercial delivery.

Specifically, the DOA states the following regarding this deliverable: "This task defines the reference architecture for the integrated DS2 approach. The architecture will be enhanced by establishing synergies and adopting key principles from other initiatives, such as those from Gaia-X concerning federation, distributed consensus, decentralisation, transparency, controllability, portability, regulation by automation and interoperability across data and services. IDSA Reference Architecture and BDVA guidelines will also be considered related to data sovereignty, governance, trusted communication, security, and handling of Big

Data. The DS2 Reference Architecture is designed based on established standards such as IEEE 42010 Systems and Software Engineering and will be delivered in different views in terms of representation. The tasks will also define the technical and non-technical specifications based on this architecture for WP3-6 and all components that will realise an integrated DS2 framework. The specifications will consider the user requirements, specific needs of pilot applications (WP7)".

1.2 Glossary and Abbreviations

A definition of common terms related to DS2 as well as a list of abbreviations, is available at https://www.dataspace2.eu/results/glossary

1.3 External Annexes and Supporting Documents

Supporting Documents:

- DS2 Description of Action
- DS2 D2.1 - DS2 Problem space definition

1.4 Reading Notes

In terms of Section references:

- Form 1.2.3.4 references a point in the main body of the document
- Form x.1.2.1. references a section in the module annex representing the same section in multiple Module templates

There are over 20 DS2 Modules presented in a similar template form, and which are located on the Github (https://github.com/ds2-eu). The Orchestration Module (ORC) is used to illustrate many concepts and is located in the main document in Section 11.1 at the end of the main part of the document for convenience.

In the end, it is important to mention that in this document, the term "dataset" is used instead of "data product," as both refer to structured data collections that are essential for analysis, insights, and decision-making within the system.

2 USE CASE DESCRIPTIONS

This chapter presents three pioneering use cases that illustrate how technology and data-driven solutions are being applied across different sectors to address pressing environmental and societal challenges. These use cases are part of larger efforts to create smart, sustainable cities and agricultural systems by leveraging innovative technologies such as IoT, AI, and data spaces. A key enabler of these solutions is the seamless sharing of data across different stakeholders, fostering collaboration and enhancing decision-making through interoperable, multi-sectoral dataspaces.

1. City Scape Case: Focused on Cluj-Napoca, one of Europe's fastest-growing tech cities, this case aims to centralize and harmonize data to help achieve a Net Zero City. The project aims to reduce greenhouse gas (GHG) emissions, improve energy efficiency, and optimize public services by enhancing dataspaces within the city. Collaboration among stakeholders from various sectors—such as public authorities, research institutes, companies, and civil society—will be essential for collecting and sharing city-level data. This cooperative data exchange will enable more accurate, data-empowered policymaking and create a more sustainable urban environment.

Figure 1. City Scape use case

2. Green Deal Case: In the city of Murska Sobota, this use case focuses on monitoring air quality and raising public awareness about pollution levels, especially particulate matter (PM10). Using IoT sensors and AI tools, the use case gathers real-time data on pollutants and GHG emissions, aiming to improve local air quality and provide actionable insights for residents and policymakers. By facilitating data sharing between sectors such as agriculture, transportation, and industry, stakeholders can collaboratively address the root causes of pollution. Integrating these datasets in multiple dataspaces will allow for a more comprehensive understanding of the environmental impacts, ensuring a holistic approach to public health and environmental policy.

Figure 2. Green Deal use case

3. Precision Agriculture Case: This case focuses on increasing crop production while minimizing resource usage through precision farming techniques in Northern Greece. Utilizing IoT sensors and data analytics, the project creates a digital ecosystem for efficient field monitoring and crop management. The connection of datasets from different farms through data spaces enables cross-sectoral data sharing between agronomists, farmers, researchers, and technology providers. This sharing of data across multiple stakeholders allows for more accurate, real-time AI-generated recommendations for optimizing water, fertilizer, and pesticide use, while maintaining data security and sovereignty.

Figure 3. Precision Agriculture use case

These use cases not only highlight the potential of data-driven innovation to address environmental and societal challenges but also emphasize the critical role of data sharing among different stakeholders. By fostering collaboration across multiple dataspaces, these projects can integrate diverse data sources, enhancing decision-making processes, improving resource management, and driving sustainable development across various sectors.

2.1 Use case 1 – City Scape

Primary Goal of the City Scape case: To facilitate the centralisation, sharing and harmonisation of data relevant for the ambition of achieving a Net Zero City for citizens, public authorities, research institutes and companies.
As one of the fastest-growing and most innovative cities in Europe (mainly due to ICT and tech-based innovation), Cluj-Napoca wants and needs to address the biggest challenges of today's society in an effective, sustainable, state-of-the-art manner. Among these challenges, being a green, resilient, and sustainable city has been a top priority in recent years, which brings a set of needs of outmost importance for the Cluj-Napoca Municipality and its partners in the city's development ecosystem (industries, RTOs, citizens', civil society organisations etc). In this sense, Cluj-Napoca is one of the first cities in the world to initiate the adoption of a NetZeroCity Action Plan and in 2023 it was selected as one of the 30 Pilot Cities for the EU's 100 Net Zero Cities challenge.

To this end, the City Scape pilot will enrich two dataspaces DS2 City and ZeroNet that serve the above-mentioned goal, one at the city level and one at the level of a residential building.

The two dataspaces will benefit from the technological solutions developed and implemented in the DS2 project in the following areas:

- Reducing the GHG emissions and increasing the energy efficiency of the city. 78% of GHG emissions in Cluj-Napoca are generated by the built environment, with the residential sector accounting for over half of these emissions. Transport accounts for roughly 21% of the GHG emissions. The challenges are to find data-based scenarios and solutions in order to increase the energy efficiency of buildings, to reduce the traffic congestion and to reduce pollution.

- Improving the data infrastructure, data agile economy and complex data empowering policymaking. Currently data is used for informing policy related to achieving net-zero goals (energy consumption, green energy production potential, traffic, waste etc). However, is based on statistical proxying and lacks the granularity, consistency and interoperability needed for complex modelling. Thus, a better and more accurate city-level data collection, open data availability and the interoperability and service integration of data represent a strategic priority in Cluj-Napoca Digital Transformation Strategy.
- Optimisation of public services based on citizen's needs and increased awareness at all levels as to the feasible and recommended solutions to change the state of play.

Building upon existing initiatives and Horizon projects (such as REFLOW – which defined the city's transition towards circular economy and REFLOW OS, ZEROW – which started implementing a platform for Food Loss and Waste (https://www.zerow-project.eu/innovation/flw-monitoring-and-assessment and Net Zero Cities, UIA Future of Work), the DS2 City and ZeroNet Dataspaces will provide a cross-sectorial framework for reducing CO2 emissions, by gathering and sharing relevant data from various sources and sectors (eg: traffic, heat islands, waste data, electricity consumption, electromagnetic emissions etc).

2.1.1 City Scape use case participants

This chapter presents the main participants who will be active in the use case and describes their role and relevance to the use case.

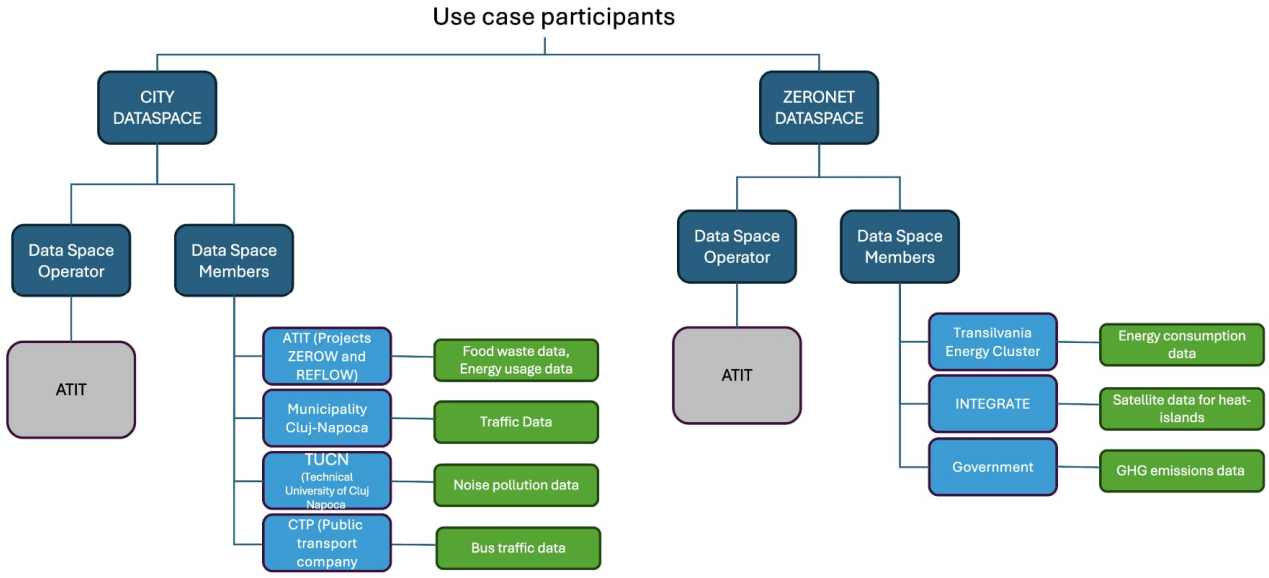


Figure 4. City Scape use case participants

Actors	Role	Relevance
Municipality of Cluj-Napoca (Public Institution/Local Authority)	Data owner and Data provider	The municipality of Cluj-Napoca is a data owner, providing traffic data for the City Dataspace
TUCN (Technical University of Cluj-Napoca) (RSO)	Data intermediary, Data Consumer	Intermediating data on pollution and traffic. TUCN will also act as a data consumer, facilitating researchers to access traffic and city pollution data.
CTP (Public Transport Company from Cluj-Napoca)	Data provider	CTP provides data related to public transport

TREC (Transilvania Energy Cluster)	Data provider and Data intermediary	Providing data on energy consumption
INTEGRATE	Data intermediary	Indeco Soft intermediates through the INTEGRATE project, satellite data related to heat islands
ATIT (Transilvania IT Cluster)	Data intermediary	ATIT intermediates data related to Food Loss and waste and energy usage provided by existing initiatives and Horizon projects like ZEROW and REFLOW OS. ATIT is also the operator of the two dataspaces.
Government institutions	Data provider	Provides data related to GHG emissions

Table 1: Description of City scape participants

2.1.2 Use case implementation

The goal of this use case is designed to optimize the efficiency of resource usage, aiming to reduce climate-impacting emissions and to correct missing or lacking information regarding climate neutrality across citizens, industries, and the public sector. This use case seeks to empower stakeholders at all levels by providing data-driven insights that can influence sustainable practices, while fostering a deeper understanding of the paths toward achieving a Net Zero City.

However, the implementation faces certain constraints. One significant challenge is the need to model the interactions between various sectors' data with the city's net zero objectives. While a model has been developed for a residential digital twin solution—building upon the Nearly Zero Energy Buildings (NZEB) regulations—extrapolating this to a city-wide level requires further refinement. Additionally, the availability of large, granular datasets remains a constraint, as detailed, real-time data is critical for accurate modelling and decision-making.

The DS2 project is expected to provide integrative technology that is currently missing. At present, there is no comprehensive system that can integrate diverse datasets from different sectors or offer recommendations that reflect the complexity of these interactions. The DS2 project will bridge this gap by developing a harmonized data environment that incorporates various data types, enabling more accurate assessments and suggestions for achieving the city's net zero goals.

A data space is essential for the operationalization of these efforts. It will allow the Municipality to effectively manage the green and digital transitions by enabling stakeholders to access, share, and analyze relevant data. The data space also supports the development of innovative, state-of-the-art solutions in core sectors of the economy and society, ensuring that all players—from public authorities to industry leaders—are aligned in their efforts to achieve sustainability goals.

2.1.3 Datasets availability

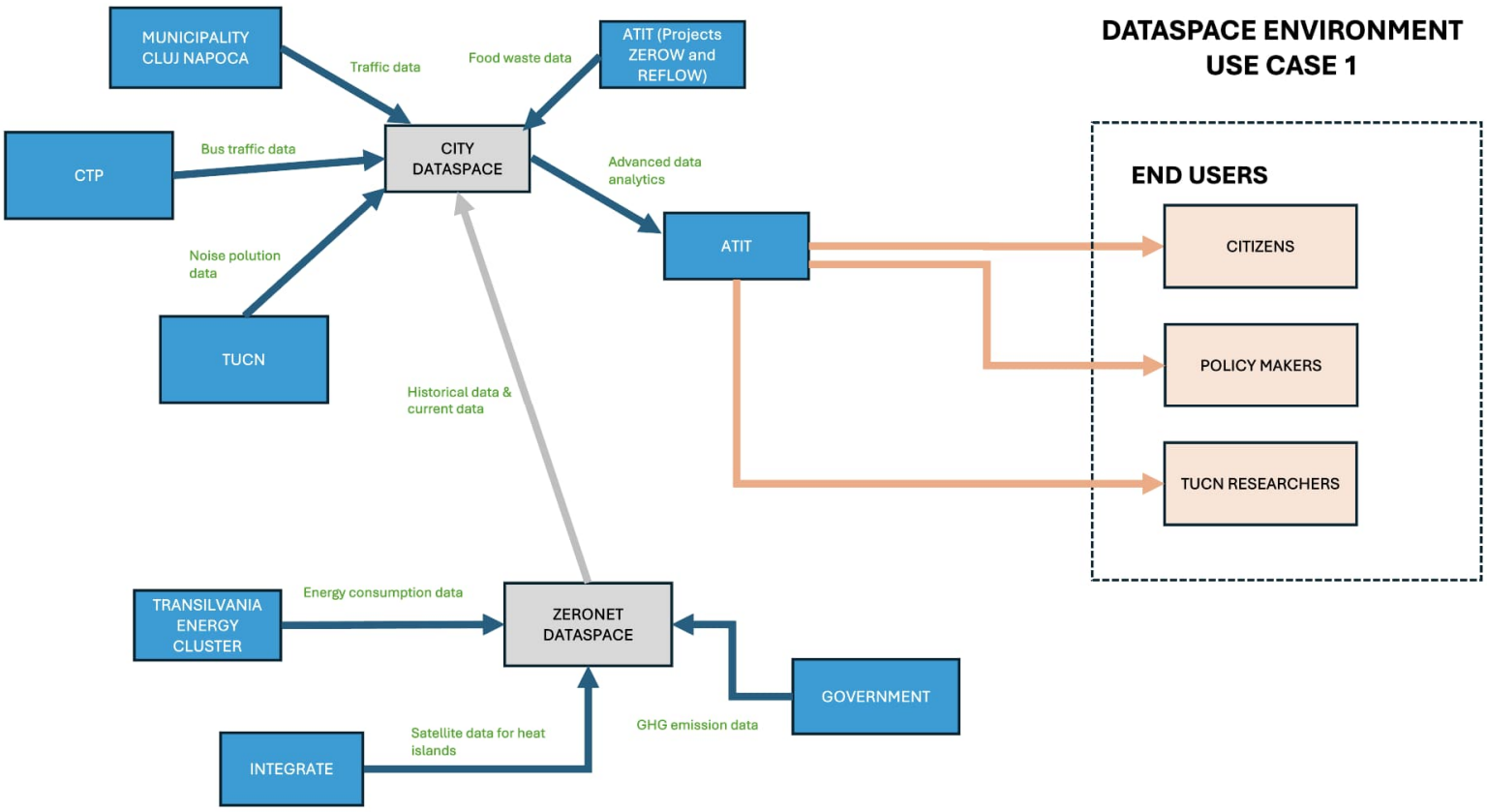


Figure 5. Dataset availability in City Scape use case

DATASET	OWNER OF THE DATASET	DATA TYPE AND DESCRIPTION
Traffic Data	Municipality of Cluj-Napoca	Data related to traffic and city mobility such as parking spaces, car traffic data, bike users, to be served into the City Dataspace.
Public Transport Data	CTP (Public Transport Company)	Data related to public transportation gathered from busses to be served into the City Dataspace.
Noise Pollution Data	TUCN (Technical University of Cluj-Napoca)	Data related to noise levels across the city are gathered by sensors. TUCN acts as a provider for these data to serve them into the City Dataspace.
Food Loss and Waste Data	ATIT	ATIT intermediates data related to Food Loss and waste and energy usage provided by existing initiatives and Horizon projects like ZEROW and REFLOW OS. The data is served into the City Dataspace.
Energy consumptions data	TEC (Transilvania Energy Cluster)	Data related to energy consumption that is served into the ZERONET dataspace
Satellite data	IndecoSoft	Data related to heat-island formations gathered by IndecoSoft using earth observation data, served into the ZERONET dataspace.
Government data	Public data	Data related to GHG emissions served into the ZERONET dataspace

Table 2: Description of City Scape datasets

2.2 Use case 2 – Green Deal

Primary Goal of the Green Deal case: To gather air quality data and inform citizens about elevated levels of PM10 particles in the city of Murska Sobota.

Air pollution, global warming, and other pollutants have a great impact on the environment, which has become a major global concern for many years. The most relevant sectors/areas contributing to air pollution and global warming are considered to be traffic, energy, industry, agriculture, and households. There are different harmonised methodologies in place for monitoring pollutants. However, the majority of them are linked to gathering statistical/reporting data from predominant actors and representatives of such sectors.

Determination of root causes and their direct influence in specific local environments (micro locations or hot-spots) are dependent on gathering/processing data from many sources (such as IoT, Satellite, citizen data, public data sources and 3rd party data sources). The challenge is that those data are mostly gathered and processed by different actors and siloed into particular sectoral/geographical framework and hence the use and benefit of DS2 in assisting to 'network' such data.

CO2 has the largest contribution (as much as 82.8% in 2018) in the total share of GHG emissions in Slovenia. CO2 is mainly produced during the combustion of fuel and from industrial processes. It is followed by methane – CH4 (11.1%), which mainly originates from waste and agriculture, and di-nitric oxide – N2O (4.3%), which is also mainly produced in agriculture. Nitrous oxide emissions from road traffic are also noticeable. Among the sectors which are not included in emissions trading (mainly those are all energy producers and large energy dependant industries), transport is the most important polluter, contributing 52.9% of all emissions in 2018.

The next important source is agriculture, which contributed 15.6% of emissions. In 2018. With a share of 11.9%, emissions from the use of fuels in households and in the commercial-institutional sector are the third most important source. In 2018, these emissions were 8.3% lower than the previous year and as much as 46.4% lower than in 2005. This reduction in recent years is largely influenced by milder winters, as well as improved insulation of buildings and increased use of wood for heating, as CO2 emissions from biomass are not considered. All other sources that still contribute to emissions outside the ETS are: Remaining use of fuels in industry (6.7%), waste management (4.0%), remaining process emissions (4.1%), and others (4.7%).

The source of all this data is, in the majority, provided by Slovenia's statistical office based on regular/yearly reporting covering the whole country and not backed up by a gathering of IoT or other respective data sources. By providing frequent and systemic gathering and processing of the range of data, the Green Deal pilot will establish direct communication with residents and local stakeholders by providing a notification system about air pollutants and GHG emissions. Additional explanatory data on the root causes and impacts of key localized activities—such as agriculture, traffic, industry, and household heating—will enable awareness-raising and the implementation of operational and policy-related strategies.

MOMS is currently collecting data on air pollution with help of sensors in the city of Murska Sobota. Within the scope of the project, there is a plan to install additional sensors around MOMS to have increased and more precise data gathering. The primary goal is to gather air quality data and inform citizens about elevated levels of PM10 particles in the city of Murska Sobota. Additionally, it notifies public institutions such as kindergartens, schools, and elderly care facilities via email in the case of increased PM10 levels, allowing them to adjust their indoor activities accordingly.

Figure 6. Current online notification system and Sensors installed on one of the public buildings in MOMS

Energy consumption and waste are major contributors to greenhouse gas (GHG) emissions. Consequently, the city plans to incorporate information from the Municipality of Murska Sobota in its use case. The city can access state-owned data as well as data from waste and utility companies owned by the Municipality of Murska Sobota.

Traffic, households, and industry contribute to air pollution and greenhouse gas (GHG) and also the agricultural sector also plays a significant role. Therefore, the Municipality of Murska Sobota (MOMS) is seeking additional datasets to provide a comprehensive overview of the city's pollution. By integrating data from the agricultural sector, MOMS aims to offer citizens and other users a detailed and enhanced understanding of air pollution in both the city and the surrounding region. To achieve this, MOMS plans to collaborate with Digital Innovation Hub AGRIFOOD DATA SPACE (DADS), which possesses numerous agricultural datasets that can help assess the impact of air pollution in the city.

Finally, the goal of the use case is to use the AI tool provided by the University of Maribor, which can help with enhancement of the different datasets to receive better and quality data for the users at the end. AI tool can significantly enhance urban air quality monitoring by integrating diverse datasets, such as industrial emissions, energy consumption, agricultural activities, and earth observation. By processing this data, AI can predict pollution levels, detect anomalies, and provide high-resolution air quality maps. This comprehensive approach enables real-time monitoring, informed policy-making, and effective public health advisories, ultimately leading to improved air quality and urban sustainability.

2.2.1 Green Deal use case participants

This chapter presents the main participants who will be active in the use case and describes their role and relevance to the use case.

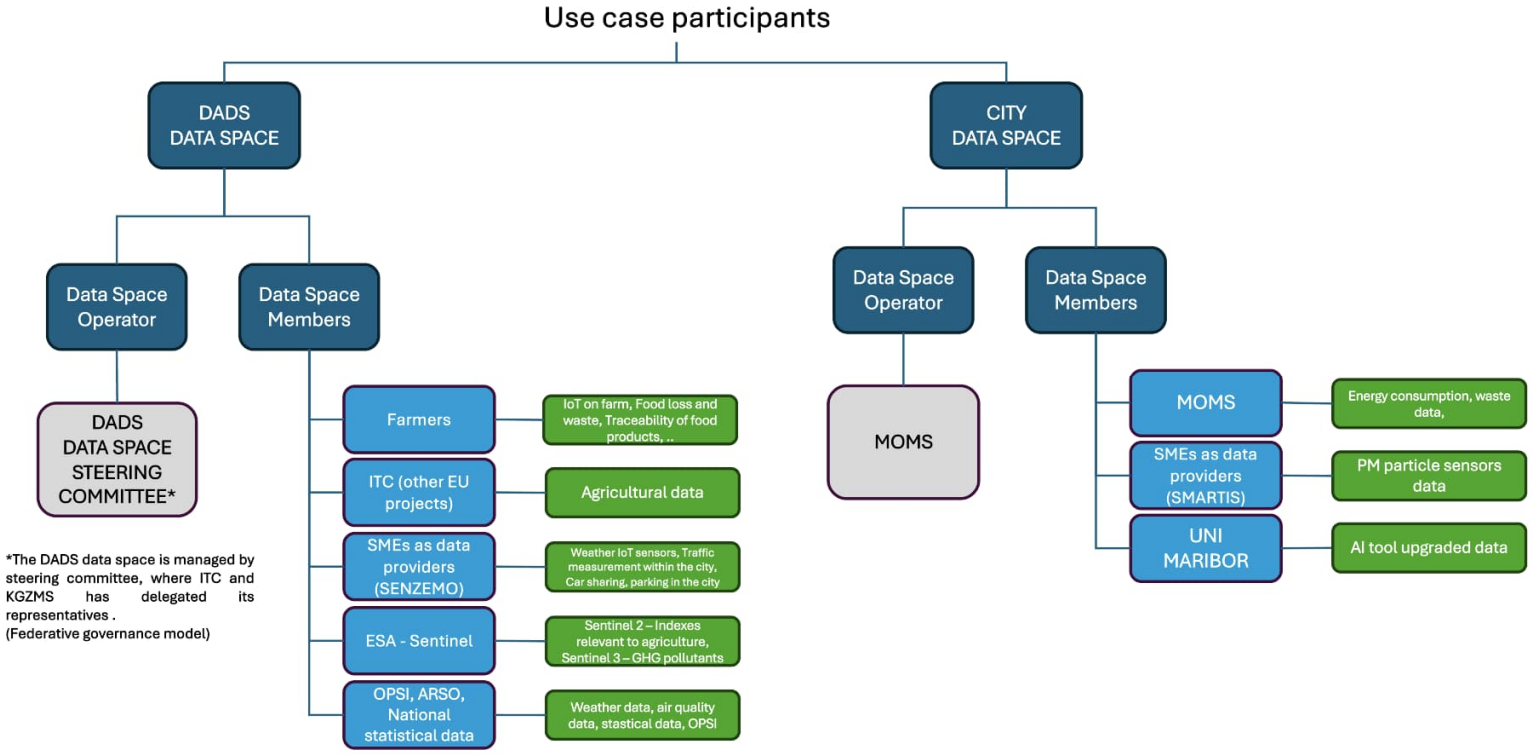


Figure 7. Green Deal use case participants

Actors	Role	Relevance
MOMS – Municipality of Murska Sobota (Public institution)	Data owner, Data consumer	MOMS is a data owner, providing data about: • Energy consumption of municipality owned buildings • Waste data in the municipality MOMS is also a data consumer, where they will consume the data provided by UNI MARIBOR and their AI TOOL, which will enrich the data on air quality in the city. To enhance public awareness and support policy and operational strategies, the MOMS will create a web portal to present the city's air quality. This portal will include explanatory data on the root causes and impacts of the most influential local agricultural, traffic, industrial, and household (heating) activities. Additionally, a notification system will be introduced to alert citizens and public institutions (such as kindergartens, schools, hospitals, and elderly homes) about air pollutants and GHG emissions. Users will be informed about increased air pollution through the website and email notifications.
ITC – Innovation Technology Cluster Murska Sobota (Business support organization)	Data provider	ITC's role is as an operator of the DADS dataspace and as a data provider when it comes to different applications and data which are provided by the farmers. They are a bridge to connect data from farmers to the dataspace.
Farmers (individual organizations)	Data owner	Farmers own data regarding their production (amounts, type of production,...) and this data is shared in within the DADS dataspace. There

		is also data about food traceability, food loss, and waste generated by the supply chain.
UNI MARIBOR - University of Maribor (Higher education and research institution)	Data consumer	The University of Maribor is a developer of an AI tool that can enhance different datasets to receive better and higher quality data for users. They act as a user in the MOMS dataspace and their role is to also connect to the DADS dataspace to search for datasets related to the agricultural sector.
Other EU projects (initiative, business support level)	Data owners	Different projects are linked to the DADS dataspace, mainly generating data about the farm production and agricultural data. These projects are ZEROW, GREEN.DAT.AI and DIVINE. ITC will act as a data owner as partner and pilot region in those projects.
SMEs as data providers	Data providers	Different SME's have their own applications and services and acts as data providers within the DADS and MOMS dataspace. For example, IoT sensor provider (SENZEMO), which is collecting data from sensors, stores it on its database and provides access to the data through DADS. On other hand, SMARTIS is the developer of the IoT sensor for measuring air quality and will be acting as data provider in the MOMS dataspace.
Citizens (general public)	Data users	Citizens will act as data users, accessing the MOMS web portal for detailed information on the city's air quality. By subscribing to the email notification service, they will also be promptly informed about any increases in air pollution.
SMEs as data users	Data users	SMEs will have access to rich data that will be generated by the AI TOOL. By this they will have the opportunity for collaboration to develop innovative solutions (development of various tools and mobile applications,...).
Government institutions (Ministry, Statistical offices,...)	Data users	Access to real-time and historical air quality data enables the government to make informed decisions and develop effective policies and regulations to mitigate air pollution. This data can also be used to monitor the effectiveness of implemented measures and adjust strategies accordingly. By providing timely information about air quality and pollution levels, the government can help citizens take preventive measures to protect their health.

Table 3: Green deal use case participants

2.2.2 Use case implementation

The primary goal of the Green Deal use case is to enhance urban management and governance by leveraging data from multiple sources through a federated approach. In particular, data from the MOMS and DADS data spaces will be integrated to facilitate informed decision-making. The federation of these data spaces is a key aspect of the project, ensuring that diverse datasets—ranging from sensor networks, public institutions, and citizen feedback—are harmonized and made accessible. This approach enables several critical optimizations:

- Efficient Resource Allocation: By federating data from MOMS and DADS, decision-makers gain access to a unified, comprehensive dataset, allowing for better allocation of resources such as funding, personnel, and infrastructure to tackle urban challenges more effectively.
- Improved Service Delivery: By analysing data from multiple sources, public services can be tailor to meet the specific needs and preferences of residents, enhancing the quality and efficiency of service delivery.
- Enhanced Urban Planning: The federation of MOMS and DADS data spaces provides a holistic view of the urban landscape, aiding planners in making informed decisions regarding land use, transportation, infrastructure, and environmental sustainability.

- Proactive Problem Solving: By leveraging real-time and historical data, city officials can identify emerging issues and trends early, allowing for proactive interventions and preventive measures to mitigate potential challenges.
- Community Engagement and Empowerment: Through the incorporation of citizen feedback and behaviour data, greater community engagement and empowerment can be fostered and residents enabled to actively participate in the decision-making process and contribute to the improvement of their neighbourhoods.

The use case aims to revolutionize urban governance, promote sustainability, and elevate the quality of life for all residents in the region by optimizing the collection and integration of information from diverse sources. To achieve this, a dedicated data space is essential for securely and effectively sharing data, integrating varied datasets, enabling real-time monitoring and analysis, and ensuring strong data governance and security. This centralized platform for urban data management empowers cities to fully leverage data-driven insights, addressing complex urban challenges and significantly improving residents' quality of life. Through this approach, cities can harness the full potential of data to drive informed decision-making and foster sustainable, resilient communities.

By participating in the DS2 project, the Green Deal use case aims to address following key areas:

- Federation of Data Spaces: Since the DS2 project centres around data space federation, a critical focus is on federating the MOMS and DADS data spaces. This federation will enable seamless data sharing and integration between these two data spaces, enhancing the ability of stakeholders to access and use comprehensive datasets for improved decision-making and urban management.
- Business and Stakeholder Engagement: The use case seeks to explore and define innovative approaches for establishing a viable business model for the data space. This includes identifying strategies to enhance the incentives for data owners to share their data, thereby fostering greater collaboration and engagement among stakeholders.
- Technical Development: On the technical front, the DS2 project should aim to develop standardized components that will enable stakeholders to achieve compatibility and interoperability across various data-sharing initiatives. Examples of these components include standardized identity management systems, connectors, data catalogues, and logging services. Additionally, tools for data quality improvement, such as those for data harmonization, orchestration, and transformation, could be developed to ensure the integrity and usability of shared data.

2.2.3 Data sets availability within the Green Deal use case

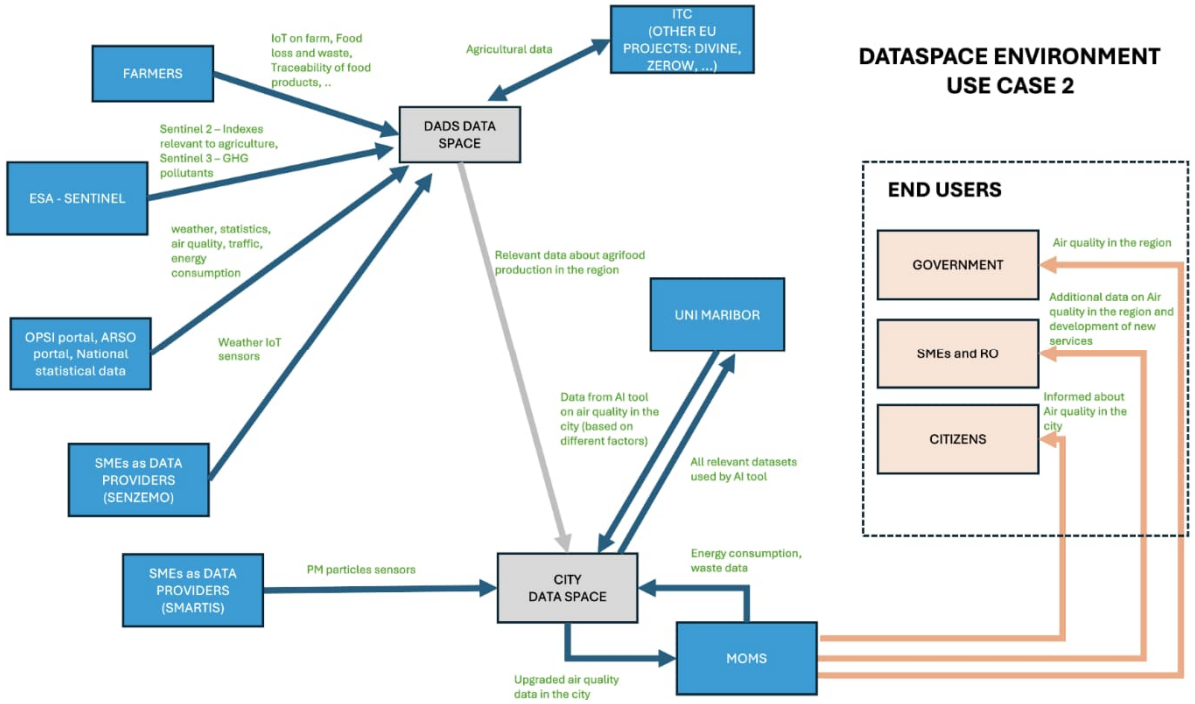


Figure 8. Data sets availability in the Green Deal use case

Diagrams shows only part of relevant datasets in the Slovenian use case, that are currently available and accessible via both data spaces. The table below is providing also the ones that partners of the use case are working on to connect to existing data spaces by the end of 2024.

Dataset	Owner of the dataset	Data type and description
ZEROW PROJECT	*FARMERS AND SUPPLY CHAIN ACTORS*	Data about food supply chain including food loss and food waste created. ITC as project partner acts as a data provider to serve those data into the DIH AGRIFOOD DATASPACE.
DIVINE PROJECT	*FARMERS*	Economic data related to milk production in Slovenian farms participating in benchmarking sessions organised by farming advisory service KGZMS. ITC as a partner in the project will act as a data provider to serve those data into the DIH AGRIFOOD DATASPACE.
GREEN DAT.AI PROJECT	*FARMERS*	Data related to the farmers production that are part of the project GREEN.DAT.AI. Project partners acts as data providers to serve those data into the DIH AGRIFOOD DATASPACE.
EARTH OBSERVATION DATA	*PUBLIC DATA*	Earth observation data for Murska Sobota city from the public datasets of the European GNSS (Sentinel 2 – Indexes relevant to agriculture, Sentinel 3 – GHG pollutants). The data will be taken from ESA SENTINEL and will be provided by ITC to the dataspace.
WEATHER DATA	*PUBLIC DATA*	Weather data from national weather agency and data portals.
ENERGY CONSUMPTION	*MOMS*	Energy consumption (heating) data from public/private buildings of the city administration.
AIR QUALITY DATA	*PUBLIC DATA*	AIR quality data from the national weather agency and data portals. This data encompasses various pollutants such as

		particulate matter (PM10, PM2.5), nitrogen dioxide (NO2), sulphur dioxide (SO2), ozone (O3), and carbon monoxide (CO).
SENSEDGE IOT	*FARMERS*	Measurement from Agricultural IoT devices (weather station, soil sensor) which are setup at different farmers in the region.
STATISTICAL DATA	*PUBLIC DATA*	National statistical data relevant for the use case.
OPSI DATA	*PUBLIC DATA*	OPSI is a national open data portal. It allows access to different open data in Slovenia.
PM PARTICLES SENSORS	*MOMS*	2 PM particle sensors are installed in the city with relevant data about air quality in the city.
GEOTHERMAL RESOURCES	*MOMS*	List and register of geothermal wells/spots in the Municipality of Murska Sobota.
TRAFFIC DATA	*PUBLIC DATA*	Relevant traffic data in Slovenia.
WASTE DATA	*PUBLIC DATA*	Amount of waste generated by the citizens in the Municipality.
HEATING DATA	*PUBLIC DATA*	Data on type of heating device per citizens.
FOOD SUPPLY CHAIN DATA	*FARMERS AND SUPPLY CHAIN ACTORS*	Process flow from farm to the retail store. Events are stored in the blockchain network.
CYCLIST DATA	*PUBLIC DATA*	Counter of cyclist in the City of Murska Sobota.
OVERNIGHT STAYS	*MOMS*	Number of overnight stays in the Municipality of Murska Sobota.

Table 4: Description of Green Deal datasets

2.3 Use case 3 – Precision Agriculture

Primary Goal of the Precision Agriculture case: To increase crop production while minimizing the use of resources by utilizing technology and data-driven decision support systems tailored to specific area characteristics.

Precision agriculture is increasing in importance as a mechanism to increase crop production, whilst minimizing the use of resources (e.g. fertilizers, water/irrigation systems, etc.) based on the specific area characteristics (e.g. microclimate, soil quality, sunshine periods, etc.). As technology and agronomic research has improved, precision farming has demonstrated stronger results and outcomes if various technologies and sectors are stacked together. For the effective monitoring of a field crop, the deployment of low-cost sensing devices and associated edge nodes are critical for collecting data and uploading them to a core cloud. At the core cloud, data analytics and Machine Learning (ML) solutions can be implemented, aimed at decision support for optimizing the usage of the available resources to accomplish maximum yields at the minimum possible cost and environmental footprint.

The University of Thessaly (UTH) (tech and data provider) works closely with CROWNEST SA (SME, use cases provider, innovation broker, exploitation manager) to form a smart-farming alliance that seamlessly connects the entire food value chain by enabling capabilities like shared information and measurements from the field, decision support tools drawing on diverse data pools, integrated crop management across regions, etc. UTH brings significant technical expertise in IoT sensors and networks and has formed a long-standing partnership in the form of an established dataspace, DigiAgro DS, with several farm owners across the Northern Greece. Additionally, it has developed and deployed a comprehensive IoT solution in each farm, aiming at collecting high-quality data regarding environmental conditions and crop growth. On the other hand, CROWNEST is associated with experienced agricultural scientists and data analysts who bring their own models for crop productivity predictions as well as for resource management and optimization, and which provide as recommendation services to the farmers and agronomists, in an established dataspace, named AgroScience DS.

UTH and CROWNEST operate a powerful IoT-based ecosystem, called AgroNIT, that is developed for the distributed monitoring and field data collection to enable real-time decision support. The system is composed of three components: 1) the edge system; 2) the cloud computing infrastructure; and 3) the decision-making component. The architecture diagram of the AgroNIT ecosystem can be seen in the figure below.

The AgroNIT testbed has been designed to provide common access to facilities and services fostering the development and the validation through experimentation of novel cultivation practices, as well as to enable the digital transformation of Agriculture by applying advanced cultivation methods with digital tools and connected machinery. AgroNIT brings emerging digital technologies to the disposal of agriculture stakeholders (such as farmers and scientists) in order to advance agriculture practices in multi-fold dimensions: From the effective management of fertilizers and pesticides to the optimization of the irrigation processes tailored to the needs of each crop cultivar aiming to improve crop productivity whilst minimizing crops' environmental footprint, as well as to optimize the logistics of the processing industry through the real-time sampled monitoring and estimation of expected crop productivity at different areas of interest. In AgroNIT smart decision support systems provide real-time AI-generated consultancy to farmers and agronomists by relying on sensors that automate field measurements collection. Furthermore, agricultural scientists have access to a large repository of data measurements and novel data analytic services, being able to design, apply and validate new cultivation protocols.

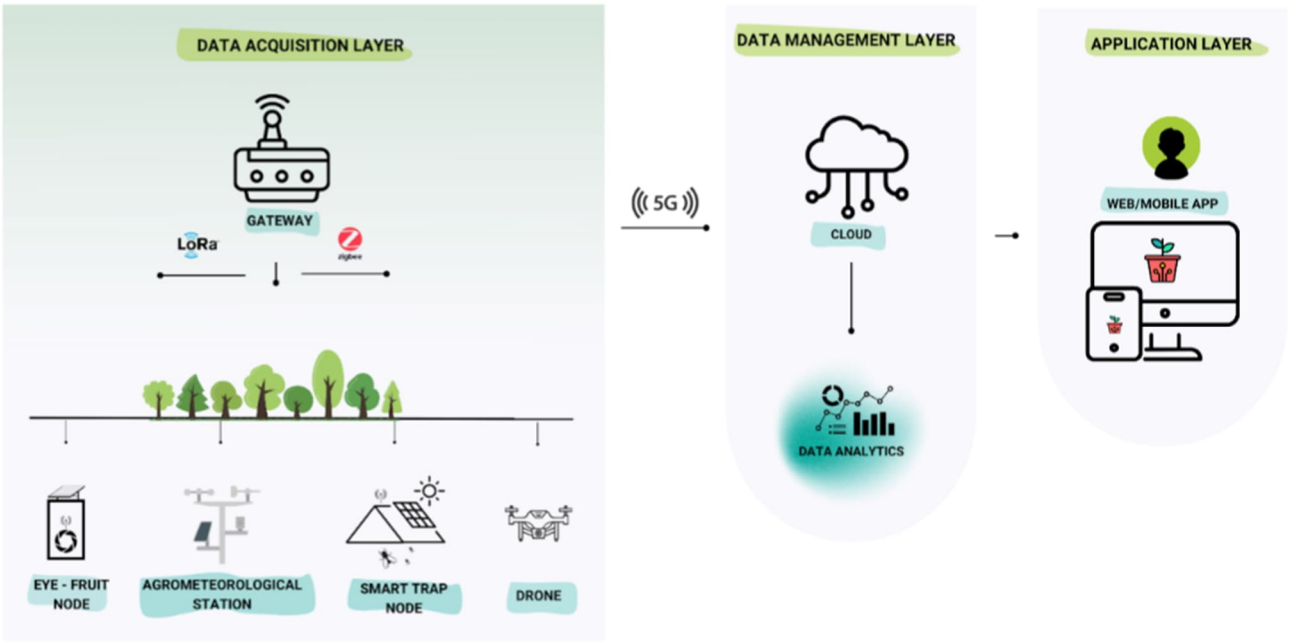


Figure 9. The AgroNIT ecosystem architecture diagram

When data is produced from different crop owners, aspects such as sovereignty, rights management, data protection and trust need to be ensured. The establishment of a dataspace based on the principles of IDSA and GAIA-X initiatives assists in overcoming these challenges. However, to involve actors from adjacent sectors and create sustainable business models, the need for data interoperability as well as the seamless integration of knowledge from several dataspaces is urgent. To this end, the implementation of tools capable of interconnecting more than one dataspace is of great importance.

Within the DS2 project, an intersectoral data environment will be created which will allow data that is hosted in DigiAgro DS, and originating from earth observation sensors, weather stations, and cameras monitoring crop growth, to be combined and consumed by AI and agroscience services that are hosted in AgroScience DS, in order to provide detailed and accurate recommendations for efficient crop management and growth. Based on this data, and processes around it, an example service that will be offered through a Federated Learning

(FL) approach, developed by CROWNEST, will enable the monitoring or even prediction of crop growth/maturity, soil water availability, as well as crop water requirements in certain types of tree crops (i.e., peach cultivars). With the FL, the data is retained, and partial ML models are trained at the edge node, minimizing data transmission to the core cloud and increasing the security of the overall solution. To apply the FL scheme, issues regarding data leakage through model updates will be investigated and subsequently the DS2 modules will be employed to ensure data protection, sovereignty, rights management, and compliance with GDPR regulations.

2.3.1 Precision agriculture use case participants

This chapter presents the main participants who will be active in the use case and describes their role and relevance to the use case.

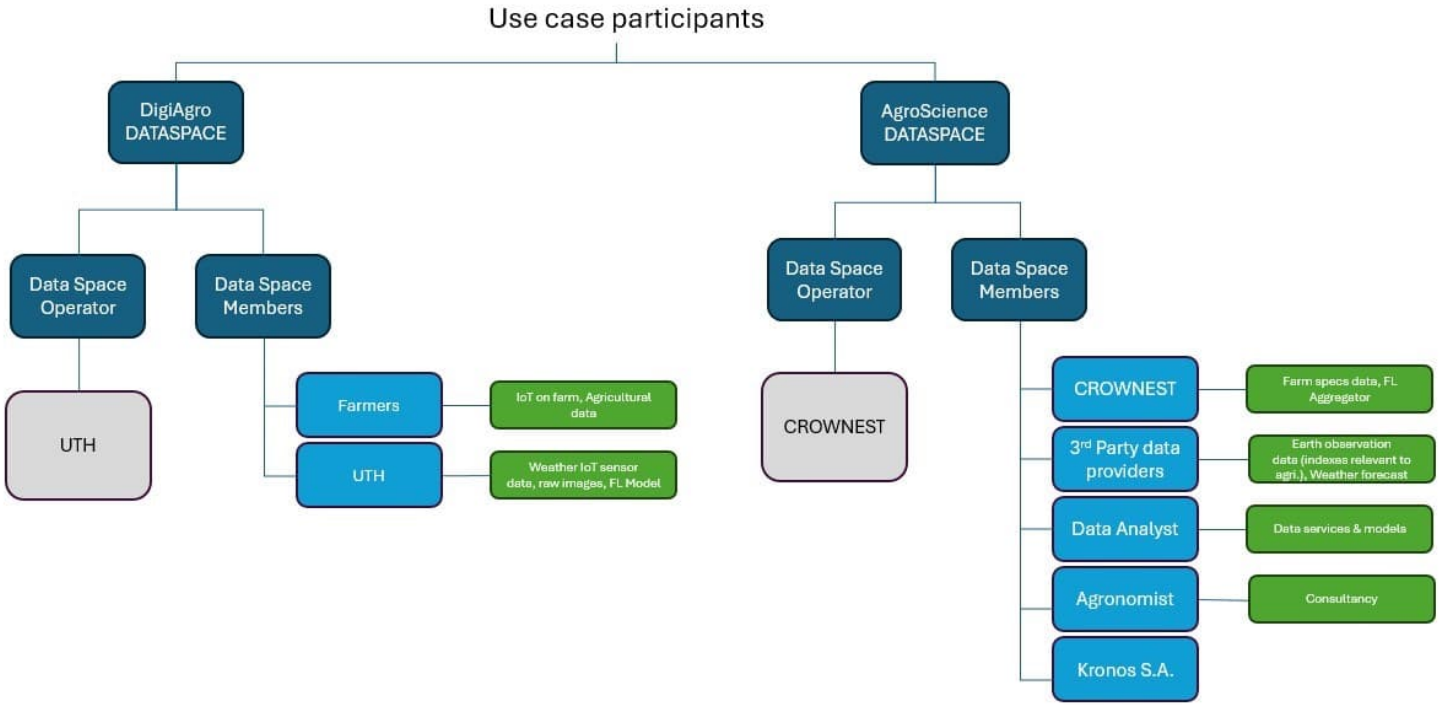


Figure 10. Precision agriculture use case participants

Actors	Role	Relevance
UTH - University of Thessaly (Public institution)	Data Owner, Data Provider, Data consumer, Data Space Operator	UTH is the operator of the DigiAgro DS. It also participates in the DS with the following roles: As data owner, providing data on: • Weather conditions (based on environmental measurements by the deployed IoT sensor nodes) • Fruit images (raw images are collected from the deployed field-based camera nodes) • FL model (AI-enabled model capable of extracting from the colour palette and size from the captured fruit images) As a data consumer, since it consumes data from the AgroScience Dataspace that is useful for the end-users (i.e., the farmers) and visualised via the AgroNIT app/platform.
Farmers (individual organizations)	Data owners, Data provider, Data user	Farmers are data owners, participating in DigiAgro DS and providing data on:

		• The soil and local climate conditions, based on measurements obtained by the deployed IoT sensors installed in-situ • Crop management-related data Farmers also consume this type of data together with precise consulting recommendations through the AgroNIT platform as users.
CROWNEST (SME)	Data Owner, Data Provider, Data consumer, Data Operator	CROWNEST is the operator of the AgroScience DS. It also participates in the DS with the following roles: As data owner, providing data related to the specifications of each farm, including relevant (static) data about the registered crop fields, associated farmers and other. As data/service provider by operating the aggregator of the FL model that is capable of extracting the colour pallet and size of fruits from the obtained fruit images. As data/service consumer by consuming model updates from each individual FL model that is deployed at the edge through the interconnection with the DigiAgro DS.
3rd Party data providers	Data provider	3rd Party data providers like Weather Forecast Agency and space agencies or Earth observation satellite agencies participate in the AgroScience DS with data such as weather forecasts and satellite imagery.
Data Analysts	Data Consumer, Data Service provider	Kronos SA is a large canning industry that is interested in receiving information on crop maturity rate and expected productivity/yield, in real-time, across its operation area, aiming to optimize the logistics behind the receipt of raw material (peach fruits) during the harvesting period.
Agronomists	Data Intermediary, Data consumer	An agronomist (aka agricultural consultant), participating in AgroScience DS, can directly use or consume data mainly to transform it to useful, ready-to-use~~r~~ information in the form of consultancy for the associated farmers.
Kronos SA (Large processing industry)	Data user	Kronos SA is a large canning industry that is interested in receiving information on crop maturity rate and expected productivity/yield, in real-time, across its operation area, aiming to optimize the logistics behind the receipt of raw material (peach fruits) during the harvesting period.

Table 5: Description of Precision Agriculture participants

2.3.2 Use case implementation

The primary goal of this use case is to address the challenge of creating sustainable business models by incorporating data and services from complementary sectors that are managed within different dataspaces. Specifically, the precision agriculture use case aims to enable farmers and agronomists to extract value from sensor data gathered from IoT deployments on the fields, which is then combined with weather and satellite data. This diverse data is processed and analysed by AI and optimization models provided by AgroScience Data space, allowing for more accurate and efficient agricultural consultation, resource optimization, and precise estimation of critical agricultural indicators.

The integration and use of project will bring value to a variety of stakeholders. Farmers benefit from gaining sovereignty over their data while receiving precise, personalized consultancy that promotes efficient crop

management and minimizes resource waste. Agronomists have access to advanced AI and AgroScience models, empowering them to provide more accurate and effective recommendations. Data scientists benefit from the ability to access vast amounts of data and offer their models as services. Meanwhile, the processing industry receives real-time information on crop maturity and expected yield, enabling them to optimize logistics for raw material procurement and processing.

The optimization objectives of the use case are centred around improving resource utilization, increasing crop yield, and enhancing cost efficiency. Through precise monitoring and management of growth conditions, the use case supports maximized yield and optimized resource allocation, which reduces unnecessary expenditures. The use case also promotes sustainable farming practices, aiming to reduce the use of chemicals, minimize environmental impact, and improve soil health through optimized crop rotation. In addition, it helps mitigate risks related to weather variability by utilizing accurate forecasts and real-time data. This leads to improved decision-making and better yield prediction, supporting supply chain planning and market positioning. Moreover, seamless data integration from various sources is essential for comprehensive analysis and ensures interoperability across different dataspaces.

However, several constraints need to be addressed before full implementation of the use case. Ensuring high-quality, timely, and accurate data from IoT sensors, weather forecasts, and satellite imagery is crucial for the reliability of AI and optimization models. There is also a need for smooth interoperability between diverse data sources and dataspaces, which requires adopting standardized formats and protocols. The system must be scalable to handle growing data volumes and user interactions, while also addressing data security and privacy concerns. Regulatory compliance, especially around data ownership, sharing, and usage, must be ensured, and optimizing data processing workflows is critical to avoid delays in decision-making.

The DS2 project is expected to deliver key capabilities that are currently missing. It will ensure robust data security, protecting the integrity and confidentiality of data while facilitating secure data sharing across different dataspaces. It will also enable the seamless integration of data from multiple sources, allowing for comprehensive analysis and the generation of valuable insights. Furthermore, the project will foster collaboration among stakeholders in agriculture, processing industries, and data science, opening up new opportunities for sustainable business models. The scalability of the DS2 framework will ensure that data analysis capabilities can grow as the agricultural IoT framework expands. Additionally, the project will enhance data accessibility, enabling the discovery and access of datasets and services across multiple dataspaces.

2.3.3 Data sets availability in the precision agriculture use case

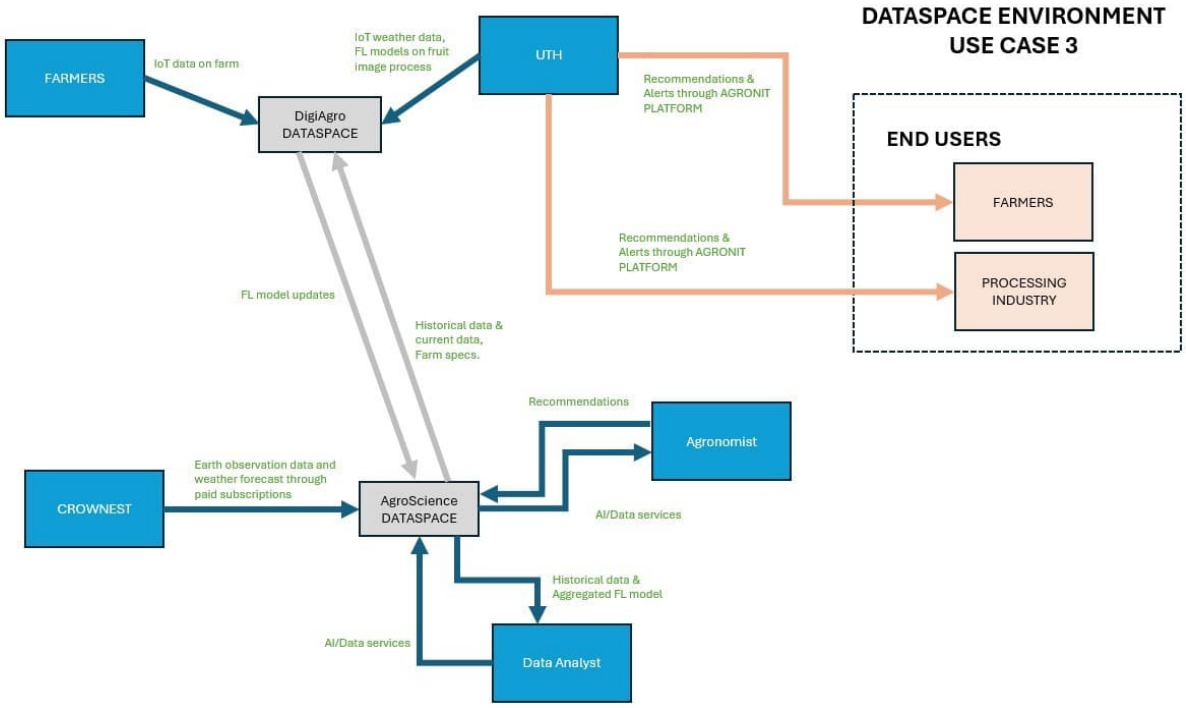


Figure 11. Data set availability in the precision agriculture use case

DATASET	OWNER OF THE DATASET	DATA TYPE AND DESCRIPTION
IoT data from farm	Farmers	Timeseries data coming directly from the farm, related to soil conditions.
IoT weather data	UTH	Timeseries data coming from the IoT weather stations infrastructure deployed by UTH that includes data related to temperature, humidity, wind speed, solar radiation, etc.
FL model updates	UTH	Local FL models' updates from training on image data provided by cameras placed on the field.
Earth observation data / Satellite imagery	CROWNEST	Satellite imagery provides a broad view of agricultural fields and can be used to monitor crop health, detect anomalies, assess vegetation indices, and track changes over time. Platforms like NASA's Landsat, ESA's Sentinel series, and commercial providers like Planet and DigitalGlobe offer high-resolution satellite imagery for agricultural applications. 1. Crop health monitoring systems like NDVI (Normalized Difference Vegetation Index) and EVI (Enhanced Vegetation Index) use satellite or aerial imagery to assess plant health, biomass production, and stress levels in crops. These indices can help farmers identify areas of concern and take corrective actions to improve crop productivity. 2. Earth observation data can be used in conjunction with machine learning algorithms to develop crop growth/yield prediction models. 3. Earth observation data can be used to monitor pest and disease outbreaks in crops by detecting changes in vegetation patterns, colour variations, and growth anomalies.
Weather forecast data	CROWNEST	48-hours local weather forecasts with a 3-hour interval are provided to the end-users (farmers, agriculturists) daily,

		including forecasts on rainfall, air temperature & humidity, wind speed and direction, as well as user-dependent recommendations on the suitability of execution of key practises (irrigation, spraying, fertilizing).
AI/Data Services	Data Analysts	AI models and optimization services for crop growth prediction, resource waste minimization, etc.
Recommendation Services	Agronomists	Personalized recommendations to farmers for efficient management and scheduling of agricultural tasks.
Farm Specifications	CROWNEST	Data about the registered crop fields, associated farmers and other.
IoT-sensed image data	UTH	Crop growth monitoring using micro-cameras.

Table 6: Description of datasets in Precision Agriculture use case

2.4 Use case 4 – X-Intersector

Whilst the first three use cases have been identified already in the project submission phase, this case was purposefully not so that it can be dynamically constructed based on the findings of the project. Each of the above cases are "Intersector" meaning that each one deals with multiple variant sectors pertaining to that case. This case is labelled "X-Intersector" where X means Cross – So "Cross-Intersector ". The ideal being that it takes elements from all cases to compose at project-time a new case. It will also place more emphasis on the technological outputs of WP3-6 since as these run in the project they will undoubted adapt and evolve due to the natural nature of such projects and thus the case will be adapted to, for example, focus on more challenging modules or modules which require more in depth validation.

2.4.1 Potential of using datasets from different use cases

The potential of the X-intersector in using datasets from different use cases, whether within the same sector or across sectors, is immense. Integrating these datasets can enhance insights, improve decision-making, and drive innovation across various domains. Below you will find an example scenario on how data from the use cases could be shared or combined, while exact implementation will be defined within the setup phase of the use cases in M18.

Scenario:
The product in question is a fruit-based product (e.g., a juice or jam) sold in Cluj-Napoca. The fruit is grown in Greece, processed and combined with other locally produced ingredients in Slovenia, and transported across Europe to be sold in Romania. By combining the datasets, a comprehensive carbon footprint can be calculated:

1) Agriculture and Production Footprint:

The carbon footprint from farming operations in Greece (based on IoT and precision agriculture data) is combined with the manufacturing and processing data from Slovenia. This includes the energy used in both farming and production, the transportation of raw materials within each country, and the environmental impact of production.

2) Transportation and Logistics Footprint:

The transportation footprint is calculated by linking the data from both Greek and Slovenian agriculture and production hubs to logistics data from transport across Europe. This includes GHG emissions from the different modes of transport (trucks, ships, etc.) and route optimization based on European mobility and traffic datasets.

3) Retail and End-Consumer Footprint:

Once the product reaches Cluj-Napoca, smart city data helps assess the emissions from the retail environment (storage, refrigeration) and from local mobility patterns during last-mile delivery. Additionally, public transportation or consumer travel to the store is factored into the product's total carbon footprint.

2.4.2 *Setting Up: Interoperability of use case across data spaces via IDT toolkit*

Use case 4 will validate the IDT toolkit from the perspective of a typical data space participants new adoption of DS2, their selection of modules. And where they share data with participants from several dataspace in different sectors. This use case will primarily demonstrate the technical functionality the toolkit and selected modules to enable and support the interoperability between different data spaces.

In use case 4, the setup and configuration of the toolkit components will be validated. This setup acts as a preliminary instruction for participants to configure their own inter-participant connections using DS2 components when relevant. The components may have dependencies, and such dependencies need to be resolved in the configuration phase. The method for these resolutions may be automatic or manual, depending on the case. DS2 targets a straightforward setup, pre-commercial grade, which primarily needs configuration rather than programming skills to deploy.

The planned use case scenario is as follows with (related participants in brackets):

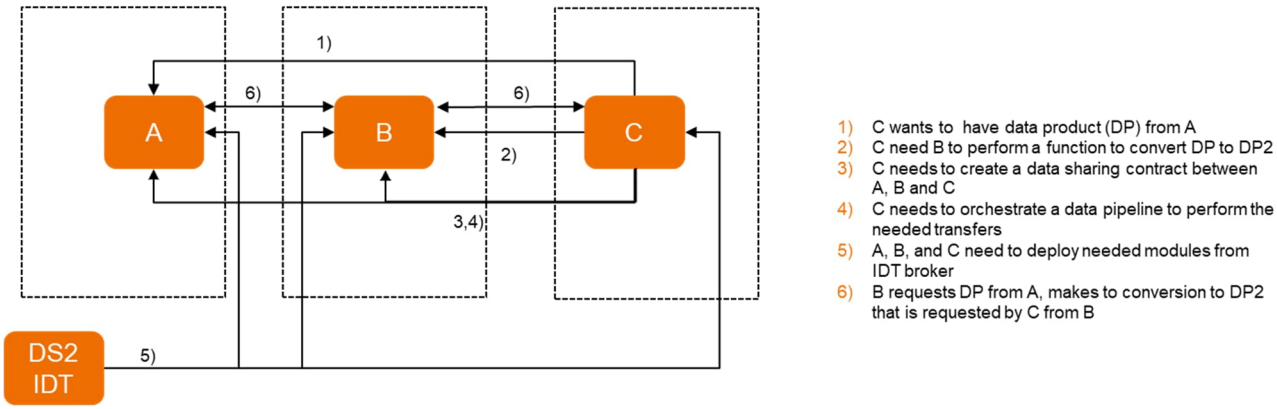


Figure 12. Use Case 4 illustration

Figure above illustrates the use case, the roles in the figure are:
C = data consumer
A = data provider
B = data space intermediary service

In this example there are two data spaces, three participants, one consumer (C) and one provider (A). B acts as a participant refining the data in the middle of the data pipeline. C initiates the data request from A, but needs B to make modifications to the data. Such modifications can be related to data model, aggregation, algorithmic, translations etc. C informs B that A will provide a data product that needs to be processed. C also initiates data sharing contract between each participant. C is in charge of the data pipeline, including the relevant DS2 modules in the end-to-end route, triggering DS2 module deployments in each of the participants, using e.g. DS2 IDT broker as a source for the modules. When the pipeline is deployed, B will request the data product from A and, after refinement, send it to C. The same process could occur with more participants or fewer participants. In minimum, there are two participants (here C and A), and there is no theoretical maximum, in principle there can be any number of participants in the data pipeline, as is needed by the data consumer.

3 OVERVIEW OF KPI'S PER USE CASES

The Key Performance Indicators (KPIs) of DS2 are categorized into two distinct sections to ensure comprehensive monitoring and evaluation.

The first section focuses on industry-specific KPIs, which are tailored to the unique needs and objectives of each individual use case. These KPIs vary significantly depending on the industry context and goals of the use case. Responsibility for defining, tracking, and analysing these KPIs lies primarily with the leader of the specific use case, who possesses the expertise and insight necessary to align these indicators with industry standards and expectations.

The second section pertains to technology-specific KPIs. These KPIs are designed to measure the performance and effectiveness of the technological solutions implemented within the use case. Monitoring these KPIs is a collaborative effort between the use case leader and the technical coordinator. Together, they ensure that the technology is performing as expected, and that it contributes effectively to the overall success of the use case. This dual focus ensures that both industry-specific and technological aspects are comprehensively evaluated, leading to a well-rounded assessment of performance.

3.1 City scape KPI's

Industry specific KPI's:

KPI definition	Baseline	Goal	Methodology / description of the KPI
Optimisation of energy consumption leading to reduction of emissions	100%	25% reduction	Estimate a 25% reduction of CO2 emissions at the observation points with the baseline specified at M13 of the project with data from a sample of 3 apartments in Manastur neighbourhood in Cluj-Napoca. The methodology will be the CO2 footprint taken at apartment level.
Increase the use and awareness regarding green, affordable solutions in the data providing residential and public building units	Percentage of awareness declared in a survey by Cluj Napoca residents	15% increase	Following the measuring of the level of awareness regarding green solutions, it is estimated that this will increase by 15%. Data taken from a survey applied at local residential units in Cluj-Napoca in months 13-15 of the project.
Development of a novel digital twin solution exploiting the different data sources	0	1	Currently there is no digital twin using multi-sector data, the integration of its data through the DS2 dataspace representing the KPI.
Contributing data into European open-source repositories	0	>= 200GB	Data used through the Cluj-Napoca Cityscape.

Table 7: Industry specific KPI's in City Scape use case

Technology specific KPI's:

KPI definition	Baseline	Goal	Methodology / description of the KPI
Availability of data from different sectors	0	6 different sectors	Currently, no relevant data is shared over data spaces. Aim at the end of the project is to use multi-sector data from 6 sectors in Romanian DS2 dataspaces. Expected sectors: mobility, weather/climate,

			economics, demographics, energy, real estate and public policy.
Human in the loop integration for DS2: Increase of data owners in the process and increased awareness on the usage of their data	Baseline as measured through a questionnaire in M13	>30% for citizens >50% for public sector	Following the baseline measurement of M13 of the project through a stakeholder survey in Cluj-Napoca, measure the data ownership and awareness level of how to use data for climate neutrality at the end of the project expecting a 30% increase in usage and/or awareness level for citizens in the sample and 50% level increase for representatives of the public sector.
Demonstration of technical building blocks of DS2 IDT	0	5 or more	5 building blocks will be used and demonstrated by the end of the project.
Decrease in the effort and complexity to make individual data connections with an external partners data source (where it is assumed other participants would also run DS2 functionality)	Discussion and agreement about access: 16 hours Bespoke implementation of services/technology: 16 hours Testing: 8 hours Total: 40 hours	>80% less effort compared to the current solution	This KPI measures the reduction in time, effort, and complexity involved in establishing data connections with external partners' data sources, leveraging DS2 functionalities. It will be measured by comparing pre- and post-implementation metrics, including time taken, user effort ratings, and procedural steps required. The goal is to achieve significant reductions in these metrics, making data integration more efficient and less resource-intensive for participants.

Table 8: Technology specific KPI's in City Scape use case

3.2 Green Deal KPI's

Industry specific KPI's:

KPI definition	Baseline	Goal	Methodology / description of the KPI
Accuracy of machine learning algorithms for calculation of GHG emissions and air pollutants	not existing currently	85%	Based on the collected data made available for further evaluation and learning of AI models via dataspace infrastructure use case will perform calculations of predicted values in areas that does not have a dedicated measuring device currently. This predicted value will then be compared with actual measured value and we expect to reach the accuracy as specified.
Square km of area, where GHG emissions and air pollutants will be measured with IoT sensors	2 stations operational, estimated area 25km2	10+ stations covering 100 km2	The current area is estimated based on 2 operational stations. Since MOMS plans to install additional stations, the area will increase according to this KPI.
Square km of area, where GHG emissions and air pollutants will be calculated using	0 – not existing currently	400 km2 – there was a typo in the proposal which indicated 100.	The area where values will be calculated corresponds to larger surroundings of Murska Sobota municipality and the area is achievable to be covered.

KPI definition	Baseline	Goal	Methodology / description of the KPI
machine learning algorithms			
Number of residents in the region and city of Murska Sobota covered and benefitted by the outcomes of the project	0 – not existing currently	50.000	The number of persons living in Pomurje region with 1336km2 is 114.000; target of 50.000 covers approximately the area of 400km2 as in other KPI above.
Contributing data into European open-source repositories	0 – not existing currently	>= 100GB	The quantity of data includes also some satellite images from Sentinel and other similar sources. This data can be contributed to European open-source repositories.

Table 9: Industry specific KPI's in Green Deal use case

Technology specific KPI's:

KPI definition	Baseline	Goal	Methodology / description of the KPI
Number of different complex data sources connected to Green Deal data spaces	4 – currently data from ARSO weather stations, ARSO AIR quality, IoT sensors and Benchmark study groups are connected.	>20 (IoT sensors, satellite, human)	Based on existing list of data sources to be included and as listed in table above.
Decrease in the effort and complexity to make individual data connections with an external partners data source (where it is assumed other participants would also run DS2 functionality)	Discussion and agreement about access: 20 hours Bespoke implementation of services / technology: 14 hours Testing: 12 hours Total: 46 hours	>80% less effort compared to the current solution	This KPI measures the reduction in time, effort, and complexity involved in establishing data connections with external partners' data sources, leveraging DS2 functionalities. It will be measured by comparing pre- and post-implementation metrics, including time taken, user effort ratings, and procedural steps required. The goal is to achieve significant reductions in these metrics, making data integration more efficient and less resource-intensive for participants.
Piloting a human-centric, complex and cross-sectoral data gathered through IDT broker in the city of Murska Sobota	Not existing currently	Successful piloting based on KPIs achieved plus interviewing pilot participants before and after the implementation of DS2	Interview/questionnaire with participants of the use case will be done before the start of pilot implementation in M13 in order to have the baseline. The piloting will be concluded by interview/questionnaire to measure if the DS2 implementation was successful.
Human in the loop integration for DS2: increase of data owners in the process and increased awareness on the usage of their data.	not existing currently	>20% for citizens, environment related policy actor	Questionnaire with stakeholders of the use case will be done before the start of pilot implementation in M13 in order to have the baseline. The piloting will be concluded by questionnaire to measure increased awareness on the usage of the data. Data

KPI definition	Baseline	Goal	Methodology / description of the KPI
			sharing platforms will be used to review the number of data owners before and after the implementation of DS2.

Table 10: Technology specific KPI's in Green Deal use case

3.3 Precision Agriculture KPI's

Industry specific KPI's:

KPI definition	Baseline	Goal	Methodology / description of the KPI
Increase in the accuracy of predictions via faster training from FL using more recent data leading to more frequently updated models and greater model accuracy.	TBD	>20% increase in the accuracy of predictions	This KPI measures the improvement in prediction accuracy by utilizing federated learning (FL) techniques and the availability of data from other dataspaces. It will be assessed by comparing baseline and post-implementation accuracy metrics. The goal is to achieve significant accuracy improvements and more timely model updates through the efficient use of FL and diversity of datasets.
Increase in estimation of Leaf Area Index (LAI) through the application of DS2 modules.	TBD	> 10% increase	This KPI measures the improvement in prediction of LAI through the data fusion from satellite images and soil moisture sensors deployed on the fields, which are data sources from different dataspaces.
Minimization of resource waste. Resources (e.g. water, fertilizers) shall be used with precision according to the collected measurements	TBD	> 30% savings in resources used	This KPI measures the minimization of resource waste by ensuring resources such as water and fertilizers are used with precision based on collected measurements. The improvement of this KPI will follow the improvement of accuracy of the agronomists' recommendations to farmers that will be achieved by the greater availability of data as well as of AI and optimization models that is achieved through the federation of DigiAgro and AgroScience dataspaces.

Table 11: Industry specific KPI's in Precision Agriculture use case

Technology specific KPI's:

KPI definition	Baseline	Goal	Methodology / description of the KPI
Decrease in the effort and complexity to make individual data connections with an external partners data source (where it is assumed other participants would also run DS2 functionality)	Discussion and agreement about access: 16 hours Bespoke implementation of services/technology: 16 hours Testing: 8 hours Total: 40 hours	>80% less effort compared to the current solution	This KPI measures the reduction in time, effort, and complexity involved in establishing data connections with external partners' data sources, leveraging DS2 functionalities. It will be measured by comparing pre- and post-implementation metrics, including time taken, user effort ratings, and procedural steps required. The goal is to achieve significant reductions in these metrics, making data integration more efficient and less resource-intensive for participants.
Availability of data sources in AgroScience DS with	not existing currently	Data from at least 10	This KPI measures the availability and accessibility of data collected from various fields within the involved dataspaces. It will be assessed by tracking

measurements collected from different fields		fields are available	the number of data sources available, the frequency of data collection, and the percentage of time these sources are accessible to users. The goal is to ensure consistent and reliable data availability, aiming for high accessibility rates and comprehensive coverage of different fields.
Human in the loop integration	not existing currently	> 30% involvement of data owners in the process and increased awareness on the usage of their data.	Questionnaire with stakeholders of the use case will be done before the start of pilot implementation in M13 in order to have the baseline. The piloting will be concluded by questionnaire to measure increased awareness on the usage of the data. Data sharing platforms will be used to review the number of data owners before and after the implementation of DS2.

Table 12: Technology specific KPI's in Precision Agriculture use case

4 USER REQUIREMENTS

This chapter provides a comprehensive overview of the user requirements for each use case. Each use case is presented with scenarios that involves specific groups of stakeholders and potential end users. These scenarios outline the current state of operations, highlighting existing challenges or limitations faced by the stakeholders. For each scenario, a detailed description is provided, illustrating what is currently taking place, including the processes, interactions, and tools being used. This helps to set a baseline, allowing for a clearer understanding of the gaps or inefficiencies that need to be resolved. Following this, the chapter identifies the key requirements—whether they be related to technology, data integration, or stakeholder collaboration—that must be fulfilled to enable the successful and efficient execution of each use case by the end of the project. The selection of different DS2 modules is provided within the requirements section for a better and clearer connection to the system architecture that is provided in part B of this document.

4.1 City Scape use case requirements

Identifier	Scenario name	Group of stakeholder / users involved	*What is currently*	Requirements	Any other relevant information, comments.
UC1.1	Residential building	• Residents, • Administration • Energy Company, (producer, managing) • Municipality • Businesses (Renergia) • Environmental NGOs • EC • Building Companies (Developers) • NZEB Certifiers	• As a resident in an old building, I am willing to share the data on energy consumption and other patterns because I want to save money (to have information about the best energy saving plan and access policies that reduce my taxes). • As a property developer I would be interested to get data on awareness and value put on energy effectiveness to create better products / fit the needs. • As an energy provider I want to receive consumption information to optimize the production and create more attractive packages.	1. Identification of money-saving options and simultaneous reduction of environmental impact. 2. Optimize behaviour so that we can achieve climate neutrality in the residential sector 3. Provision of data for business modelling on green business solutions 4. Preventing the misuse of data should be enforces by adding a trust component to the data sharing process by using the SDS module. 5. Residents are willing to share the data on energy consumption and other patterns. The PAE module will ensure that the share date is anonymized before the consumer receives it. In the case of data that is not anonymized we want to trigger a notification and block the share by using the DINS and DHARE modules. 6. Energy providers want to receive consumption information to optimize the production and create more attractive	• Confidentiality of the data, privacy of personal data (energy consumption). • Risk: policies are not flexible.

Identifier	Scenario name	Group of stakeholder / users involved	*What is currently*	Requirements	Any other relevant information, comments.
			• As an environmental activist, I am interested to see how the combination of different actions can impact on climate neutrality and the impact of different actions individually on climate neutrality. • • As a start-up (RENERGIA) I am interested in mixing data on energy to inform cases for Net Zero Certification and developers' feasibility study.	packages. The ORC module should be used to pull energy consumption and production data from multiple dataspaces. 7. Property developers and environmental activists would be interested in getting data on climate neutrality and energy effectiveness for buildings. We want to track the data usage by using the DRM module to prevent misuses of data. 8. RENERGIA, startup in the energy sector, wants to easily search and extract mixed energy data in order to validate 'Nearly-zero energy and zero-emission building' certificates. The RET module would facilitate access to the information.	
UC1.2	Traffic optimisation	• Municipality (CTP – Public Transport Company) • Citizens transiting the city/ living/ working in the city	As a Municipality committed to be one of the 100 Cities in the Net Zero mission, we are interested to understand how traffic data impacts emissions and how it can be optimised. • As citizens wasting significant time in commuting, we are interested to access data that can show optimisation of routes to decrease commuting time. • Currently, there is a lack of data on options for the reduction of traffic and also the environmental impact of traffic control	1. By using the ORC, E2C and DINS modules to we could automate data sharing and corelation for the municipality and researchers who could analyse traffic patterns and climate impact. 2. Need for data sharing to be logged using the DRM module. 3. Trust between parties should be enforced using the SDS module.	The following calculators would be needed to link traffic data to impact: • IMPACT - Air quality • IMPACT - Quality of living • IMPACT Policy optimisation
UC1.3	Heat islands	Municipality	As a Municipality committed to be one of the 100 Cities in the Net Zero mission we are interested to identify and monitor the Heat Islands in the city.	1. Share of data for the identification of "hot" areas in the city the association of those areas with climate neutrality solutions and initiatives. 2. Correlation could be done by using the ORC and RET modules to automate data extraction across multiple data sources and/or dataspaces. RET could be used to generate the sharing endpoints from the data offerors, by using human readable queries.	
UC1.4	Food waste	Municipality, Citizens, NGOs (Food Waste Combat)	As a Municipality we want to find better options for recording Food Loss and Waste (FLW) data and reporting it in a more accurate manner including the impact on the environment.	1. Accurate data is not available. Sharing and gathering data from multiple sources and sectors can lead to state-of-the-art entrepreneurial solutions. Using the catalogue module should help data searching across dataspaces. 2. Data will be analysed to ensure quality criteria are met using PAE and DINS.	

Identifier	Scenario name	Group of stakeholder / users involved	*What is currently*	Requirements	Any other relevant information, comments.
			As civic actors we are interested in accessing FLW data to encourage entrepreneurship opportunities in this area.	3. Data will be retrieved from the data source using automatically generated REST requests using RAT.	

Table 13: City Scape use case requirements

4.2 Green Deal use case requirements

Identifier	Scenario name	Group of stakeholder / users involved	What is currently	Requirements	Any other relevant information, comments.
UC2.1	Analysing and evaluating effects of traffic on air quality.	Data scientists, Local authorities (municipality...), ITC, UNI MARIBOR	1. Road traffic measurement is performed in few locations with smart traffic signs and the data is shared and accessible via DIH AGRIFOOD DATA SPACE. 2. Data about the number of cars entering and exiting the city from highway is shared and accessible via DIH AGRIFOOD DATA SPACE. 3. Quality of air and weather data is measured on few locations in the city and to be extended via the project, while the data is available through MOMS DATASPACE.	1. Relevant data sources to be obtained from both data spaces within the use case. Assessment of data quality from IoT sensors will be performed with module E2C 2. Evaluating this data in context of air quality and find patterns impacting air quality. 3. Further relevant data sources to be identified and included, if needed. 4. For discovering and assessing the available datasets in the DS2 ecosystem we will use chatbot module – DARC. 5. For offering Green Deal datasets and searching for other datasets in the DS2 ecosystem, the Data Marketplace module will be used - DMK. 6. To compose a REST call to get data, based on Open API specifications, we will use module RET. 7. Data scientists from UNI MARIBOR to develop models correlating traffic and air quality. 8. MOMS will develop visualisation of data for non-technical users in order to take actions related to traffic regulation and parking. 9. A support tool for assessing risk for sharing data and provide incentives to data owners should be used. This should prevent large risks if the participants share data among themselves. Module SDS will be used. 10. For offering Green Deal datasets and searching for other datasets in the DS2 ecosystem the Data Marketplace module will be used - DMK. 11. Managing data flows from data owners to data users, where different workflows, which the tool will provide will be used for data orchestration to compose services. (ORC) 12. Requirement to identify if transformation is needed on obtained datasets and perform preconfigured transformation, if needed. Also to	Historical data can be used; refresh time is not important, but the data quality has to be assured with respect of time and location accuracy. Data anonymisation should happen at the source and is out of scope for this scenario.

Identifier	Scenario name	Group of stakeholder / users involved	What is currently	Requirements	Any other relevant information, comments.
				extract structured data from unstructured documents, e.g.: extract data about the traffic from the report in the PDF format, we will use module DDT and CUR 13. Accurately interpret and transform data from different data sources. E.g.: different organizations use different data models, for describing same properties and concepts. 14. Real-time data analysis and detecting predefined thresholds in collected data must be used with help of DINS module. 15. Monitor transfer of the data potentially including alert based on specified criteria. DSHARE module to be used.	
UC2.2	Analysing and evaluating effects of industry and agriculture on air quality.	Data scientists, Local authorities (municipality...), ITC UNI MARIBOR, Farmers	1. Industry and agricultural production locations are identified and mapped. 2. Relevant data sources on agricultural productivity are shared via DIH AGRIFOOD DATA SPACE. 3. Relevant data about energy consumption are shared via the MOMS DATA SPACE. 4. Quality of air and weather data is measured on few locations in the city and to be extended via the project, while the data is available through MOMS DATASPACE.	1. Relevant data sources to be obtained from both data spaces within the use case. Assessment of data quality from IoT sensors will be performed with module E2C. 2. For discovering and assessing the available datasets in the DS2 ecosystem we will use chatbot module – DARC. 3. For offering Green Deal datasets and searching for other datasets in the DS2 ecosystem the Data Marketplace module will be used - DMK. 4. To compose a REST call to get data, based on Open API specifications, we will use module RET. 5. Evaluating this data in context of air quality and find patterns impacting air quality. 6. Further relevant data sources to be identified and included. 7. Data scientists from UNI MARIBOR to develop models correlating traffic and air quality. 8. MOMS to develop forms of data visualisation for non-technical users to take actions related to industry and agriculture and their influence on air quality in the city of Murska Sobota. 9. A support tool for assessing risk for sharing data and provide incentives to data owners should be used. This should prevent large risks if the participants share data among themselves. Module SDS will be used. 10. Immutable (permission blockchain (Hyperledger fabric) based) transactional log of access to the data should be used. Because the data is from the industry stakeholders, it is very important that we have an immutable log of who and when has accessed their data. In the same way we will also record data transformation logs, data about data policies. (DRM) 11. Service for describing and defining datasets and their REST endpoints, which will be shared by the use case. The description will also be used for metadata description of the connectors. The analysis	Historical data can be used; refresh time is not important, but the data quality must be assured with respect of time and location accuracy. Data anonymisation should happen at the source and is out of scope for this scenario.

Identifier	Scenario name	Group of stakeholder / users involved	What is currently	Requirements	Any other relevant information, comments.
				of the provided information, will also be put in the catalogue module. (CAT) 12. Managing data flows from data owners to data users, where different workflows, which the tool will provide will be used for data orchestration to compose services. (ORC) 13. Requirement to identify if transformation is needed on obtained datasets and perform preconfigured transformation, if needed. Also to extract structured data from unstructured documents, e.g.: extract data about the traffic from the report in the PDF format, we will use module DDT and CUR. 14. Accurately interpret and transform data from different data sources. E.g.: different organisations use different data models, for describing same properties and concepts. 15. Real-time data analysis and detecting predefined thresholds in collected data must be used with help of DINS module. 16. Monitor transfer of the data potentially including alert based on specified criteria. DSHARE module to be used. 17. Searching and assessing of available data and DS2 tools, modules and policy in the ecosystem, using AI algorithms for chatbot like services but and at the end also configures the modules for user (DARC).	
UC2.3	Analysing and evaluating effects of households on air quality	Data scientists, Local authorities (municipality...), ITC, UNI MARIBOR	1. Household heating types to be identified and mapped. 2. Data about households (including flats) to be arranged including number of participants, building area and other relevant data. Dataset shared through the MOMS DATA SPACE. 3. Quality of air and weather data is measured on few locations in the city and to be extended via the project, while the data will be available through MOMS DATASPACE.	1. Evaluating this data in context of air quality and find patterns impacting air quality. Assessment of data quality from IoT sensors will be performed with module E2C 2. Further relevant data sources to be identified and included. 3. Data scientists from UNI MARIBOR to develop models correlating traffic and air quality. 4. Development of data visualisation forms for non-technical users to take actions related to effects of households on air quality in the city. 5. A support tool for assessing risk for sharing data and provide incentives to data owners should be used. This should prevent large risks if the participants share data among themselves. (SDS) 5. Immutable (permission blockchain (Hyperledger fabric) based) transactional log of access to the data should be used. Because the data is from the industry stakeholders, it is very important that we have an immutable log of who and when has accessed their data. In the same way we will also record data transformation logs, data about data policies. (DRM) 6. Service for describing and defining datasets and their REST endpoints, which will be shared by the use case. The description will also be used	Historical data can be used; refresh time is not important, but the data quality has to be assured with respect of time and location accuracy. Data anonymisation should happen at the source and is out of scope for this scenario.

Identifier	Scenario name	Group of stakeholder / users involved	What is currently	Requirements	Any other relevant information, comments.
				for metadata description of the connectors. The analysis will of the provided information, will also be put in the catalogue module. (CAT) 7. The tool that facilitates the management of data flows from data owners to data users by providing different workflows specifically designed for data orchestration should be used. (ORC) 8. Requirement to identify if transformation is needed on obtained datasets and perform preconfigured transformation, if needed. Also to extract structured data from unstructured documents, e.g.: extract data about the traffic from the report in the PDF format, we will use module DDT and CUR. 9. Accurately interpret and transform data from different data sources. E.g.: different organizations use different data models, for describing same properties and concepts. 10. Real-time data analysis and detecting predefined thresholds in collected data must be used with help of DINS module. 11. Monitor transfer of the data potentially including alert based on specified criteria. (DSHARE) 12. Searching and assessing of available data and DS2 tools, modules and policy in the ecosystem, using AI algorithms for chatbot like services but and at the end also configures the modules for user (DARC).	
UC2.4	Awareness raising of citizens and stakeholders in the Municipality.	Citizens, public institutions and all general public, MOMS, ITC, UNI MARIBOR	1. MOMS is using the data from two sensors in the city. This data is shown on Municipality website, where citizens and other stakeholders can observe current level of pollution in the city.	1. Data scientists from UNI MARIBOR to develop models correlating effects of scenarios 1-3 on air quality in the city. Assessment of data quality from IoT sensors will be performed with module E2C 2. A learned model based on scenarios 1-3 to be used for detecting situations that are leading to bad air quality. MOMS will display the data on relevant webpages and other existing notification systems in use. 3. A support tool for assessing risk for sharing data and provide incentives to data owners should be used. This should prevent large risks if the participants share data among themselves. (SDS) 4. Real-time data analysis and detecting predefined thresholds in collected data must be used with help of DINS module. 5. Searching and assessing of available data and DS2 tools, modules and policy in the ecosystem, using AI algorithms for chatbot like services but and at the end also configures the modules for user (DARC).	Historical data can be used; refresh time is not important, but the data quality must be assured with respect of time and location accuracy. Data anonymisation should happen at the source and is out of scope for this scenario.
UC2.5	Air quality notification system.	Citizens, public institutions and all general public, MOMS, ITC, UNI MARIBOR	1. Municipality and meteorological agency are providing alerts on their webpages based on existing	1. A learned model based on scenarios 1-3 to be used for detecting situations that are leading to bad air quality. The output of the system will be displayed on relevant webpages and other existing notification systems in use.	Live or near-real time data to be used; refresh time is important; the data quality must be assured with respect of time and location

Identifier	Scenario name	Group of stakeholder / users involved	What is currently	Requirements	Any other relevant information, comments.
			data from air quality measurements.	2. A support tool for assessing risk for sharing data and provide incentives to data owners should be used. This should prevent large risks if the participants share data among themselves. (SDS) 3.Real-time data analysis and detecting predefined thresholds in collected data must be used with help of DINS module. 4.Searching and assessing of available data and DS2 tools, modules and policy in the ecosystem, using AI algorithms for chatbot like services but and at the end also configures the modules for user (DARC)	accuracy. Data anonymisation should happen at the source and is out of scope for this scenario.

Table 14: Green Deal use case requirements

4.3 Precision agriculture use case requirements

Identifier	Scenario name	Group of stakeholder / users involved	What is currently	Requirements	Any other relevant information, comments.
UC3.1	Fruit procurement – sourcing	Farmers, Processing industry, UTH, CROWNEST, Data Analysts	1. A processing industry uses information derived from different fields (i.e., field-based crop images) to: a. Evaluate fruit quality. b. Assess growth variability across the operations area. c. Determine the optimal fruit sourcing process / logistics to implement. 2. The goal is to secure their processing capacity and throughput. 3. This data is made available to the processing industry through AgroScience DS, which includes: a. Machine learning algorithms developed by data scientists. b. Optimization models contributed by CROWNEST researchers.	1. Retrieve data from DigiAgro DS where sensor data from crop owners is collected and stored using E2C module 2. Ensure that DigiAgro DS provides comprehensive and high-quality data from farm sensors and weather sensors. Usage of DINS module is planned. 3. Use AgroScience DS for data processing services and machine learning algorithms to analyse the collected data. 4. Ensure compatibility between DigiAgro DS and AgroSciences DS data formats and structures.	–DigiAgro DS: Farm sensor data, Weather sensor data –AgroScience DS: Farm specs data, Data services (data processing services, machine learning algorithms)
UC3.2	Extreme weather phenomena forecasting and crop management.	Agronomist, UTH, CROWNEST, Farmers, Data analyst	1. An agricultural consultant (agronomist) reviews the following data originating from DigiAgro DS and AgroScience DS respectively: a. Sensed environmental data. b. Local weather forecast data provided by the national meteorological agency. 2. The consultant then uses the data processing services originating from AgroScience DS to:	1. Ensure sensed environmental data and weather forecast data are available and readily accessible using E2C module. 2. Integrate environmental data and weather forecast data seamlessly between the two dataspaces for comprehensive review by the consultant. 3. Utilize AgroScience DS for data processing services to analyse and evaluate the intensity of	–DigiAgro DS: Farm sensor data, Weather sensor data –AgroScience DS: Weather forecast, Farm specs data, Data services (data processing services, machine learning algorithms)

Identifier	Scenario name	Group of stakeholder / users involved	What is currently	Requirements	Any other relevant information, comments.
			a. Evaluate and determine the intensity of a weather phenomenon (e.g. frost). b. Generate recommendations on crop management to mitigate the effect of the weather phenomenon on crop growth. 3. The end user (crop owner) will then consume related information in their native language to apply the recommended crop practices.	weather phenomena. DMK and DRM modules will be used to support this process. 4. Provide the recommendations in the native language of the end user (crop owner) for ease of understanding and implementation.	
UC3.3	Crop productivity	Farmer, Agronomist, UTH, CROWNEST, Data analyst	1. AgroScience DS offers third parties a pre-trained federated learning (FL) model to identify, isolate, and classify visible fruits. 2. AgroScience DS offers additional machine learning models to estimate fruit yield at the tree level in kilograms (kg). 3. AgroScience DS has deployed a FL aggregator model in the dataspace, which collects and combines trained models from FL agents running on field-based camera nodes at the edge originating from DigiAgro DS. 4. The end-user (farmer) is interested in quantifying a tree's fruit load by using the services offered and aims to receive the results in a direct and comprehensive manner.	1. Develop and offer a pre-trained federated learning (FL) model in AgroScience DS to identify, isolate, and classify visible fruits in the images. 2. Deploy a federated learning aggregator in AgroScience DS to collect and combine trained models from federated learning agents. 3. Deploy additional machine learning models in AgroScience DS to estimate the fruit yield at the tree level in kilograms (kg). 4. Ensure seamless integration of data from DigiAgro DS with the models in AgroScience DS.	–DigiAgro DS: Farm sensor data, Weather sensor data –AgroScience DS: Weather forecast, Farm specs data, Data services (data processing services, machine learning algorithms)

Table 15: Precision Agriculture use case requirements

5 USE CASES DS2 MODULES MAPPING

The table below is providing the link between the requirements within use cases and modules identified by the technological partners that will be developed as part of the DS2 project. This chapter addresses only the modules that are optional and to be explored by the use case. Modules that are essential to the DS2 project and represent the foundation, environment, and system modules are not shown in this table since they will be explored and tested by all use cases. The detailed description of the individual modules can be found in part B of this document.

MoSCoW prioritization methodology is used to prioritize requirements per module and use case scenarios. The module will be tested only in scenario marked with full word (e.g. M, S, C), those marked in brackets (e.g. [M], [S]) are to be tested only if the resources will allow it.
The acronym "MoSCoW" stands for "must-have," "should-have," "could-have," and "won't-have (this time)."

- M = Must-have: These items are essential for the success of the use case in DS2 project. There can be no compromise on whether they are included because, without them, the use case will not be possible to implement. In short, this is a top-priority MoSCoW requirement.
- S = Should-have: These items are those that are important but not essential like those in the "must-have" category. Elements in this category are considered a secondary priority; that is, they are important, but not crucial to success in the use case.
- C = Could-have: These items would be nice to have but are not essential. Still less important than the two preceding categories, these elements are considered a third-level priority in the DS2 framework. If including them will have negative consequences on cost or meeting deadlines, they should be omitted. It is only when they don't negatively affect other project elements that they should be included.
- W = Won't-have (this time): These items are those that are not essential and can be excluded from the project without jeopardizing its success. Being the lowest priority category, omitting them won't hurt the project and they would be included if project conditions became more favourable.

Name of the modules	DS2 Tier	DS2 Nature	UC1.1	UC1.2	UC1.3	UC1.4	UC2.1	UC2.2	UC2.3	UC2.4	UC2.5	UC3.1	UC3.2	UC3.3
DS2 Sovereignty Decision Support Tool (SDS)	T3	Optional	S	[S]	[S]	[S]	M	[M]	[M]	[M]	[M]	C	C	C
DS2 DRM Module (DRM)	T3	Optional	S	[S]	C	C	C	M	C	C	C	[S]	S	[S]
DS2 Catalog Module (CAT)	T3	Optional	C	C	C	S	M	[M]	[M]	C	C	[M]	M	W
DS2 Orchestration Module (ORC)	T2, T3	Optional	M	[M]	[M]	[M]	M	[M]	[M]	C	C	[S]	[S]	[S]
DS2 Data Retrival Module (RET)	T2	Optional	S	C	C	C	W	W	W	W	W	W	W	W

Name of the modules	DS2 Tier	DS2 Nature	UC1.1	UC1.2	UC1.3	UC1.4	UC2.1	UC2.2	UC2.3	UC2.4	UC2.5	UC3.1	UC3.2	UC3.3
DS2 Data Detection and Transformation Module (DDT)	T2	Optional	C	C	C	C	S	[S]	[S]	C	C	S	S	W
DS2 Model Development Toolkit Module (MDT)	T2	Optional	C	C	C	C	W	W	W	W	W	W	[S]	W
DS2 Curation Module (CUR)	T2	Optional	C	C	C	C	S	[S]	[S]	C	C	C	C	C
DS2 Catalog Module (CAT)	T2	Optional	M	[M]	[M]	[M]	M	[M]	[M]	C	C	M	[M]	[M]
DS2 Data Marketplace Module (DMK)	T2	Optional	C	C	C	C	M	[M]	[M]	C	C	C	M	[M]
DS2 Edge to Cloud Module (E2C)	T2	Optional	[M]	[M]	C	M	[M]	[M]	[M]	[M]	M	[M]	M	W
DS2 MultiCloud module (MCL)	T2	Optional	C	C	C	C	C	C	C	C	C	C	C	C
DS2 Data Inspection Module (DINS)	T2	Optional	[S]	[S]	C	S	[M]	[M]	[M]	[M]	M	M	[M]	C
DS2 Data Share Controller (DSHARE)	T2	Optional	[S]	C	S	[S]	M	[M]	[M]	C	C	C	C	C
DS2 DARC Module (DARC)	T0	Optional	C	C	C	C	[M]	[M]	[M]	[M]	M	C	C	C

Table 16: Use cases DS2 optional modules mapping

6 TECHNICAL PERSPECTIVE AND PRINCIPLES

6.1 Problem Space Revisit

The DS2 D2.1 Problem Space Definition document presented an overview of data spaces and within the technical perspective Section (Section 5) high-level information was provided on partner knowledge and Background to be used, and potentially relevant sister projects. Within the Problem Space Definition Section (Section 6) an initial breakdown of Modules/features into tiers was also covered. Part of this also included initial consideration on the license conditions re Background and updates to this can be seen in the Module attachments to this document although IPR will generally be specified in the exploitation deliverable of T1.4.

Following on from this, the DS2 technical perspective related to the RTD activities of WP4-6 and to a certain extent WP3 (which largely deals with Policies), remains grounded on the DOA high-level architecture picture whose core is built around an Intersector Dataspace Toolkit (IDT) enabling the sharing of data from providers and consumers in disparate sectors and dataspaces through a set of pick-and-mix Modules. In short, the DS2 technical objectives and implementation will:

- Allow Data Sharing between dataspaces.
- Allow the installation of dataspace enabling Modules (DS2 Modules) to enable or enhance their functionality in dataspaces and whose Modules may interact across dataspaces and/or only in one dataspace.

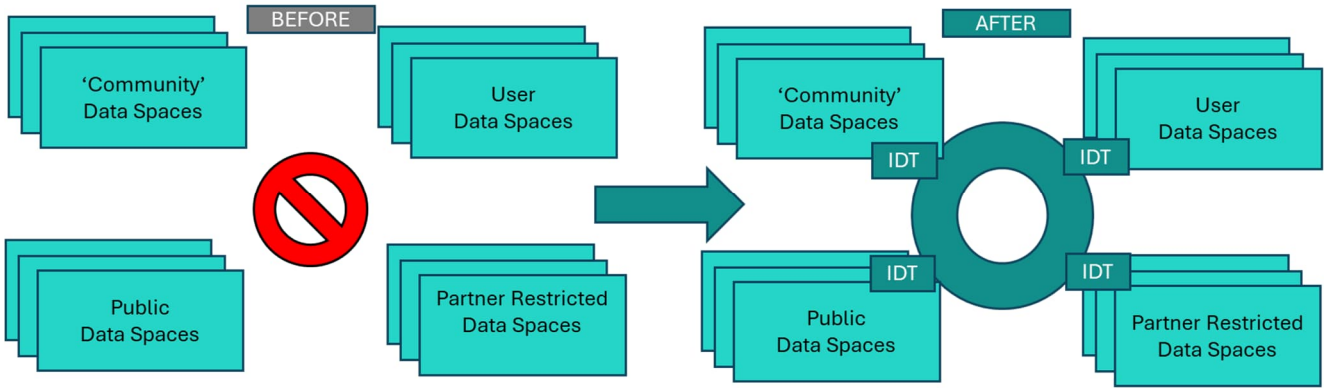


Figure 13. DS2 Before and After

The DS2 Modules (Dataspace Participants Apps) will be deployed through the mentioned IDT which will enable both these objectives. Each Module will be containerised according to the approach, methodology and parameters dictated by WP6.3, The RTD parties then place these on a central web portal where dataspace participants can register and select the Modules which fit their business needs through the portal marketplace. They will then be packaged in the IDT toolkit along with several default Modules primarily related to Data provider/consumer trust. Finally, IDTs are then deployed at the dataspace participants and their functionalities configured and invoked. They can add further existing and future Modules, as necessary. The technical approach of DS2 is open and generic such that this marketplace can be open to new Modules built outside or beyond the project itself. Within DS2 the Modules will have a common visual identity – ie a common look and feel.

Partners will make their developments on an agreed set of technical principles and look to make those development based on existing initiatives (primarily IDSA in the first instance) as well as their own knowledge and code background base. In reality, many of the Modules themselves will be based on existing background and/or open-source adapted to the needs of DS2 technologies and processes. Thus, DS2 can be considered very application and innovation orientated and supported by strong applied research basis.

The Problem Space document suggested a series of technical feature necessities for data sharing as per the diagram below, and through this document the project specifies how this will be achieved and where DS2 Modules fit.

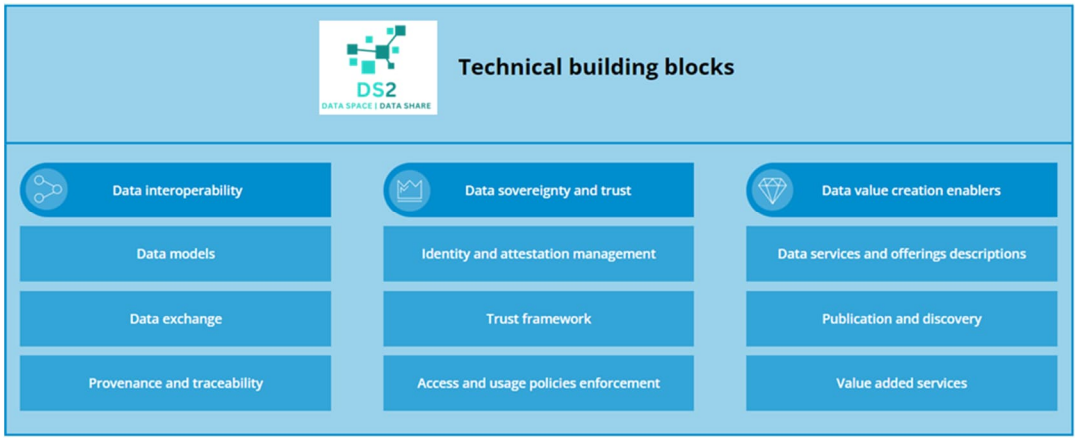


Figure 14. Technical Building Blocks

6.2 Technical Principles

DS2 will adopt existing approaches where possible and build incrementally; the default will be that DS2 adopts where possible the IDSA paradigm and if lacking or absent then will investigate GAIA-X, then other reasonable initiatives such as DSSC, and then perhaps only then build from scratch. DS2 will make, where relevant, technical contributions back to these initiatives as well as to standardisation through a CEN/Workshop agreement (CWA) or technical discussion in clustering events.

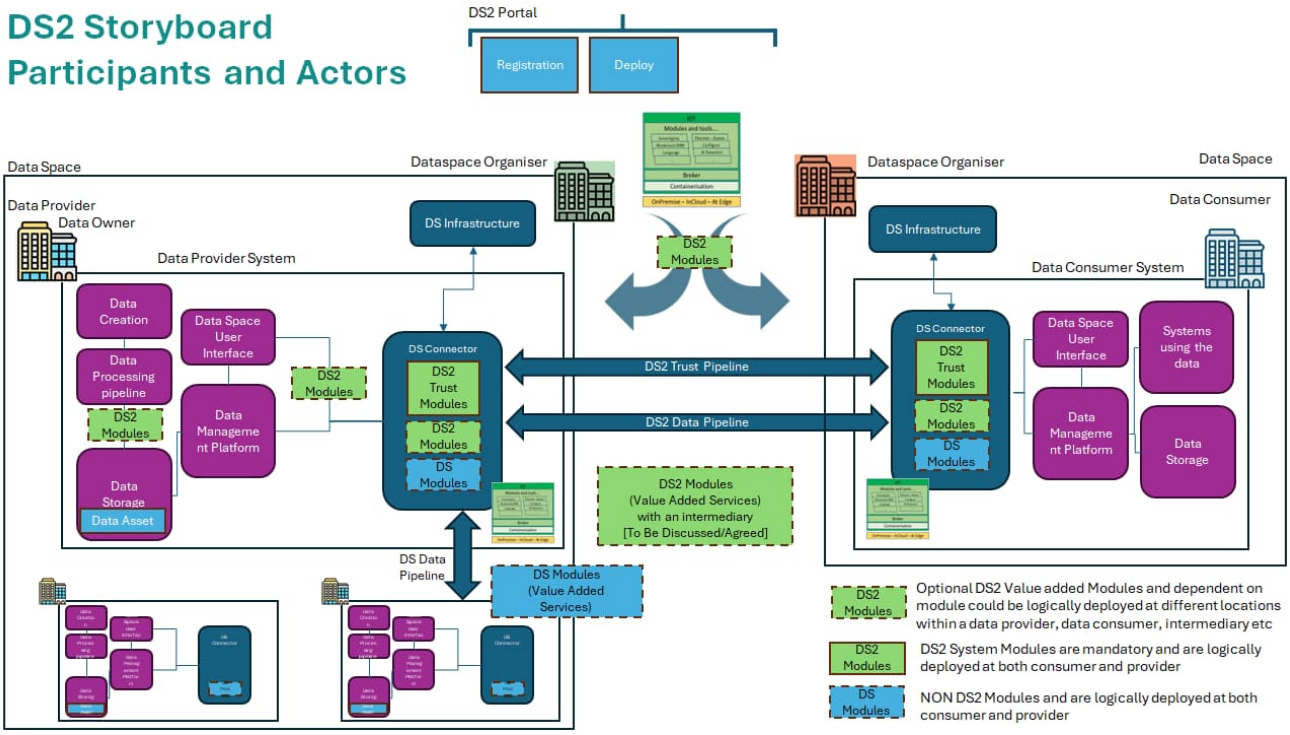


Figure 15. DS2 and Dataspace Participants and Actors

Principles adopted:

- What is meant by a DS2 Module?:
  - These are the self-standing software 'executables' which provide value added functionality to support data sharing both within dataspaces and between dataspaces (BDS). For example, Data Transformation, Data Sovereignty control.
  - In theory, the DS2 architecture can provide for an unlimited universe of Modules, but only certain (non-overlapping) Modules are chosen to be developed within the project.
  - The scope, genericity, breath, and duplications of Modules is not restricted by the DS2 architecture which should be open to allow any Module to be added, including post project Modules.
  - DS2 Module functionality is governed by Module owner as long as it complies with the IDT principles
  - It is the intent that all Modules are deployed via the IDT when applicable.
  - DS2 Module providers are not developing Modules specifically for use cases and must be dataspace agnostic in nature albeit that all DS2 use case needs should, where possible, be served.
  - The choice and configuration of Modules is determined by the owners of the dataspace and/or its use cases although there may be a minimum set of 'default' mandatory Modules.
  - Modules need to conform to the DS2 API model and common APIs (eg from IDT or the Security Module(s)) and other common requirements such as its Visual Identity, Dashbutton use etc. Where there are common needs, Modules should use the same API (eg View DRM Entry) and where possible, Modules of similar functionality should support common APIs to allow portability and non-lock-in.
  - Each Module needs to conform to the architecture deliverable, and any exception implemented only when presented and 'approved' by the DS2 Technical Manager and/or EXEC.
  - As long as it is conformant with the DS2 architecture/broker/IDT framework (APIs, deployment paradigm, containerisation etc), each Module can be independent of the development technology used.
  - There should be no strong dependencies between Modules and each should be useful and operatable standalone both within a DS2 context and for potential exploitation outside of the DS2 framework. If there is value-add of different Module interaction's, this is also possible and welcomed.
- All activity (methodology, architecture, Module functionality etc) must enable and/or contribute to inter-dataspace sharing and not be solely serving a single dataspace. This does not however mean that a Module needs to always work across dataspace but simply that it needs to contribute to it. Such activity (methodology, architecture, Module functionality etc) may also enable and/or contribute to in- dataspace sharing as well, and where this is possible this has benefits, but it is not at the jeopardy of achieving inter-dataspace sharing.
- As a basic principle of DS2, there is no central DS2 coordination point where all data and Modules are supposed to run and interact – i.e. there is not one central dataspace platform. However, in practice DS2 has determined supporting functions may be necessary through a "DS2 Portal" (eg registration to download an IDT, IDT deployment, a contact point for support, the ability to find dataspace and data...). Individual Module functionality is not part of the portal although all DS2 Modules are only downloaded through it to be deployed by a dataspace participant. Individual Modules may need to call on "DS2 Intermediaries" which provide the necessary centralized function only for that Module. For example, if a Transformation Module could provide more intelligent or faster transformation through the use of a service from an Intermediary, it is possible, but those services are in the ownership/domain of the Module owner and not the central platform.

- All Modules must be part of the IDT and be suitable for Dataspace participant Deployment. In the context of between data space sharing, the Modules could be deployed at external service providers, at each dataspace (by Consumer, Provider, or dataspace Provider) or in both depending on the exact functionality. However, the deployment paradigms are prechosen by developers and the selection of the Modules will be based on the choice/needs of the dataspace participants where they are deployed.
- DS2 does not support multiple participant combined scenarios out-of-the box

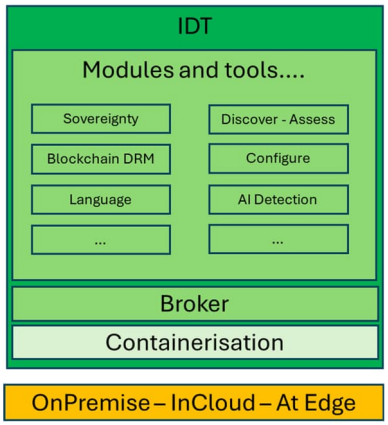


Figure 16. Modules as part of IDT

- DS2 IDT deployment, as per the DOA all Modules must support Kubernetes (K8s) /Docker Containerisation technology. Thus, Modules must be Dockerised, and a Helm chart provided. IDTs will then initially be installed at a DataSpace/DataSpace Participant with default DS2 Modules, Rancher (including UI), K8s core features. As the user selects new Modules, they will then be visible/installable via their local IDT which will ensure and ease portability, inter-Module operation, maintenance updates and other classic K8s features.
- DS2 is not a commercial project aimed at specific DS2 use case requirements. The use case are typical examples of what might be user priorities, but the aim is to do this generically across all dataspaces and ensure that DS2 user needs of D2.2 are met. Thus, the requirements need to be interpreted more widely and generically. Technical partners were selected because of their expert knowledge of the technology and market, those features of Modules can be expanded well beyond DS2's use cases with the provision that these use cases take precedence.
- DS2 needs to consider the trust creation and access control (contracts, transfer control, enforcement etc) of the data space technologies used on participating dataspaces to ensure interoperability.
    - IDT is used for:User, organisation/participant, and data space registration.
    - Deploy selected Modules.
    - Communicate status and license information back to the DS2 Portal.
    - Provide a DS2 connector which connects to data space interoperability modules and other DS2 connectors in the ecosystem.
    - Provide link to logging services implemented by intermediary.
- IDT will manage the deployments of the Modules but not provide dataspace-enabling (value added) functionality which is provided by the Modules.
- DS2 technical providers should take advantage of the DS2 "experimentation dataspace (T6.1)" to test and validate Module both on their own, with other necessary or optional Modules and/or as necessary
- Where possible DS2 Modules should be open-source.
- DS2 will support the technical principles of the IDSA rulebook:
    - Not reinventing the wheel: use proven technologies.
    - Integrate existing systems: integrate data spaces into existing systems to the extent possible.
    - Integrate or use existing standards: align national and international specifications, technical standards, and established processes.
    - Industry and domain independent: make data spaces applicable as a concept as a horizontal standard.

- Easy to use: Low deployment threshold for companies and initiatives with a focus on portability and replicability.

- Modules will need to:
  - Be Unit, Functional, and User Tested in one of more DS2 use cases.
  - Implement needs of the portal, IDT, containerisation, marketplace and any other agreed functionality in the DS2 architecture.
  - Be resourced to do the above.
  - Be uploaded on the PORTAL Module marketplace (Including GitHub) at least as an executable (and if open-source code) with accompanying documentation, “how-to”, videos, Module descriptive information, helm charts which is needed for the deployment and ‘marketing of the Module.

7 CONCEPT INFORMATION

This section introduces:

- Tiers
- Nature of Modules

7.1 Tiers

To simplify the understanding of DS2 Modules and the reference architecture, the modular components of DS, they have been divided into Tiers as described below. The sequence of the Module annexes on the GitHub is in the same order.

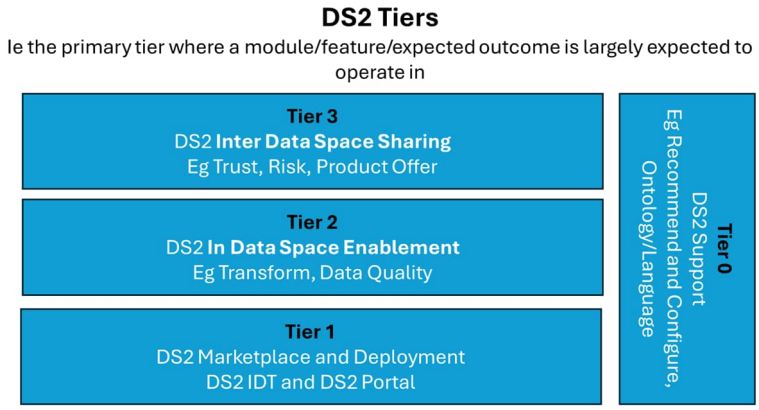
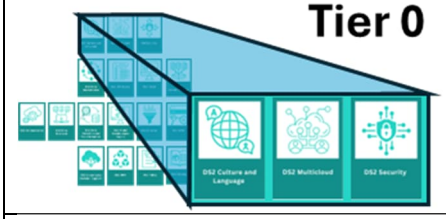
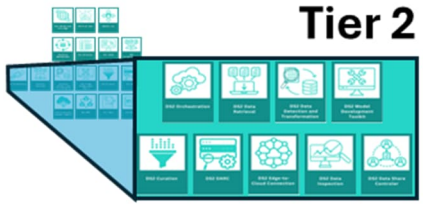


Figure 17. DS2 Tiers

	DS2 Support orientated: These represent a range of Modules that are perceived as cross-cutting to the other tiers and their Modules. They include: • Data Culture and Language • Multicloud • Security
	DS2 Marketplace and Deployment orientated: These support the acquisition, porting, and deployment of Modules at participants or service intermediaries: • Marketplace • IDT Broker • Portal • Containerisation • Also included are features: Dashbutton and Platform
	DS2 In-Data Space Enablement: The Modules facilitate consumer - provider participants in sharing their data. The Modules are in general used by a single participant only in their local environment but in some cases, there are additional features to be used between data provider and consumer. • Orchestration • Data Retrieval • Data Detection and Transformation

	• Model Development Toolkit • Curation • DARC - Data Space Discovery, Assessment, Recommend and Configure Module • E2C Connection • Data Inspection • Data Share controller
	DS2 Inter Data Space Enablement: These Modules can be allocated to the Tier 2 situation, but their added value is much greater in a cross-sector, cross dataspace scenario • Sovereignty Decision Support • DRM – Digital Rights Management Module • Policy • Catalog • Identity

Table 17: Tier list and description

7.2 Module Nature

This categorisation has been established to try to help users understand which Modules they are most likely to need for their data processing and operations. 'Nature' is cross-cutting to Tiers although in general optional Modules are associated with Tier 2 and 3 and the other categories with Tier 0 and 1. The Natures are:

- OPTIONAL
- SYSTEM
- FOUNDATION
- SUPPORT

In Detail:

Optional: These are typically Tier 2 orientated whose choice of installation is determined by the user. Generally, they deal directly with data or dataflow although DRM and DARC are also marked as SYSTEM since they are used alongside other Modules. They are focused upon use cases which are the only ones listed in the Requirements tables of chapter 4. They would be acquired from the DS2 Portal on a case-by-case basis, adding additional Modules as needed, and deployed via IDT.

DS2 Sovereignty Decision Support Module	SDS	K3.1	**T3**, T2	Optional
DS2 DRM Module	DRM	K3.3	**T3**	Optional/System
DS2 Orchestration Module	ORC	K4.1	**T2,** T3	Optional
DS2 Data Retrival Module	RET	Outcome	**T2**	Optional
DS2 Data Detection and Data Tansformation	DDT	K4.2	**T2**	Optional
DS2 Model Development Toolkit Module	MDT	Outcome	**T2**	Optional
DS2 Curation Module	CUR	K4.3	**T2**	Optional
DS2 Catalog Module	CAT	Outcome	**T3**	Optional
DS2 DARC Module	DARC	K5.2 & K5.3	**T2**	Optional/System
DS2 Edge to Cloud Module	E2C	K6.2	**T2**	Optional
DS2 Data Inspection Module	DINS	Outcome	**T2**	Optional
DS2 Data Share Controller	DSHARE	Outcome	**T2**	Optional

- System: System Modules are used to deploy or support other data related Modules, but which are not directly part of data sharing itself. DMK and MCL are marked as Optional since use is conditional on the context or user needs.

DS2 Culture and Language Module	CLM	K5.1	**T0**	System
DS2 Data Marketplace Module	DMK	Outcome	**T1**	System/Optional
DS2 MultiCloud Module	MCL	Outcome	**T0**	System/Optional
DS2 Security Module	SEC	Outcome	**T0**	System
DS2 Portal	PORTAL	K6.4a	**T1**	System
DS2 Containerisation Module	CONT	K6.1	**T1**	System

- Foundation: These Modules are intrinsic to the concept and operation of DS2 and DS2 implementation would not be possible without them. IDM is in green since it is a new Module not originally envisioned in the DOA as it was originally considered that existing Data Space identity providers along could take care of cross-DS identification without adaptation.

DS2 Policy Agreement and Enforcement Module	PAE	Outcome	**T3**	Foundation
DS2 IDT Broker	IDT	K6.3	**T1**	Foundation
DS2 Identity Module	IDM	Outcome	**T3**	Foundation

- Supporting: These are not Modules as such and are not shown in the main Module listings. They provide value-added feature support to 1 or more Modules. Information about them can be found in related Modules as follows:
  - [DASH] in IDT: Provides system wide holistic navigation
  - [PLATFORM] in Portal: Provides a centralised hosting service and test environment
  - [AILIB] in MDT/DDT: Provides model libraries
  - In addition, [EXP], The experimentation facility, is a special case and information around that can be found in Section 110.4.3.4.4.

DS2 AI Detection Module Libraries	[AILIB]	K4.2		Supporting
[Experimentation Facility]	N/A	Outcome	**N/A**	Supporting
DS2 Dash Button	[DASH]	Outcome		Supporting
DS2 Platform	[PLATFORM]	K6.4b		Supporting

8 TECHNICAL OUTCOME OVERVIEW

8.1 Outcome List

The list of technical outcomes produced via DS2, and their associated properties, is as per the table below. Primarily this includes software Modules deployed by participants (eg Orchestration) and also includes Modules which service a central function (eg Portal), or feature which are explained specifically (eg Dashbutton). Modules are generally indicated in capitals – eg ORC for Orchestration and features are bracketed eg [DASH] for Dashbutton. With this introduction, along with an understand of other DS technical concept, their fit in to the Big Picture Architecture in chapter 6 can be better understood.

Concerning the Modules the subsequent section presents an extract of each Modules purpose. The other properties in the table are discussed thereafter and all Modules, with the exception of the DS2 Orchestration Module, are presented to the same template on the GitHub (https://github.com/ds2-eu). The Orchestration Module is shown as an example in the main part of the document in Section 11.1 and also helps to identify the nuances of the standard template and is referred to in the texts in the rest of the document as an example.

The table columns are as follows:

- Primary Properties
  - Name
  - Abbreviation
  - Task
  - Results – Either DOA KER ID or project identified additional outcome
  - Tier - 0 to 3 – See Section 7.1
  - Nature - Optional, Supporting, Foundation, Environment, System
  - Owner – Owner(s) of the Module
- Deployment - Y=Yes, N=No, N/A=Not Applicable, C=Could
  - Within Single DS
  - Within Participant
  - Across DS without Intermediary
  - Across DS with Intermediary
- Additional Colour Scheme – See Section 10.4.1
  - Orange: Merged or Rebranded
  - Brown: Removed
  - Green: Special Case

Name	Abbreviation	Task	Result	Tier	Nature		Within Single DS	Within Participa	Across DS without Intermediary	Across DS with
DS2 Sovereignty Decision Support Module	SDS	T3.1	K3.1	**T3**, T2	Optional	UOS	C	C	C	Y
DS2 DRM Module	DRM	T3.3	K3.3	**T3**	Optional/System	ATC	Y	Y	N/A	Y
DS2 Orchestration Module	ORC	T4.1	K4.1	**T2,** T3	Optional	ICE	C	Y	Y	C
DS2 Policy Agreement and Enforcement Module	PAE	T4.1	Outcome	**T3**	Foundation	INDRA	C	N/A	Y	N/A
DS2 Data Retrival Module	RET	T4.1	Outcome	**T2**	Optional	IBM	Y	Y	Y	N
DS2 Data Detection and Data Tansformation	DDT	T4.2	K4.2	**T2**	Optional	SWAG	Y	Y	Y	Y
DS2 AI Detection Module Libraries	N/A	T4.2	K4.2		Merged into DDT	DIGI				
DS2 Model Development Toolkit Module	MDT	T4.2	Outcome	**T2**	Optional	INDRA	N/A	Y	N/A	N
DS2 Curation Module	CUR	T4.3	K4.3	**T2**	Optional	IBM	N	Y	N	N
DS2 Portability Module	N/A	T4.3	K4.4		Removed					
DS2 Interoperability Module	INT	T4.3	K4.5a	**T2**	Optional	IBM				
DS2 Catalog Module	CAT	T4.3	Outcome	**T3**	Optional	VTT	C	N	Y	Y
DS2 Culture and Language Module	CLM	T5.1	K5.1	**T0**	System	INTU	Y	N/A	C	Y
DS2 Data Space Discover and Assess Module (Part	N/A	T5.2	K5.2		Merged into DARC	ATC				
DS2 Recommend and Configure Module (Part of DA	N/A	T5.3	K5.3		Merged into DARC	ATC				
DS2 DARC Module	DARC	T5.2 &	K5.2 & K5.3	**T2**	Optional/System	ATC	Y	Y	Y	Y
DS2 Data Marketplace Module	DMK	T6.1	Outcome	**T1**	System/Optional	INDRA	C	N	N	Y
DS2 MultiCloud Module	MCL	T6.1	Outcome	**T0**	System	DIGI	C	C	Y	N/A
DS2 Edge to Cloud Module	E2C	T6.2	K6.2	**T2**	Optional	SWAG	Y	Y	Y	N
DS2 Data Quality Module [Merged in to E2C]	N/A	T6.2	Outcome		Merged into E2C	SWAG				
DS2 Security Module	SEC	T6.2	Outcome	**T0**	System	DIGI	Y	Y	Y	Y
DS2 Data Inspection Module	DINS	T6.2	Outcome	**T2**	Optional	INDRA	N/A	Y	N/A	N/A
DS2 Data Share Controller	DSHARE	T6.2	Outcome	**T2**	Optional	i4RI	N/A	Y	N/A	N/A
DS2 IDT Broker	IDT	T6.3	K6.3	**T1**	Foundation	i4RI & ICE	C	C	Y	Y
DS2 Dash Button	[DASH]	T6.4	Outcome	**T1**	System	i4RI & ICE				
DS2 IDT Integrated Broker, Modules & Toolkit		T6.4	K6.4			Rebranded into PORTAL				
DS2 Portal	PORTAL	T6.4	K6.4a	**T1**	System	ICE	C	N/A	N/A	Y
DS2 Platform	[PLATFORM]	T6.4	K6.4b	**T1**	System	ICE				
DS2 Containerisation Module	CONT	T6.4	K6.1	**T1**	System	ICE	N/A	Y	N/A	Y
DS2 Identity Module	IDM	T6.4	Outcome	**T3**	Foundation	ICE	C	N	C	Y

Figure 18. Module Matrix

8.2 Module Purposes

The Modules:

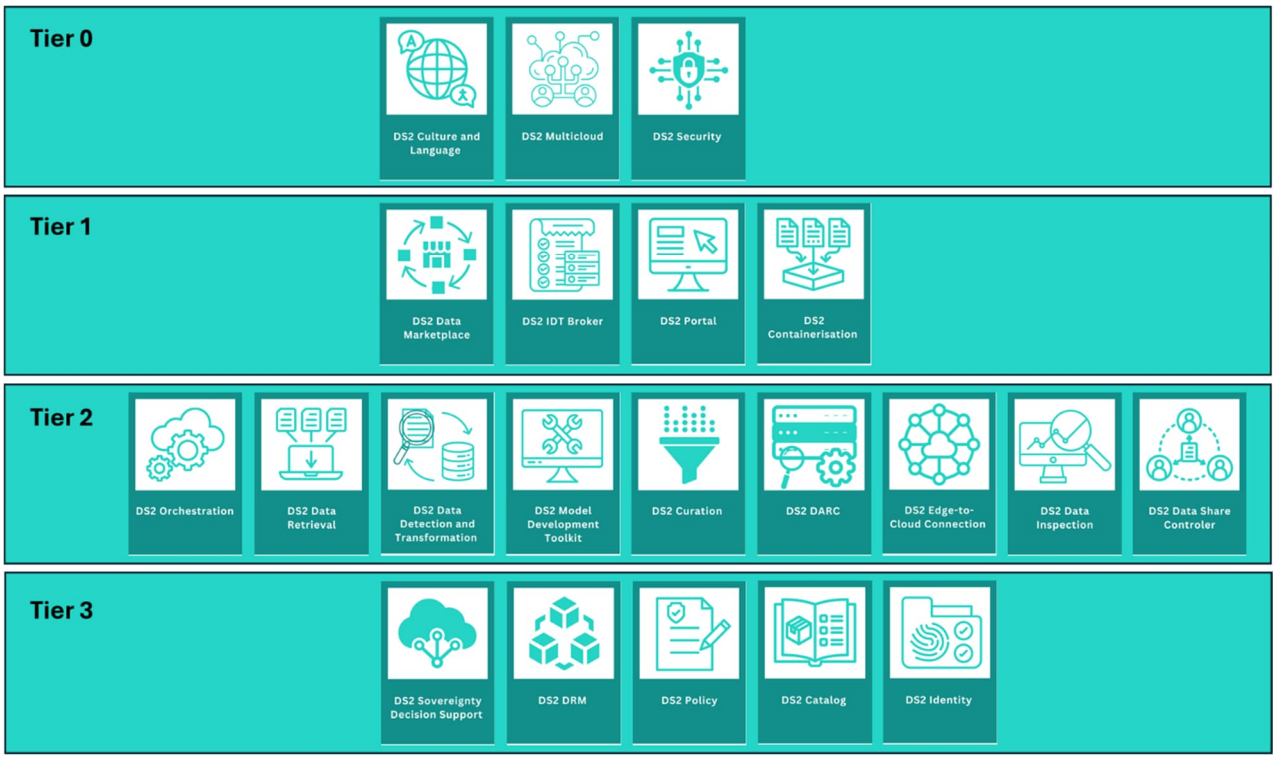
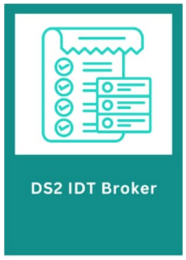
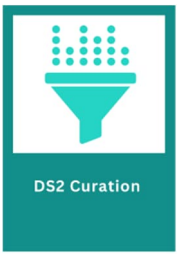


Figure 19. Tiers and Modules

From the detailed Module template; the list below extracts the "purpose" to introduce each Module, along with its common tag and icon.

Module	Purpose
Tier 0	
 CLM	The Culture and Language Module (CLM) aims to give a data consumer a better understanding of what data offers exist both within their own data space and in other data spaces, which may come from different sectors or countries and in different languages. This will increase the potential for use of a data offer. This is achieved through transformation of human language in shared information and queries into a rich, searchable hierarchical ontological description of the offered data set.
 MCL	The DS2 Multi-cloud Module (MCL) enables efficient transfer of discreet data, vast amounts of data, and streaming data between participants of dataspaces from data stores that are distributed across multi-cloud storage infrastructure. MCL includes intelligent data placement and caching at dataspace provider participants with a dataspace consumer participant requesting such data and provide services through use case applications(s). It will also ensure data exchange happens over secure connections using the DS2 Security Module (SEC).

 SEC	The DS2 E2C security Module (SEC) covers data security, data protection, and privacy with a focus on securing the edge-to-cloud data enablement and ensuring data quality and privacy. This involves implementing secure communication protocols, robust authentication mechanisms, encryption, anonymisation, and continuous monitoring of events in the DS2 ecosystem. The DS2 architecture is designed to manage large volumes of data, facilitating data-driven decision-making while maintaining stringent data security and data privacy standards. The proposed component works in conjunction with all DS2 architecture components.
Tier 1	
DMK 	The Data Marketplace Module (DMK) will provide a marketplace for data and data models. It will allow the registration of data from a catalog, record all transactions, and communicate transactions to any external system if required (e.g. Data Rights Management Module, Clearing House). Data will not be stored in the Data Marketplace Module. It will the capability to support not only datasets but also algorithms.
IDT 	The Intersector Dataspace Toolkit (IDT) is the core enabler of DS2 who purpose is to be deployed in front of participants data source/spaces and network connected to any other IDT-enabled data source. As such its aim is to run all DS2 Modules, including the DS2 Connector, the core Module for Inter-Dataspace communication and data transfer, and the Containerisation Module for DS2 Module deployment. IDT contains the core Kubernetes runtime to run all containerised Modules and a series of additional open-source software for Module management.
 PORTAL	To provide a user and developer friendly portal allowing dataspace participants to register and select DS2 Modules which can then be packaged into a IDT environment and then subsequently deployed by participants enabling both In-Data Space and Inter-Data Space operations. As such it includes functionality for developers to include Modules, users to find those Modules, to trigger the packaging through links with the containerisation Module, as well as supporting functionality for dataspace support, dataspace resources, registration and identity management, and administration. It also provides support for the Data Marketplace.
CONT 	To allow easy and automated packaging and deployment of Modules on the IDT Kubernetes runtime subcomponent environment. The containerisation Module leverages on custom Helm Chart descriptors to automatically convert them into full Kubernetes Helm Charts representing the Module, based on standard base templates located in the DS2 Portal Marketplace. The Helm Charts are then deployed on the IDT Module.
Tier 2	

ORC 	To design and then orchestrate at runtime In-Dataspace, Inter-Dataspace, internal, and third-party services which facilitate common data-orientated operations such as transformation of data, checks on data, data updates etc. The orchestrator contains a flexible GUI to design workflows and decision points on these services and run time component to implement the workflow.
RET 	A DS2 data catalogue for a data space can potentially reference a large number of data sources accessed by different backend applications. All requests for data must go through a data space connector both at the consumer and producer side. For a connector to access stored data, the connector must generate the required API code to communicate with the storage backend. This can represent a major learning and programming effort even for an experience programmer and is certainly not a task which the average non-technical data space consumer can easily take on. It is the goal of the Data Retriever Module to vastly simplify the effort required to access a data source by automatically generating the required REST call for the connector, based on a textual request in natural language.
DDT 	Dataspaces allow for data to be shared between data providers and data consumers. A lot of data comes from sensors and devices at a high rate. To allow for a well-defined data structure and quality during the data generation and exchange, DDT is a Module that can analyse data on the fly.
MDT 	The main purpose of the DS2 Model Development Toolkit Module (MDT) is to provide a set of tools to allow the users to develop algorithms based on the CRISP-DM standard to assist in the whole development cycle (training, test, etc.) and package the algorithms which can be deployed as executable software component.
CUR 	Data obtained from disparate sources runs the risk of remaining siloed unless it is curated to match the format of similar data from other data sets. Manual curation of datasets, however, can be a labour-intensive task, and not suited to DS2's dynamic nature of federating dataspaces. The aim of this task is the automatic creation of pipelines to curate data through machine learning. The Data Curation Module is invoked on two or more data sets and aims to identify through machine learn data transformations which need to be performed on fields in order to allow interoperability between the data sets – for example, the conversion of time-data formatting. The required transformation(s) will be automatically selected from a transformation library, and a processing pipeline will be created and executed to curate the data.

DARC 	To inquire, discover and assess though the conversational UI of an AI-driven agent, the In-Dataspace and Inter-Dataspace data-oriented capabilities and limitations of DS2 Dataspaces as well as DS2 Modules (software prerequisites and/or recommendations) which will compose the "ideal use" scenarios/ paths that will fit the needs of the end-users. Then, to complement this by recommending to end-users, though the conversational UI of an AI-driven agent, the best DS2 Modules for the implementation of the selected "ideal use" scenario/ path, based on the outcomes of the T5.2 Module end-user interaction. Finally, to demonstrate the autoconfiguring, through APIs, a subset of Module based on these recommendations so that they can easily create their DS2 pipeline and start using and sharing data.
 E2C	Dataspaces allow for data to be shared between data providers and data consumers. This includes data coming from sensors and devices at a high rate. This Module is used to establish a secure edge-to-cloud connectivity for data providers to cloud-based IoT platforms like Azure IoT or Cumulocity IoT or AWS IoT via a MQTT bridge. The data providers will decide which data to share and with whom. In addition, the data quality can be monitored as well.
 DINS	The Data Inspector Module (DINS) facilitates the configuration and deployment of processes for real-time data analysis, ensuring data quality and compliance with thresholds set by the parties involved. It performs several key functions: generating notifications based on the values of the exchanged data, executing reactions such as sending requests and notifications to external tools, and integrating with models developed to enhance its capabilities. It is a complement to the Data Share Controller which focuses on control information with both Modules using the Data Interceptor.
 DHARE	To provide a user orientated view of control plane information related to a specific exchange of data to monitor its status and to potentially limit or block it. It will access data through a Data Interceptor component which it shares with the DS2 Data Inspection component (DINS) which operates more at the data level. It can be seen an In-Dataspace enablement Module. Its role is especially important in an Inter-DS environment to provide extra monitoring and control of the data exchanges when partners are less known.
Tier 3	
SDS 	Contracts and identity management technologies provide a foundation for secure data sharing, but they are not sufficient on their own to establish trust in the process. Further information needs to be shared between provider and consumer to ensure that the decision maker can take an risk-informed decision when sharing data. From a practical point of view this means that a mechanism for sharing and then analyse some information on the infrastructures and data management systems needs to be setup.

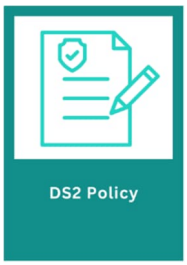
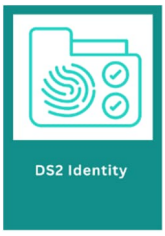
DRM 	To enhance the management and security of digital asset transactions through a robust blockchain-based Data Rights Management (DRM) system. It is designed to perform critical functions, including the notarization, tracking, and validation of all data rights transactions both within individual Dataspaces and across multiple participating Dataspaces
PAE 	The primary function of the Policy Agreement and Enforcement Module (DS2 PAE) is to ensure compliance with the established policies and regulations governing data exchange among users in different data spaces. Henceforth, policies, regulations, and agreements are synonymous with the term policy. The policies are evaluated as the control plane stage of data sharing in the Connector. The policies serve two main purposes: Access Control and for Usage Control. Access Control determines whether access to data is granted or denied. Usage Control dictates how the data can be used once access is granted.
CAT 	The Catalogue Module is a Module designed to support the exchange of data within and across different data spaces. It ensures robust data governance, secure data exchanges, and compliance with sovereignty requirements. The main goal of DS2 catalogue is to enhance the functionalities of catalogue systems within existing reference architectures, enabling them to support both intra-data space and inter-data space operations. This includes defining data models for data product offers, data product offer searches, and interactions with members of other data spaces, thereby fostering collaboration across different data spaces. It is listed as an optional Module since it is technically possible to manage this at the participant level and this creates a lot of overhead for data consumers and providers, however in reality it is a core essential Module of DS2 for most practical dataspace sharing scenarios.
IDM 	The DS2 Identity Module (IDM) is a 'system' since it relies on a central DS2 Identity Module deployment, IDTs which are pre-configured to communicate with this Module, and access to/from individual dataspace identity Providers. It aims to provide a practical framework for the creation and validation of participant identities for inter-Dataspace activities based on the existence of existing dataspaces and their own individually selected Identity authorities. This is linked to the IDT Connector and allows for a federated approach of the connectors whilst relieving participants from connector interoperability, outside dataspace change and maintenance issues, and minimising or eliminating the changes to their existing environment.

Table 18: List of modules

8.3 Module Fit re Problem Space

From superimposing the Modules on the problem space envisioning of the technical building blocks DS covers them as follows showing a good coverage of the functionalities needed and with a focus on individual participant enabling value added services:

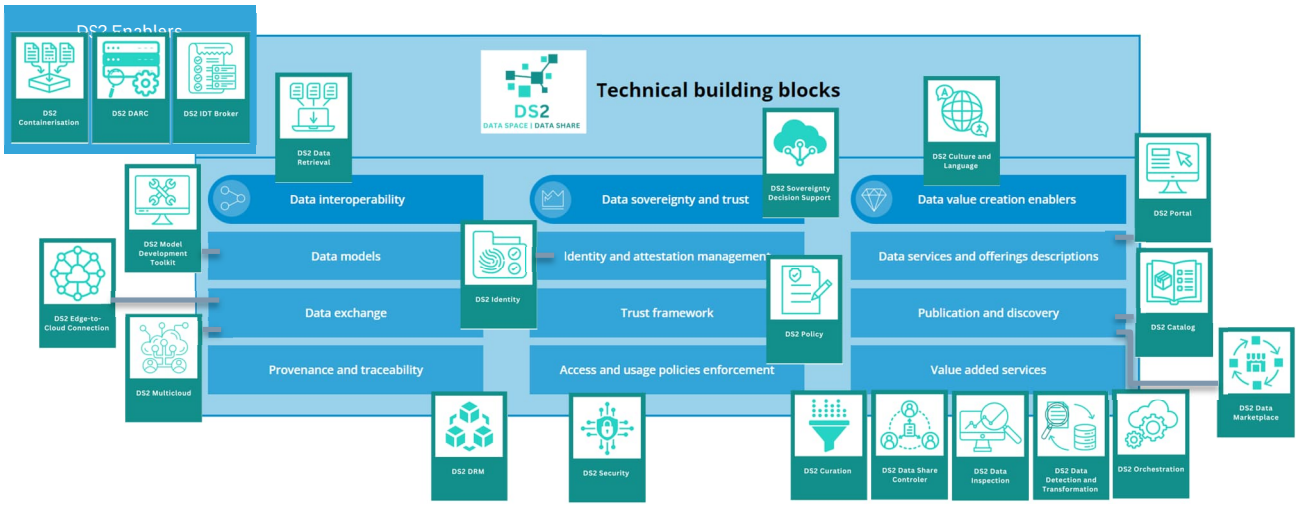


Figure 20. Module Fit

9 BIG PICTURE ARCHITECTURE

9.1 General

One State-of-the-Art item in DS2 Description of Action referred to the “Intersector Dataspace Challenges & Reference Architecture” (O2, WP2) indicates that the challenges of wide data sharing and its subsequent real utilization are still in their infancy and there are several challenges hindering the ultimate benefits of data spaces particularly when considering interaction of data spaces across different domains. Mentioned items and DS2 Module examples include handling vast amount of data (MCL), data sovereignty (SDS), data interoperability (RET), and human language (CLM) issues at all stages of data life cycle.

DS2 promised that its architected Modules would realize the ultimate benefits of data sharing taking, where possible, advantage of other initiatives (eg IDSA, GAIA-X) and standards in the area if existing. DS2 proposes a reference based on federation, distributed consensus, decentralization, transparency, controllability, portability, regulation by automation and interoperability across data and services. DS2 also promised to enable a pick-and-mix of multiple software Modules developed, and then selected primarily by dataspace participants, and potentially dataspaces, to be part of their deployment to enable data sharing.

Thus, as per the technical principles, the nature of DS2 is holistic, modular, pick-and-mix, and platform independent. Then, the DS2 universe must be recognised as infinite such that it can contain an unlimited number of modules of distinct and/or overlapping functionality, with optional/mandatory links to other modules if the module developer determines it appropriate. The module acquirer simply needs to select which module(s) they determine match their requirements, based on for example functionality match, price, support level etc, and/or how such a module may interact with other modules they wish to use.

Hence, there is not one architecture which each and every module fits into but instead each Module each has their own architecture and independent functionality but where 1+1=3 functionality increase can be obtained. However, every Module needs to fit the DS2 Architectural reference model. The sections below show this high-level reference architecture, pick-and-mix nature, the Modules, and environment interactions and the win-win interconnections where utilising single Modules can add value, but interconnected Modules may add more, as well as looking at deployment and use.

9.2 Reference Architecture

9.2.1 General

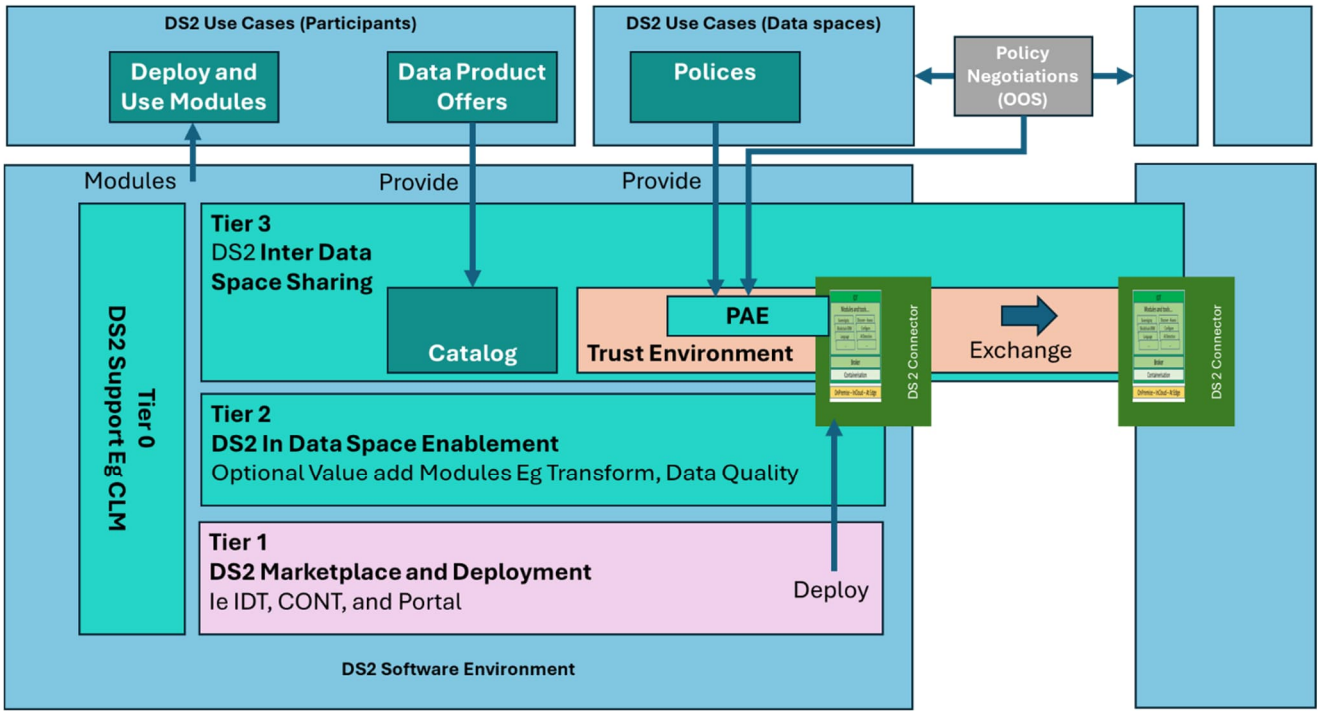


Figure 21. Reference Architecture

The top-level reference architecture is illustrated above and as stated is primarily a framework for the deployment of standalone optional Modules as well as those which are near-mandatory for Inter Data Space sharing at Tier 3. It delivers on the DS2 mission of a "A modular, secure, trust-sensitive, platform neutral environment for the networked sharing of data". The focus in this section is on Tier 1 (Marketplace and Deployment) and the Trust environment of Tier 3 since there is a greater level of interaction and dependency.

- Tier 0 (Supporting eg CLM) Modules are broadly self-contained and simply support other Modules on a case-by-case optional basis and deployed as per the general Tier 1 case. As largely independent Modules functionality should be assessed from the individual Module templates
- Tier 1 (Marketplace and Deployment eg IDT) concern how Modules are made available from the DS2 PORTAL, selected, and deployed at participants. Its deployment mechanisms apply to all Modules in Tier 0, 2, 3, with the Modules being packaged into IDT with its embedded connector
- Tier 2 (In Dataspace Enabling Modules eg ORC) are, similar to Tier 0, self-contained Modules but providing specific functionality which individual participants may find attractive to deploy. Whilst self-contained in some cases advanced functionality can be added by operating with other Tier 0,1, or 3 Modules
- Tier 3 (Inter data space sharing eg PAE) represent a near mandatory set of Modules, with several inter-relationships connected with the DS2 Trust environment (and also the provision and acceptance of Data Offers via the CAT Module
    - Trust Environment: This is made up primarily of Modules PAE, IDM, SDS, DRM although the PORTAL and [PLATFORM] will support their operation and other Modules (eg DINS, DSHARE) may take advantage of some APIs particularly from the PAE Policy Module. In addition, in

terms of PAE this receives policies from the participant and in a DS-DS sharing context from the Dataspace and Dataspace pair once a DS-DS arrangement is negotiated noting that the negotiation itself is outside of the scope of DS2. See following Section 9.2.2.

- Catalog Module: In terms of the Data Offers they need to be defined by the participants and provided to CAT. Details on CAT are provided on the GitHub (https://github.com/ds2-eu) since it is broadly self-contained.

9.2.2 DS Trust Environment

Whilst many elements of DS2, especially those orientated to individual participants for data sharing (eg DDT) may not necessarily rely on inter-participant 'trust', many others will need this especially in a DS-DS data sharing environment. DS2 defines this as its trust environment which is composed of the following:

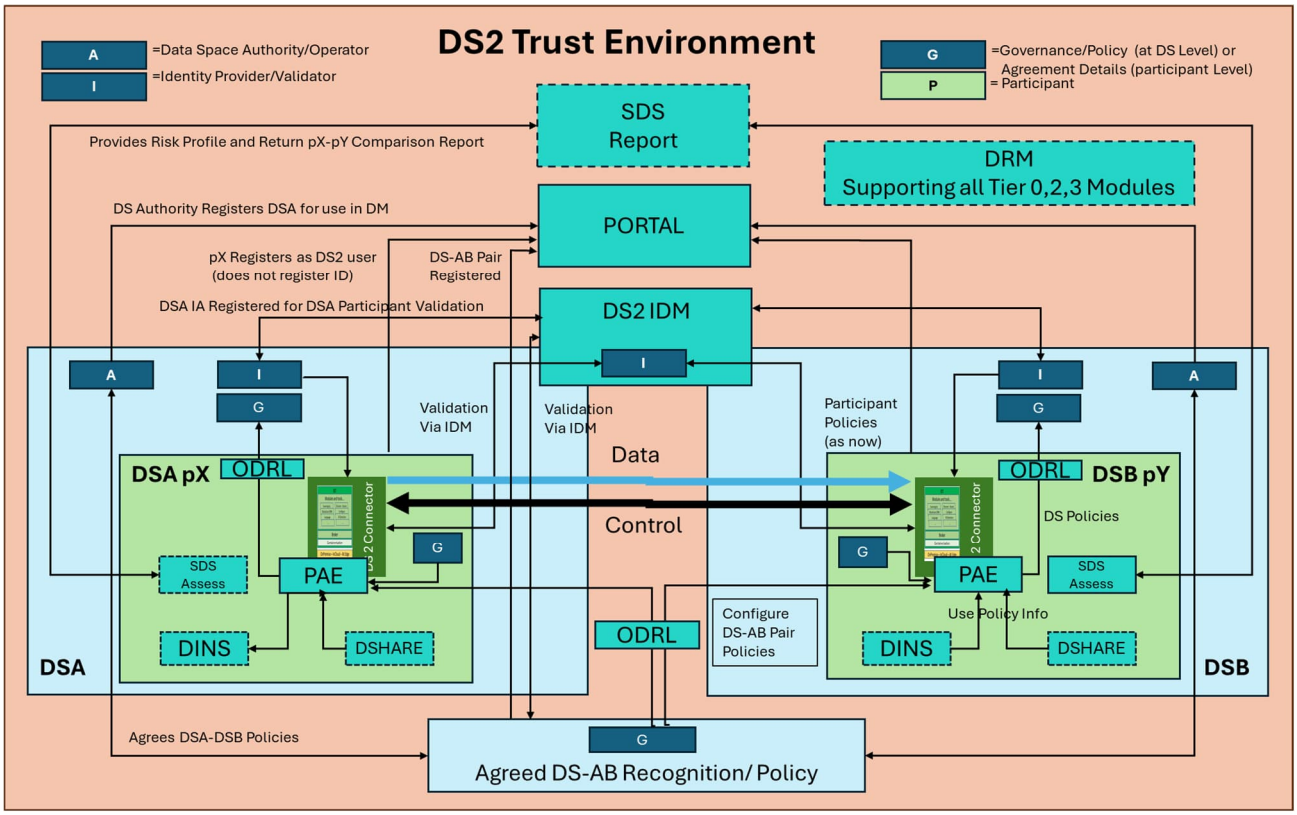


Figure 22. Trust Environment

In brief:

- Data sharing entities can register as users in the DS2 portal which gives access to eg DS2 Modules via IDT and other assets about DS2. This is necessary for IDT deployment and DS-DS sharing
- Authorities of Dataspaces who may wish to intersect can, outside of DS2, make policies/agreements (eg accept each other policies as of, eg negotiate a (minimum) common policy):
    - The basis for this discussion is on the methodologies/templates of WP3 and the results would include ODRL structured definition of the agreements
    - Those authorities making those agreements, and the resulting DS-pair, are both registered in the PORTAL and the DS2 identity Module IDM and tokens specifying this are shared with relevant participants
    - For validation, then via IDM, the eg Provider participant shares the Dataspace ID, Data pair ID, and Participant ID via the Connector control plane and once received are validated by

IDM which in turns requests local Dataspace provider identity authorities to validate participant IDs

- The DS2 PAE Module takes dataspace, dataspace pair and dataspace governance policy rules in ODRL format and attempts to apply them and trigger enforcements noting:
    - Access control actions (eg Allow/deny datasharing) can be enforced in the Module
    - Usage control actions (eg delete after three days) enforcement actions are outside of its scope since they are rule dependent and need to be performed elsewhere
- The DRM Modules may be deployed by participants for notarisation (and retrieval) of events and information. Other Modules may hook-in to it if deployed
- Other Modules such as DINS, DSC may take advantage of structured policy information through DINS
- The optional SDS Module can be used by individual participants to i) Assess their own risk profile independently; ii) if consumer/provider participants agree, share their risk profile, and use a comparison service from DS2 which will highlight risk differentials

9.2.3 DS2 Module Deployment

In general deployment is related to:

- Underlying Technology – in this case Modules:
    - DS2 "Trust environment" which is broadly composed of identity (IDM), policy (ODRL Rules no Module)), risk (SDS) and agreement/enforcement (PAE) parts – as described beforehand
    - DS2 containerised Module deployment via the portal (PORTAL), Intersector Dataspace Toolkit (IDT), and Containerisation (CONT) Modules – as described below
- Deployment location which is a combination of the following and addressed in Section 9.2.4:
    - In Data space or Inter Data space
    - Local or via a service intermediary
- User Orientated steps: Orthogonal to the above are the following which are covered in subsequent sections along with an indication of a practical sequence in Section 9.5.5:
    - Pre-deployment step – Policies and Registration
    - User deployment
    - Implementation

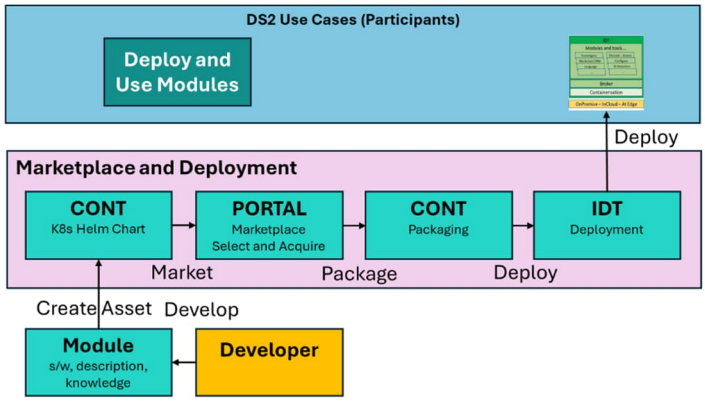


Figure 23. Marketplace and Deployment

A DS2 Module developer, either within the DS2 project or outside of it, will create a Module according to the principles and guidance of this document amongst others. The Module need accompanying material such as documentation and a product descriptor – See more on GitHub (https://github.com/ds2-eu) as well as of course the necessary quality/test regimes, security scans etc. Once registered on the DS2 Portal the developer uploads this material along with a K8s Helm Chart assisted by the CONT Module. This new DS2 asset then becomes available on the PORTAL Marketplace and participants can search and select it, as necessary. On 'checkout' the PORTAL calls on CONT again to package the Module as an IDT deployable asset and if a new instance will also include other detail Modules – specifically the mandatory trust environment Modules. The Participant then deploys IDT and the selected Modules configuring them, as necessary.

9.2.4 Deployment Location Scenarios

DS2 Modules will be (largely) deployable either OnPremise (typically at a Participate) or InCloud (On the DS2 [PLATFORM], a Service Intermediary of DS2 or a Dataspace provider, or a hosting Service of a Participant). The 'location' itself is largely Participant driven except in the case of System Modules operating in the center of DS2 – notably the PORTAL.

In the context of DS2 the location in which a Module fit is in the data sharing Big Picture diagram, an example for ORC is shown below and similarly so for each Module template on the GitHub (https://github.com/ds2-eu) - Sections x.1.2.1.

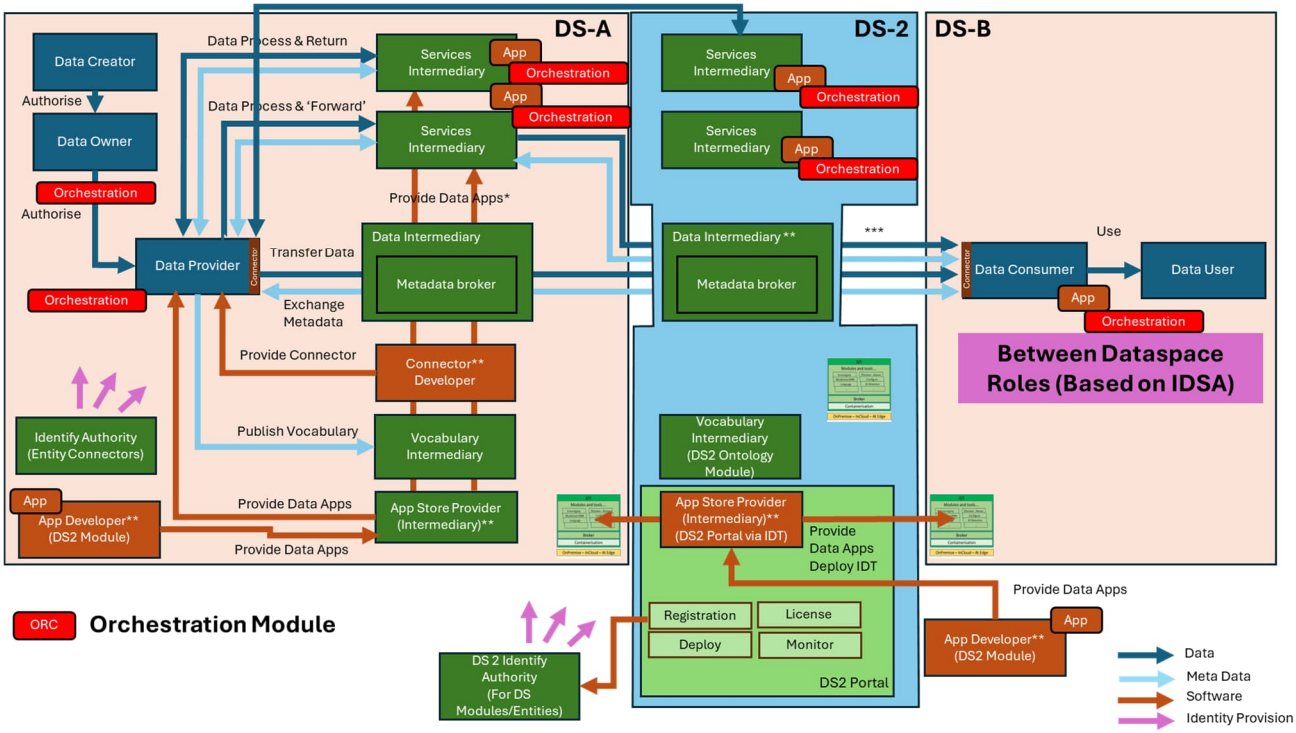


Figure 24. Big Picture

This high-level indicates different deployment scenarios for each Module noting that many Modules have multiple scenarios but not all will be validated in DS2 whose validation will focus more on the more complex cross sector scenarios. In the example above it can be seen that ORC is potentially deployable in many locations exampled in the table below which shows these 'Where' deployment situation along with a Comment column and an example from the Orchestration Module to put this into context. This information is also summarised in the Outcome list of Section 8.1 and further detailed in the relevant Module subsections.

In terms of the coding for the Outcome List:

- Yes: Module is used here, is further documented in the Module, and will be implemented and validated
- No: Module is not(normally) used here and will not be discussed
- Could : Module could be used here, but typically stated when an In-Data Space deployment is the same or highly similar to an Inter Data Space (across Data Space) scenario, so implementation is the same and validation has no added value
- N/A: Not Applicable and not discussed

Where	Comment	ORC Example
Within a single Dataspace for use between participants in that Dataspace only	Module could be deployed within a single dataspace with an inference that it acts 'between' participants (of the same dataspace) so typically it is deployed by a service intermediary of the dataspace. However, it can also be deployed by participants and instead they would directly engage the Modules features with other participants. This is a generally a subset of one of the 'across dataspace' scenarios below but involves 'more-known-to-each-other' participants from the same data space and participants that have already accepted the same dataspace rules only. As such in the context of DS2 and as a subset this is generally a COULD or a NO or N/A	Could. This is not piloted/validated in DS2 but the in-Dataspace version would work similar to the two Across Dataspace scenarios without the need for more complex identity checking
Deployed and used by a single participant to enable the participant in either an In-Data space or Inter- Data space scenario	This is the main In Dataspace scenario. Regardless of the participant, the Module is used standalone by the participant. This is typically relevant for Optional Modules and Tier 2 Modules which operate on data or a data flow to process data in some way between business application/store and to/from a connector	Yes: NB can still orchestrate external services (eg non-dataspace in other dataspace participants) if user requires and at users (probably necessary) choice can implement DS2 trust mechanisms for service action
Across Dataspaces without Service Intermediary	Module is deployed at a participant in one dataspace (or a service intermediary of the participant in a dataspace) and there is a direct connection between that a the "same" Module within another participant – eg ORC. They key is that there is no intermediary operated by DS2 or any other 'central' intermediary providing brokering features.	Yes: Would then mean orchestration Modules "liaise" and interoperate with each other

Across Dataspace with Intermediary	This is the main Inter Dataspace scenario. The Module is deployed between two dataspaces by a (inter) dataspace intermediary, which is typically, DS2 and the Module is generally hosted by the DS2 PLATFORM but not necessarily so. It can also be that there are 'endpoints' or 'end-interfaces' and consumer and provider participants that link to this	Yes. Potentially could be performed by a Service Intermediary [not covered in DS2

Table 19: Outcome list

9.3 Interaction Matrix

As part of the Module definition an important table is the "Gives To/Receives From table" which should reflect in a 1:1 way what is in a Module external interface. From past experience this has seen to be one of the most critical definitions in multi-partner research project since one developer often 'assumes' that a) something/one else will do what is needed and b) the something has exactly the same scope/data/control that they had expected. Extensive DS2 discussion around this has been one of the main elements to ensure this consistency. There are certain interfaces that are not documented in the Modules themselves since they are more common but are documented in the matrix so Modules owners can ensure they implement – for example IDT or Dashboard.

Thus, the table directly below highlights the interfaces, and the second table overleaf consolidates them enabling leads to ensure common ground. A partial example table, taken from Orchestration, follows and should be read as follows… "The Module being document (ORC), Gives To Module PORTAL the WHAT text."

With Module/Feature	Receives From/ Gives To	What
Portal	Gives To	When published on the portal information (technical, "How to's" etc etc) will be provided according to the general model
Tier 3	Receives From	Authenticity of participant information in participant-participant scenario
Tier 3	Gives To	Participant identity
Services	Receives From	Service End Point information of Other external Modules/Services. DS2 Service Intermediaries, DS2 Modules/Services
Services	Gives To	Run time triggering of end point APIs
DRM	Gives To	Log/Control information for notarisation
DRM	Receives From	Log/Control information from notarisation

Table 20: Gives To/Receives From

The currently ascertained interaction matrix is coded as follows:

- G: Gives To
- R: Receives from
- C: Common – Not necessarily documented as a Common Mandatory link in Modules but is a mandatory implementation
- [C]: As "C" but optional
- Header Row:
    - Blue: DS2 Modules (eg IDT)
    - Orange: Strongly Related 'external Modules or DS2 features (eg Dashbutton)

- Purple: A Module tier – ie all (most) Modules in that Tier. Nearly always applies for Tier 1 since Modules need to connect to the Portal/Dashbutton/etc but some Modules cross-cut multiple other Modules – eg DRM
- Green: External systems

Further notes:

- Some Module texts refer to the 'MetaData Broker' which is a sub-component of CAT

Module – Module Relationships

Module in Question in Column	SDS	DRM	ORC	PAE	RET	DDT	MDT	CUR	CAT	CLM	DARC	DMK	MCL	E2C	SEC	DINS	DSHARE	IDT	PORTAL	CONT	IDM
SDS		[C]									G								R		
DRM				G																	
ORC		GR[C]								GR	GR								G		
PAE	R	R[C]								GR		R									
RET		[C]							R												
DDT		[C]								R											
MDT		[C]														G					
CUR		[C]	R						R												
CAT		G[C]								R	GR								G		R
CLM		[C]							GR		GR										
DARC	GR	[C]							R	GR											
DMK		R		G					R										R		
MCL		[C]							GR												
E2C		[C]																			
SEC		[C]																			
DINS		[C]					R										GR				
DSHARE		[C]		R												R					
IDT																			R	GR	
PORTAL												GR								GR	GR
CONT																		GR	GR		
IDM																		GR	GR		

Figure 25. Module-Module relationship

External Modules and Features + Tiers (Entire Groups of Modules)

Module in Question in Column	[POL] [Regulations, Policy]	[CONN] DS2 Connector	[AILIB] Model Libraries	[Dash] Dashbutton	[Platform] Platform Resources	Tier 0	Tier 1	Tier 2	Tier 3
SDS				C					R
DRM		R		C				R	R
ORC				C					GR
PAE	R			C					
RET		R		C					
DDT				C					
MDT			R	C					
CUR				C					
CAT				C					G
CLM				C					
DARC	R			C				GR	
DMK				C					
MCL		G		C					
E2C		GR		C					R
SEC		R		C					
DINS	TBD			C					
DSHARE		GR		C					
IDT		GR		C		GR	GR		GR
PORTAL				C	GR				
CONT				C					
IDM	R			C			R		

Figure 26. External Modules and Features + Tiers (Entire Groups of Modules)

External Interfacing Systems and Users

Module in Question in Column	DS Identiy Providers	Business Appication	Data Source	External Apps	External Storage	External Services	Sensors	Data space Participant
SDS						R		G
DRM								
ORC						GR		
PAE								
RET								
DDT			R					G
MDT								
CUR								
CAT								
CLM								
DARC								
DMK								
MCL				GR				
E2C			R	GR		GR		
SEC							GR	GR
DINS								
DSHARE		GR						
IDT				GR	GR			
PORTAL								
CONT								
IDM	GR							

Figure 27. Module Give-Receive Relationships

9.4 Interaction Pictorial Links

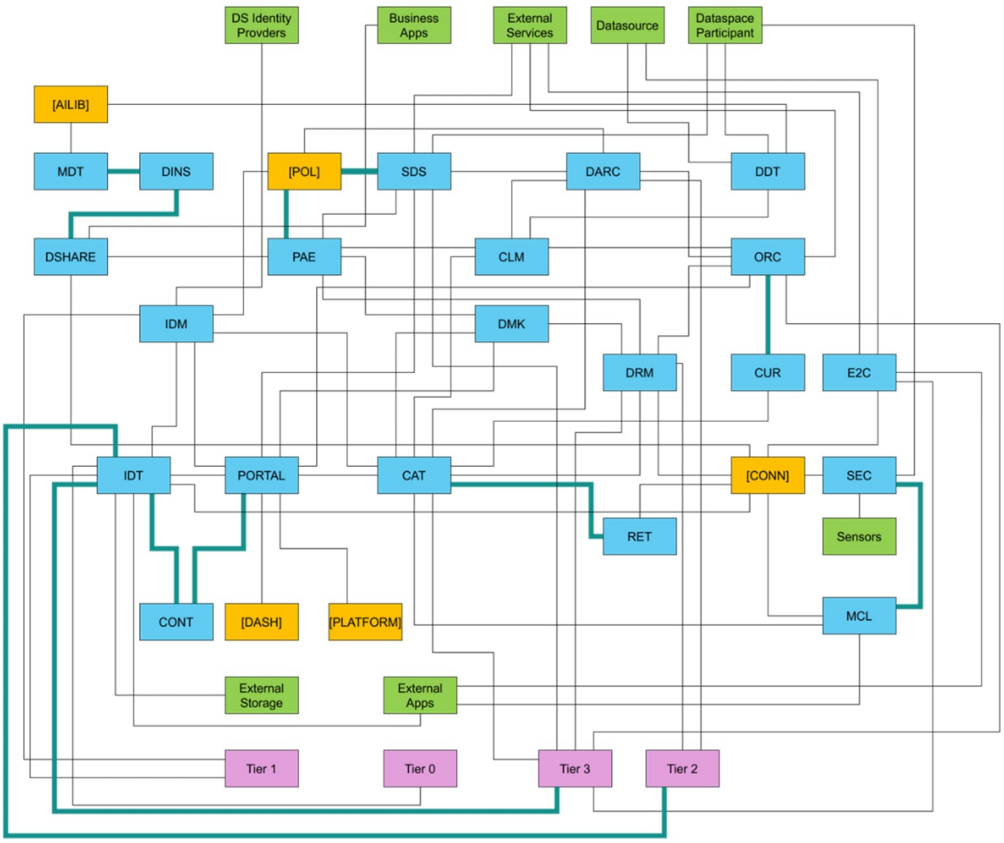


Figure 28. Module Relationships

Above is a pictorial diagram of the links above according to the same colour coding schema. It helps developers in particular understand the specified Module relationships. The bolder arrows represent strong (mainly mandatory) linkages whereas the lighter ones represented added-value relationships.

9.5 User Deployment and Implementation Steps

This section covers user orientated deployment steps.

9.5.1 *Pre-deployment step – Policies and Registration*

Pre-deployment steps are orientated around:

- Pre-deployment – Policies
- Pre-deployment – Registration

9.5.1.1 Pre-Deployment – Policies

As part of the on-boarding process, specific global Data access and usage policies may be defined. There are two types of policies, Access Control which defines who can access the data and Usage Control that define how the data can be used. Access Control policies are usually applied when deciding to share data and Usage Control policies are applied when sharing the data, so they involve actions when the data is already being shared.

At the Dataspace level, a specific Dataspace can define specific global policies using a policy definition language, ODRL. Once existing and defined, those policies can apply to all DS2 transactions that imply a participant from that Dataspace. These aspects are detailed in the outputs of WP3.

Next, at the Dataspace-to-Dataspace relationship level, further policies may be defined that will apply on top of the Dataspace ones, for transactions between participants of those Dataspaces.

Finally, on top of these Dataspace orientated policies, user defined policies for data sharing will apply as in a regular intra Dataspace transaction between Connectors.

In addition, as part of DS2 ecosystem, the Sovereignty Decision Support (SDS) Module is used to perform a risk analysis at the participant level. A subset of the result of such analysis might be shared between participants, using a reporting format, to further provide additional decision points for data sharing. These policies can be considered as part of the global Dataspace ones and will also apply when deciding to share data.

In summary, there will be “policies” at the following levels:

- Participant: Defined by a data provider as a requirement to share data
- Dataspace at DS2: Policies defined at the Dataspace level that will apply to all transactions related to participants within that Dataspace
- Dataspace to Dataspace relationship at DS2: policies defined at the Dataspace-to-Dataspace relationship that will apply to all transactions between participants of those Dataspaces
- Participant Risk Assessment: Defined and published for a data provider and consumer related to risk status in participants IT infrastructure and software using the SDS Module

The policies will be evaluated and can be enforced using the Connector capabilities when agreeing a contract but also with the enhancements from the Policy Agreement and Enforcement (PAE) Module, that will consider all the policies for a given transaction, typically for Access Control type policies, but further research will be performed for Usage Control type policies using the PAE and the Data Share Controller (DSC) Module

and the Data Inspection (DINS) Module. The enforcement, for example in PAE, needs to be carefully considered since the benefit of policies and then in applying rules only apply where providers want to enforce in their connectors and that consumers want to enforce in theirs.

9.5.1.2 Pre-Deployment – Registration

Another initial pre-deployment step, as part of the on-boarding process of a Dataspace and Dataspace participants in DS2, is to create and assigns the corresponding IDs to be able to identify and validate a participant within a Dataspace and a Dataspace-to-Dataspace transaction which is handled and more-fully described in the IDM Module

The pre-deployment requirement for inter participant sharing using DS2, prior to acquiring and installing the DS2 technology also includes the PORTAL Registration of:

- The Dataspace(s) itself as an entity (eg Dataspace A – DSA and DSB)
- Once both Dataspace are registered, registration of a relationship to share data with another DS2 registered Dataspace (eg DS-AB)
- Access details to any dataspace dataspace-participant identity provider
- Users or organisations wishing to explore, and potentially access/deploy DS2 technology register via the PORTAL including, where applicable and optionally:
    - Information of (registered) dataspace(s) they may belong to
    - Information of their participant ID in a particular dataspace noting that DS2 will in principle rely on Identity Authorities of specific dataspaces to both create registration IDs for new participants and for their validation

9.5.2 User Deployment

Once the pre-deployment steps are completed, and Dataspaces, Dataspace relationships, organisations, and Users are registered in the DS2 Portal, participants can start the Deployment stage, that is, the stage where participants will acquire and install both the required and optional DS2 software.

Once logged in to the Portal, the core enabling Module of DS2 is IDT. IDT is the key software that is used to run all other DS2 Modules, including the DS2 Connector. Users will acquire IDT from the DS2 Portal Marketplace as mentioned earlier and install it on their infrastructure. Requirements and steps for installation will be detailed in the on-line documentation on the Portal who scope is detailed on GitHub (https://github.com/ds2-eu).

During the process of the IDT installation, all the core Modules required will also be installed, such as the DS2 Connector required for Data transactions, the Containerisation Module, required for installation of additional DS2 Modules, and some others.

Having IDT and all the core Modules installed is the minimum requirement to enable DS2 transactions, but further functionalities and Modules can be used to enhance and extend the data sharing experience in DS2.

Users will go back to the Marketplace to further explore additional optional Modules that fit their needs for their data share use cases, for example ORC, and will be able to acquire and install them on top of the IDT. Those Modules in the Marketplace can be offered for free or at a cost depending on the Module provider.

9.5.3 User Implementation

The installation and configuration of IDT and DS2 Modules depends on the Dataspace and Organisation themselves, but DS2 intermediary organisations could also provide some consultancy services for this, or even provide a cloud service for organisations not willing to go through this process. Larger organisations are likely to have their own IT systems and IT staff, able to install and run their own IDT and Modules whilst SMEs

may be more willing on using the as a service model. These services do not depend on DS2 as a project but rather in individual DS2 participant organisations willing to provide them.

For users and organisations installing and running their own IDT(s), the IDT installer will be acquired via the Marketplace as explained in the previous section. Then the installer will be run and will guide users through the different installation and configuration steps for IDT (eg addition of a participant identifier). In the case of the as a service model, this step will not be necessary since IDT will be ready in the intermediary.

Regardless of the implementation model, InCloud as a service or OnPremise, once IDT is up and running, participants will be able to install and configure additional Modules via the Marketplace, on top of their installed IDT. Regarding the additional Modules, some configuration(s) will be automatic, relying possibly on previous configuration information already performed when configuring IDT during set-up, but depending on the Module being installed, additional configuration may be required.

DS2 and its Modules will provide detailed information on how to install and configure those Modules and link them with other Modules in value-added scenarios. In addition, there will be some DS2 services and Modules, such as chat bots and others, aimed to experimentally autoconfigure, to some extent, such Modules.

9.5.4 User Process Implementation Steps

Beyond the deployment and implementation of the modules, users will also need to set up their data sharing environment which broadly consists of:

- Data Creation and Preparation (Provider)
- Data Product Management (Provider)
- Data Contract Management (Provider and Consumer)
- Data Processing and User (Consumer)

The precise steps and target are shown in the diagram below:

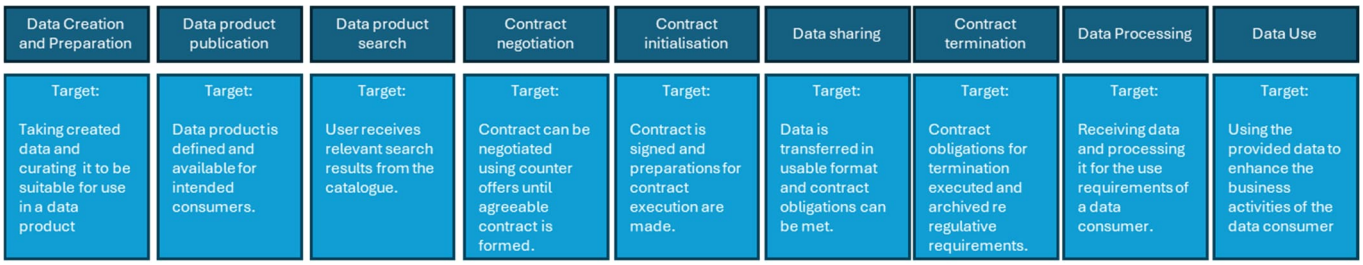


Figure 29. High Level Process Steps

The steps relate to the DS2 modules and other processes as follows:

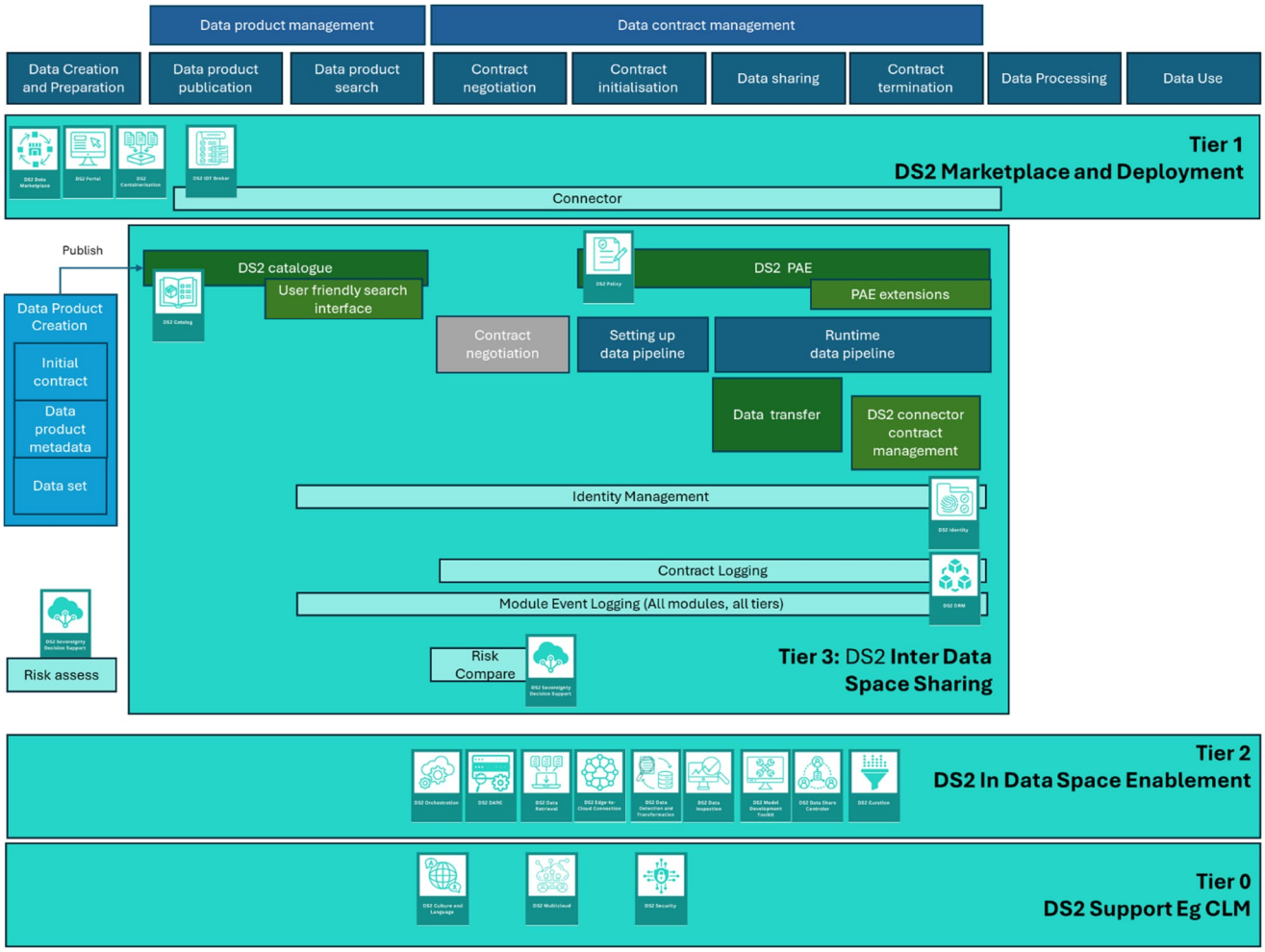


Figure 30. High Level Process Steps

9.5.5 Practical Sequence

What follows is ONE practical and demonstrative implementation scenario, but many other sequences are possible:

- Participant X (pX) of Dataspace A (DSA), orientated around the agriculture sector, would like to get information on pollution which they know is available from Participant pY in Dataspace DSB albeit that DSA and DSB, as well as pX and pY have different dataspace/participant rules and technologies
- They believe DS2 offers and approach to enable data sharing beyond their data space since they would like to use additional datasets. They start to experiment with the technology and in parallel trigger the authorities of DSA to contact DSB to see if a common or mutual policy environment can be supported
- DSA pX registers on DS2 (DS2 PORTAL) and signs-up to DS2 and its base IDT (DS2 IDT) download which includes the additional optional risk analysis Module (DS2 SDS) all in containerised packages (DS2 CONT)
- DSA pX installs and configures DS2 IDT and any selected Modules - in this case DS2 SDS which produces a risk assessment of the participants data sharing environment
- DSA pX Identifies potential participants with relevant data (DS2 CAT) and is provided with an end point of participants Y (END POINT Y) and it is noted that pY is already a DS2 adopters
- DSA pX Identifies its data pipeline and functions necessary to fulfil DSApX - DSBpY data sharing

- DSA pX realises they may need an additional Module to orchestrate the workflow (DS2 ORC) and acquires this new Module (DS2 PORTAL/IDT/CONT)
- DSA pX uses orchestration (ORC) to create a data pipeline from IDT and its connector to their backend application
- DSA pX triggers the use of IDT and its embedded connector to try to access END POINTS Y
- Since pX and pY do not know each other, then the receiving pY IDT Connector flags this as an unknown relationship and pY will only accept the request if pX can be identified which in turn means that DSA is identified and known to DSB and there is recognition of a DS-AB relationship
- pX receives the negative notification and enquires to the Authority of DSA if there is any progress on such recognition. In fact, this has just been granted
- pX configures their IDT connector with additional ID tokens for DSA itself and the DS-AB pair, as well as the existing DSA participant ID, and tries the request again (DS2 ID Triad)
- In the meantime, the identity authority of DSA has also established an interface to the DS2 IDM Module so that its participants can be validated
- DSA pX triggers the re-use of IDT and its embedded connector to try to access END POINTS Y
- DSA pY receives the DS2 ID-triad and the IDT Connector validates this via IDM which in terms uses the DSA identity authority for the individual participant validation.
- Next this triggers to process in the connector to forming the data sharing agreement
- Once the agreement is in place the data transfer can begin and the PAE functions [DS2 PAE] for policy enforcement
- DSA pY requires a risk analysis (SDS) from pX, and both provide their risk assessment reports to the SDS comparison functionality which then deliver a comparison report to both pX and pY. Both participants determine the risks are acceptable
- DSA pX and A/B execute the data sharing contract
- DSA pX monitors of policies during transfer (PAE) for potential (external) enforcement
- Data sharing between pX and pY ends
- DSA pX monitors Policies post transfer [PAE] – as above

10 MODULE DETAILS

10.1 Module Template

10.1.1 *Module Template Section Overview*

The Modules are documented in a standardised template whose scope and nuances are detailed below. The completed Module templates are located on the GitHub (https://github.com/ds2-eu) sorted per primary Tier. The ORC Orchestration Module, primarily a Tier 2 Module but can be used in Tier 3, is presented in Section 11.1 at the end and should be read in conjunction with this section in particular.

- Section x.0: Header:
    - Primary Owner: Most generally one beneficiary since this dramatically decrease issues connected with the joint development, geographically dispersed development as well as joint ownership/IPR. The only exception to this is IDT itself developed by ICE-ES and i4RI who have close working relationship
    - Tasks: DOA Tasks number and resource to which it relates
    - Tier: As documented in Section 7.1
    - Nature: As documented in Section 0
    - Results
    - DOA Text: In each Module, the verbatim DOA text is replicated and highlighted in blue is the part of the text which relates to that which the Module primarily covers. Comprehensive analysis has been performed before hand to ensure all DOA WP3 (Software), 4, 5, 6 texts are covered in some form. This has led in some cases to new explicit outcomes not originally envisaged such as the Data Marketplace DMK amongst others
- Section x.1.1: Introduction. Beyond the below the intent is that this information can be used to communicate more widely on DS2 Modules – for example on the website
    - Purpose: 4-6 lines as a general user-orientated guidance to the Module
    - Description: 10-15 lines delving deeper
- Section x.1.2.1-5: Big Picture. A variant of the Big-Picture diagram of the Technical Principles Section of this document (Section 6.2) is presented and then on it in red blocks is information where a Module is possible to operate. The picture is taken from IDSA and reformulated to also add similar concept in a DS-DS scenario - https://github.com/International-Data-Spaces-Association/IDS-RAM_4_0/blob/main/documentation/3_Layers_of_the_Reference_Architecture_Model/3_1_Business_Layer/3_1_2_Interaction_of_Roles.md. In some cases, these are multiple even if exclusive (eg ORC), ranging to other where various aspects of the same Module need to be deployed at consumer/provider participants and at the DS2 center (eg SDS). There after a short form table identifies which deployment scenarios are relevant and then for those relevant longer-form explainability text is provided. Further information on deployment scenario is in Section 9.2.4
- Section x.1.3: Component Definition. This is the heart of each Module template and is the pictorial and textural definition of the Module's subcomponents, internal flows, outwards interfaces, and APIs. This section in some cases includes additional material which helps explain the detail for more complex Modules. The graphic key is in the section below – Section 10.1.2 The primary principles are:
    - Adherence to the diagramming template of Section 0
    - 1:1 relationship between any block and narrative explaining that block

- Ensure that external blocks which are DS2 Modules have a corresponding block in the other Module
- Each arrow has a label and is somewhere covered to explain both directions of the control or data flow
- If there is Background, ensure it is clear and what is the advancement as well as indicating it in blue font in the diagram

- Section x.1.4: Technical Foundations and Background: Any Background used in the Module, along with license conditions, is highlighted here – this can range from standards through to 'product' or 'Research' which activity is based upon to accelerate the activity. There should be a 1:1 with Background indicated in the component diagram. This helps other developers and users who might intend to use a Module understand if created dependencies could cause issues – for example if a Background is a proprietary license
- Section x.1.5: Interaction of the components. To ensure cross Module dependencies are explicit. It should be 1:1 with the diagram except pervasive Modules (eg Dashbutton) are not strictly necessary. See Section 0 where the Module Interaction Matrix is generated from this table
- Section x.1.6: Technical Risks. The primary (2-4) non-generic risks of this Module. These were used to produce the consolidated Risk list of Section 10.3.1
- Section x.1.7: Security. The primary (2-4) non-generic Security issues of this Module. These were used to produce the consolidated Security list of Section 0
- Section x.1.8: Data Governance. The primary (2-4) non-generic Data Governance issues of this Module. These were used to produce the consolidated Data Governance list of Section 0
- Section x.1.9: Requirements and Functionality is a table indicating the relationship between a Module and the use cases. This is linked to the Part A requirements table. Noting the following:
  - Modules of type Optional (eg ORC) will be typically used in data processing and dataflow and hence are the focus of the tables. Each use case and subcase may typically only use certain Modules although DS2 has an aim that at least each Module should be in 2 of the 4 use cases with a target for these to be in the Sector cases UC1-3. The generic case, UC4, is a backup-plan in case this will not happen and as document will also help validate the Foundation, System Supporting Modules mentioned next
  - Modules of type Foundation, System, Supporting, (eg IDT, CONT) do not include explicit references to use case requirements unless there is an exception, since they are defacto. These Modules are not usecase specific and instead represent the foundation of the DS2 proposal and without them it would not be DS2 – hence, defacto. Beyond that the Modules are in general not specific to the data that is shared, nor the sharing flows so cannot be found in the requirements. For example, before the proposal, no user had a definitive requirement for IDT, just for a simpler way of cross sector data sharing and which they bought into the DS2/IDT concept to do this. Most, if not every, usecase will use these Modules
  - Due to the novelty and innovation of some Modules, as well as the parallel development of their architectures with the use case Part A document, in some cases it has been difficult to accurately a) State the specifics of how/and where the Module will work, and b) Provide users with sufficient education of the Module bearing in mind it was still being architected in parallel and often further research is necessary.
- Section x.1.10: Workflows: A series of 3-4 representative primary workflows for the Module including the possibility to cover both internal and external flows. This helps both users and technologists understand a how the 'static' architecture diagram may operate in practice

- Section x.1.11: Role, Resourcing and Milestones. This is a multifunctionality table with the following purposes orientated to control and manage development both for developing partners, hierarchic leads, dependent Modules/partners, and the entire consortium to feel progress:
    - To chunk the developments up in buckets of generally below 3 months for management control
    - To allocate these chunks into milestone orientated timeslots to track the development and achieve alignment with other Modules being developed
    - To assess against resource information to ensure developments are feasible against what was promised or allocated in the resourcing
- Section x.1.12: Open Issues. In some cases, not all questions could be answered since further conversation with other partners are needed or simply that further research work needs to be performed. The more significant ones are also listed in Section 11.1.12
- Section x.1.13/Annex: Other. For a few Modules supporting information may also be present – eg in the Portal information regarding the platform, and in IDT information about the Dashbutton to benefit others

10.1.2 *Architecture Diagram Key*

The key section to each Module is *.1.3 entitled "Component Definition" where an architecture view of each Module is given graphically and texturally portraying a Module's primary subcomponents, primary internal interfaces, outside-Module blocks, and appropriate API interfaces to them. The outside-Module blocks are broadly: Other DS2 Modules which are most generally optional but provide added-value, external software functionality eg connectors, or datastores, or represent human operators. The key to this is as per the diagram below although in some cases has been extended to represent a better pictorial description.

Further considerations:

- Generally, ORANGE has been used to represent the connector which is a kind of internal/external block. External since it is outside of the DS2 system, but many components are expected to extend them using its extension capabilities
- The blue data lines represent datasets from users which may be used and/or further processed. They do not represent data used in 'calls'
- For the black arrows related to control signals (ie not data plane) the arrows represent 'triggers' irrespective of if data is returned or not. Ie if subcomponent A trigger subcomponent B, which then returns information it is still a single headed arrow. If there is a functionality needed where B triggers A as well it would be two headed

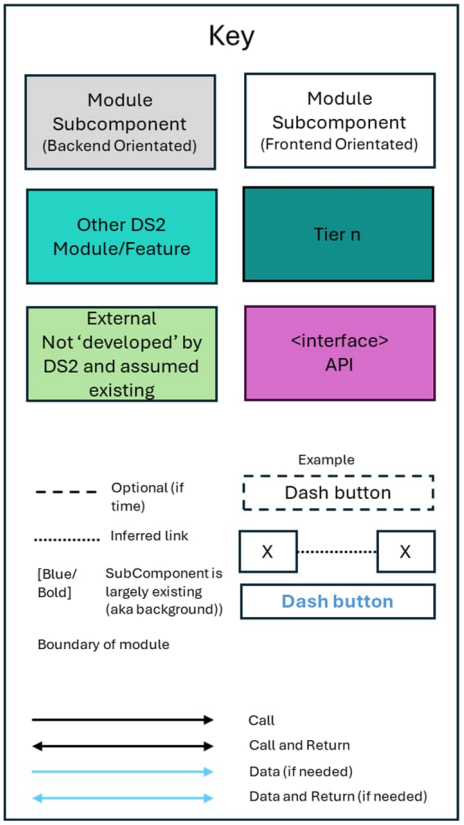


Figure 31. architecture view of each Module

10.2 Module UI Design Template

DS2 advertises itself as a holistic set of Modules and features of added value, and which is a relatively seamless and integrated experience for users. At a technical level this includes all Modules supporting IDT, using the Dashbutton features (See documentation of IDT Module on GitHub) etc but it also infers that that all Modules will be conformant, to the maximum extent possible, a common GUI template.

For Results created in the project this should be 100% achievable but many Modules will use existing libraries and other backgrounds which might not have separate presentation layer properties and hence the caveat of "where possible" although this should not be seen as 'excuse' for non-implementation.

DS2 is not a "UI design" project and with multiple parties and very different features it would be impossible to specify every element of UI design (Graphics, Tables, Graphics) so below is the primary UI scheme expected to be committed to and for other elements developers should consider approaches based on this schema in a 'best endeavours' way and provide credible reasons why this is not achievable. Best practice on this will be added to the UI guidance.

Presented below is style information re:

- Colour Pallet
- Dialog/Desktop sample
- Screen/Web sample
- From sample

Palette:

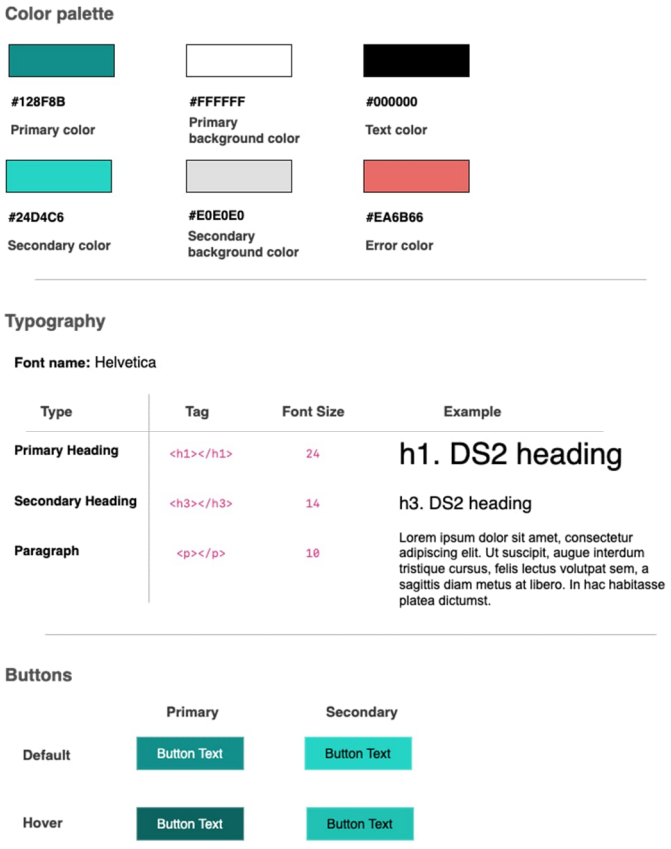


Figure 32. Palette

Dialog/Desktop Example

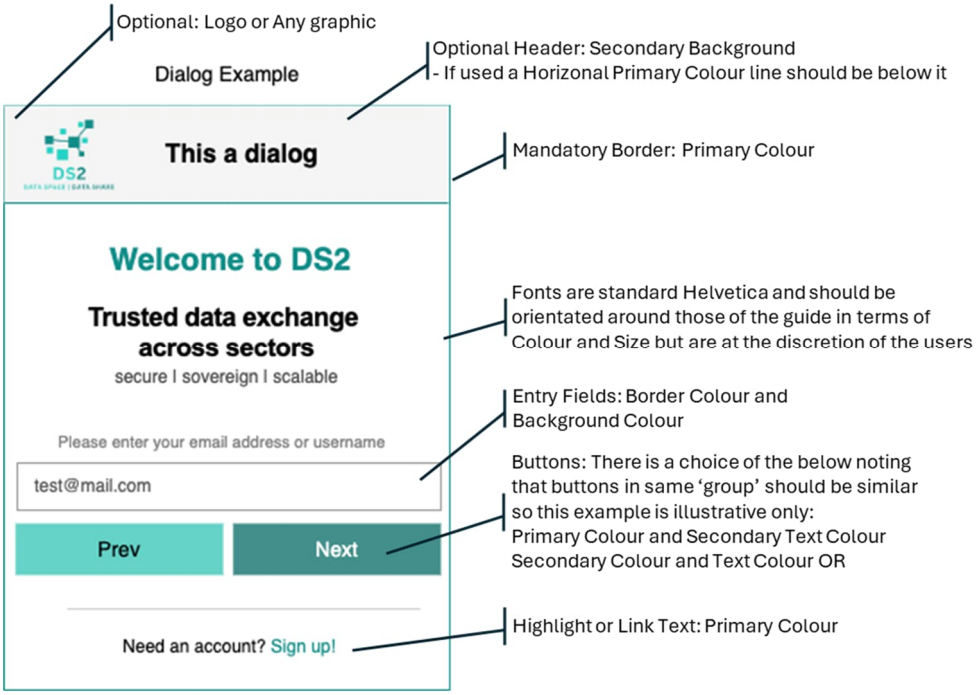


Figure 33. Dialog/Desktop Example

Screen/Web:

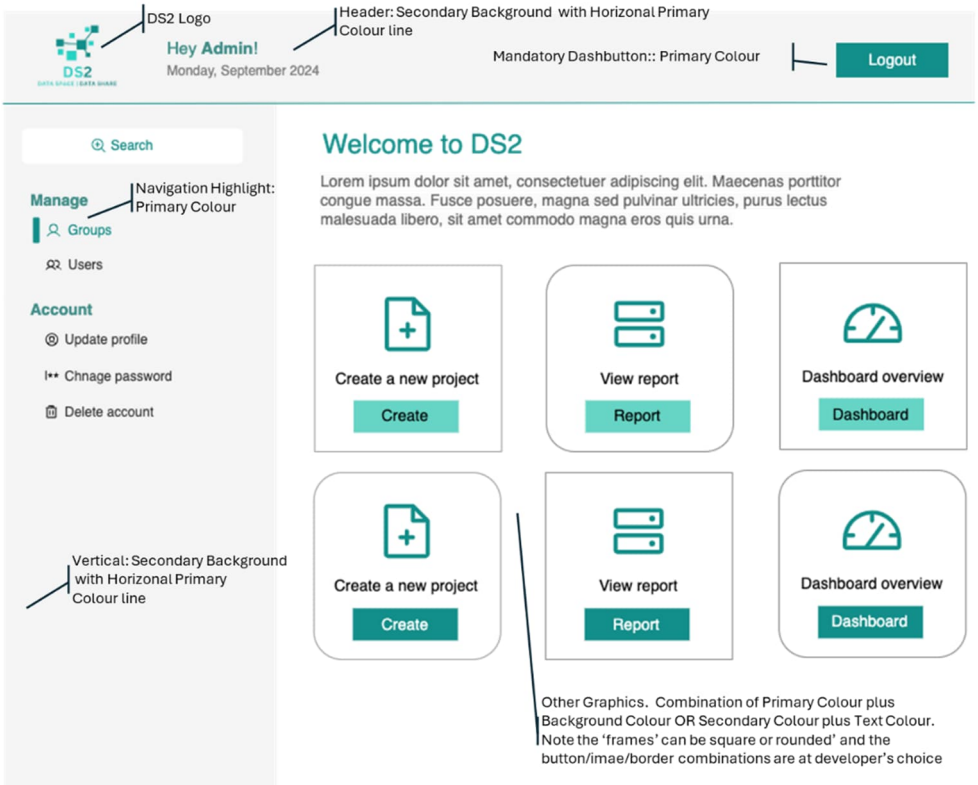


Figure 34. Screen Web Example

Web Form Example:

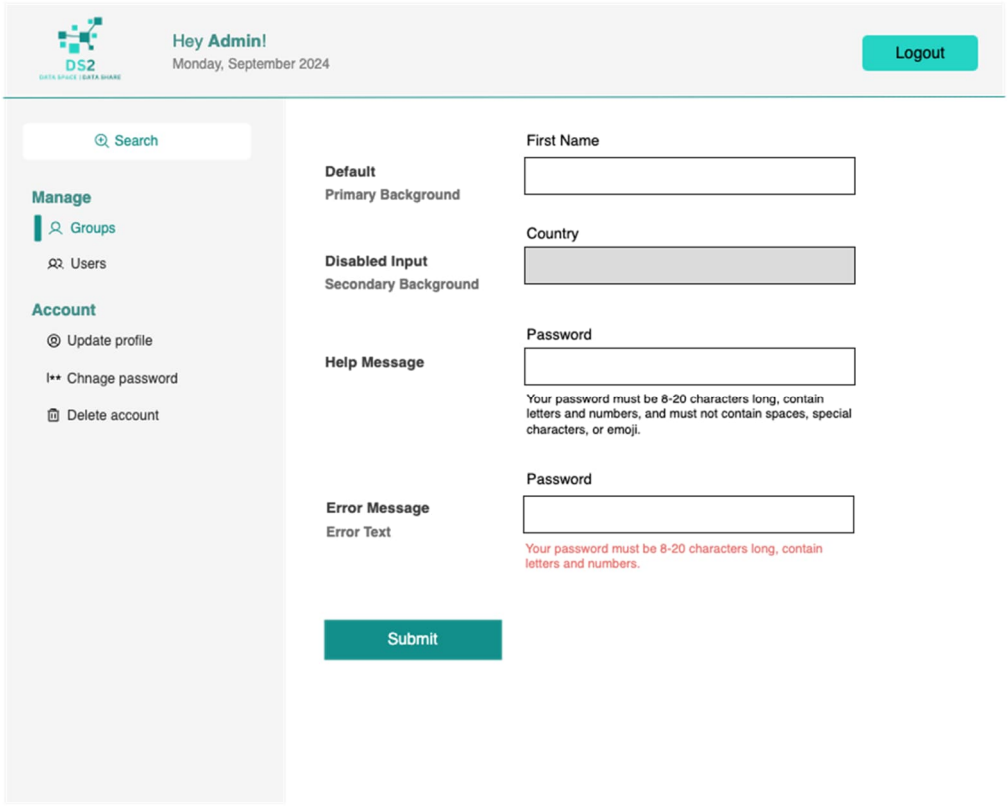


Figure 35. Web Form Example

10.3 Risks, Security, and Data Governance

Each Module template presents short tables of 2-4 primary Module risks, security concerns, and data governance for the specific Module as perceived by the developer. In terms of the risk element, developers were also requested to examine existing technical risk entries recorded under the risk management exercise of T1.2 and those recorded on Sygma. All mentioned entries for all categories have been extracted, synthesized, and consolidated into the relevant tables of the subsections below along with Modules which inferred an association with those consolidated entries. This gives a technical overview of elements which need to be controlled by the Task/Module lead and also WP lead and TM. In fact, many elements may infer an association with only some Modules but in reality, may be relevant to several or all. Where this is applicable the tables add 'ALL' in the Module column to help direct developers to a potential concert.

Beyond the mentioned lead monitoring actions and necessary developer attention; other actions:

- Risks: T1.2 Risk Record will be matched against and updated. Responsibility Technical Manager. It can be seen of particular concern is the DS Connectors – this is considered further in Section 10.4.2
- Security: No specific actions unless raised through lead hierarchy
- Data Governance: Consolidate list passed to DMP Own (VTT) for appropriate processing

10.3.1 *Risks*

Module	Risk	Description	Contingency Plan
SDS, CAT ALL	Integration and interaction with other Modules	Integration of heterogeneous Modules and systems brings some difficulty	Early definition of interfaces and integration tests to avoid later issues
DRM,	Complex Setup	DRM Architecture is complex for many Modules and can lead to challenges in setup	Carefully plan the architecture, remove unnecessary components, maintain a detailed documentation, and make use of automated deployment tools
DRM, MDT, CAT ALL	Scalability Concerns	Performance can be impacted as the system scales in terms of complexity, transaction volume number of nodes, etc.	Carefully plan the architecture, including the appropriate infrastructure, to optimize performance and scalability
DRM, CAT, MCL ALL (using connectors)	Integration with dataspace connectors - General	Modules may need to be integrated with the Dataspace Connector for different purposes. Each connector has its own technical requirements, and it must be ensured that all integrations comply with applicable specifications.	To ensure the successful integration of the DS2 uses a single EDC Connector in the IDT Module
PAE, DHARE, IDM ALL (using connectors)	Integration with non-production ready EDC Connector	The EDC Connector is an ongoing project, with the latest version currently at v0.8.0. Therefore, the interfaces may change during the project.	Design the solutions with the minimum number of dependencies to minimize the impact of potential changes in the EDC Connector's internals.
DHARE ALL (using connectors)	Novelty	Modules may be built from scratch and due to DS technology and in particular connectors being at their infancy there are expected to be difficulties intercepting control data particularly in Connectors	Operate outside of the connector as much as possible
IDT ALL (using connectors)	Choice of Connector	The choice of the connector in IDT (DS2) is a key decision and will have a significant impact on DS2 and all Modules if integration is too coupled.	Make sure the right connector is selected and try to decouple integration as much as possible with the integration architecture
ORC, DDT, CUR, E2C, DINS ALL	Changes to Data and Formats	Should the data change in form or function then the data model will need to be updated accordingly	Make sure the component is aware of changes to data. It will aim to raise a warning if the data has changed in format or schema. Try to use generic formats and unify
DARC ALL	Configuration of Modules	DARC based Configuration of Modules could be problematic if there is not much information for its language Module.	Make sure tool owners provide detailed information about Modules. Focus on Modules which have significant background and thus documentation (eg those from SWAG, ATC, DIGI)

CLM	No training or test data	Privacy or other constraints in providing data	Pre-definition of the concepts to be passed on after analysis, use of synthetic data
CAT ALL	Vendor Lock-In	Dependence on specific technologies or vendors may limit flexibility and increase costs in the long term.	Evaluate and select technologies based on open standards. Develop strategies for potential vendor transitions if needed.
CAT ALL	Technical Debt	Accumulation of technical debt due to rapid development and deployment cycles may hinder future enhancements and maintenance.	Allocate time for refactoring and code reviews. Prioritize addressing technical debt in the project roadmap.

Table 21: List of risks

10.3.2 *Security*

Module	Issue	Description	Need
SDS, DRM, ORC, PAE, MDT, CAT, DMK, DINS, DHARE, IDT, PORTAL, CONT, IDM ALL	Local Access Management: Authentication and Authorization	Managing secure and efficient authentication and authorization of users accessing the Modules and dataspace.	Local identity can be provided with the dash button, integration with DS2 IDM
DRM, ORC, CAT, DARC, IDT Broadly at least All Tier 3 Modules and soe at Tier 2	Inter-participant Trust	Establishing and maintaining trust between different participants across data spaces.	Implement secure authentication and authorization mechanisms. Utilize digital signatures and certificates to verify identities. Ensure secure communication channels (e.g., TLS/SSL). Regular audits and trust assessments. Integration with IDM
RET, CUR, IDT, CONT	Secure communication for data and metadata	Sensitive data and metadata will be exchanged between participants	Communication between participants and Modules needs to be secure.
CAT	In-Dataspace Security	Ensuring data integrity, confidentiality, and availability within a single data space.	Implement robust encryption for data at rest and in transit. Apply strict access controls and user permissions. Regularly update and patch systems. Monitor and log activities for anomaly detection.

Table 22: List of security issues

10.3.3 *Data Governance*

Module	Data Governance Issue	Description	Need
SDS, DRM, ORC, CLM, DMK, IDT, PORTAL	Managing sensitive information	Some types of data being managed by Modules may be confidential or sensitive information	Secure data store and transfer of that information. Comply with GDPR regarding storing contract and agreements
ORC, E2C, DINS	Test or Sample Datasets	Some Modules need to use test or sample datasets	A mechanism is needed so approved (or open) data samples can be shown. Synthetic data may also be used
ORC, DDT, MDT, CLM, DARC, DINS	Handling of personal data	Modules in general are not set up to deal with the monitoring of personal data but in some cases some personal data may be within the datasets being processed	User/Provider should ensure personal data transferred is transferred according to relevant regulations
RET, CAT, CLM	Metadata Catalogue	Module will access metadata associated with a requested dataset from the Metadata Catalogue	Published Metadata Catalogue information will be freely available to all entities that are allowed to access the Metadata Catalogue.
MDT, DINS	Access to Datasets	Some Modules require access to the datasets being managed	The data owner determines who can access the data.

Table 23: Data governance issues

10.4 Module Other Matters

10.4.1 *Changes to Results*

Within the DOA there were multiple technical KERs listed* along with some explicit expected outcomes albeit the latter were more general statements. In addition, inferred in the DOA text were some implicit technical outcomes* and as the project has evolved some related technical developments* which have been found necessary. Those marked * were listed in the outcome list of Section 8.1. Thus, this evolution has also meant some reconfiguration/additions /merges/removals. For transparency, the main ones are listed below with the colours referring to the outcome list table. In terms of many of the additions or outcomes many could, as mentioned, now be also listed as KERs.

- Addition:
    - IDM. It was found that a basic Identity Module would be necessary at the center of DS for knowledge of DataSpace IDs and Data Space pair IDs whilst still being reliant on Data Space identity providers for individual Participant IDs. This will be constructed by ICE as part of WP6
    - [DASH]. This is not a Module as such, but a common, easy-to-implement feature that allows users and developer navigation throughout DS2 installed features and is mandatory for all software outcomes to respect. In enables the coherent nature and navigation of DS2. It is documented within the IDT Module and will be built by ICE
- Reconfiguration, Rebranding, and Merging:

- INT: The Interoperability (KER 4.5a) is rebranded as the Data Retrieval (RET) Modules
- DARC: The proposal listed the conversational agent as being split between T5.2 (Discover and Assess) and T5.3 (Recommend and Configure) with two related KERS K5.2 and K5.3. As these topics have been researched, it is found that it is better to manage the development as 1 unit, ie effectively merging these tasks both dominated by ATC to produce the DARC Module with exactly the same scope.
- PORTAL & [PLATFORM]. KER6.4 is listed as "DS2 IDT Integrated Broker, Modules & Toolkit" which is a little vague and "uncatchy" and moreover not directly representing the intent. This has been rebranded as the "DS2 Portal" (KER 6.4a) of which one element is the "DS2 [Platform] (KER6.4b). Both are not 'normal' Modules but are focal points for accessing and deploying Modules. [PLATFORM] is documented within PORTAL.
- MDT and DDT: These relate to the T4.2 task on AI Anomaly detection and were original one KER T4.2 with development by SWAG, INDRA, and DIGI. Due to this the work has been split with the original KER maintained re the SWAG DDT Module, INDRA making a new outcome MDT, and DIGI perform library work for both as detailed next which also reduces the issues around joint development
- [AILIB]: The AI Detection libraries were inferred but are inappropriate to be a Module. They will be used, and are mentioned, in both the DDT and MDT Modules related to the same area
- The Data Quality Module, which was an inferred outcome and is now merged into the E2C Module (KER6.2) within the same task. Both activities are allocated to SWAG in the same task T6.2. The rationale due to some overlaps with T4.2 and the anomaly detection Modules/KERs and potential duplication), and the need to orientate itself to T6.2 activity only

- Removal:
  - The Portability KER4.4 related to T4.3 has no specific Module and is more of a generic (one word) statement about the needs for portability across the realm of DS2. In reality DS2 addresses portability at many levels eg System portability through containerisation and CONT, data portability through the new Module RET

10.4.2 Connectors

The key software, which is most relevant to several DS2 Modules, and in particular IDT to achieve the networking affect, is the Connector which is the entry and exit point to participants both in the control and data plane primarily between Consumer and Providers. However, other challenges connected with this is that because there are no standards or defacto implementations, connector-connector interoperability is all but invisible and the interconnection of connectors to other applications (eg DS2 Modules) is difficult.

To reduce this interoperability issue in the scope of DS2 a further selection was made to use the EDC (Eclipse Data Connector) which appears to be the most stable, (experimentally) implemented, and (lightly) supported. To enable some DS2 functionalities an IDT will enable the extension of a connector using formal mechanisms where possible. Indeed, EDC welcomely has these, but they are extremely limited, and the alternative is branching code (with a later view to integrate to main branch once validated)

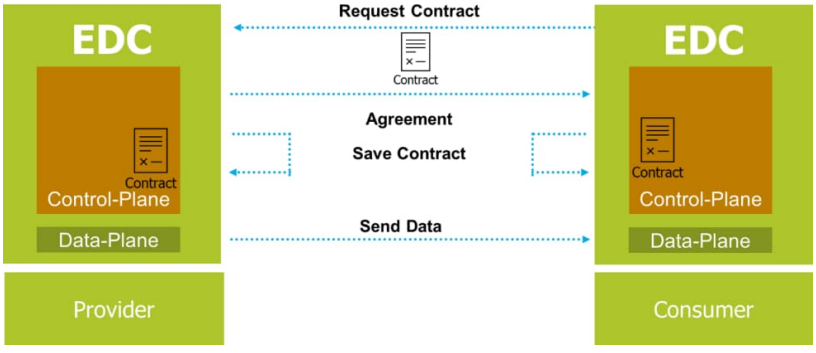


Figure 36. EDC (Eclipse Data Connector)

which extremely complicated because of number of dependencies and the lack of documentation. A final difficulty is that there are few (around 5) external developers seriously committing code to which means the level of support and interaction is low even if better than most.

The exploration of possibilities has been ongoing by several partners during this phase of the project and as of now EDC still appears to be the best choice. However, another connector called "Tractus X" which advertises itself as "Tractus-X is the official open-source project in the Catena-X data space" with Catena-X being an implementation in the Automotive sector although it is unclear if live production changes are taking place. Still, it is clear it is more trialled and documented and with further research could take the place of EDC especially considering it is theoretically based on EDC. In addition, Catena-X has also implemented so call Kits, which might be in essence similar to DS2 Modules, and could allow partners to gain further knowledge on extending EDC/Tractus-X Connector.

Next steps:

- Modules to further specify their technical needs on connectors
- Partners to explore EDC based on technical needs and issues around extendibility/branching
- DS2 partner to pool knowledge
- ICE as TM (possibly others) to explore Tractus-X and Kits use

10.4.3 *Experimentation Facility VTT*

DS2 is committed to provide an experimental data space whose infrastructure can be used as a baseline for developing customized data space solutions for specific application domains or use cases where appropriate to DS2 technical parties and users. This allows the testing of business concepts and building of proof of concepts before investing in real-life solutions. It is implemented via VTT's Data Space Innovation Lab (DSIL).

DSIL is a state-of-the-art experimentation facility designed to support the development, integration, and enhancement of data spaces. DSIL acts as a sandbox environment that fosters innovation and experimentation within the realm of secure, scalable, and interoperable data spaces. It provides a versatile platform equipped with cutting-edge technologies that facilitate the onboarding, management, and scaling of data space solutions. Within this lab environment, a variety of reference components, tools, and services are deployed, offering comprehensive support for the data space lifecycle.

At the heart of the DSIL are several key components that ensure a complete and compliant data space infrastructure, based on the standards and principles established by the International Data Spaces Association (IDSA). These components include:

- IDSA Data Space Reference Components: The DSIL includes a full suite of IDSA-compliant reference components, ensuring that the data space environment adheres to the highest standards of security, interoperability, and data sovereignty. These components are critical for the provision and consumption of assets within the data space and include:
    - Connector: A secure and trusted endpoint for data exchange between participants within the data space. It ensures compliance with IDSA's architecture and policies
    - Metadata Broker: Facilitates the discovery of data assets by acting as a metadata catalog, enabling participants to find and access available datasets within the dataspace
    - Dynamic Attribute Provisioning Service (DAPS): Manages and validates the digital identities and access control policies of data space participants to ensure secure and trusted exchanges
- Eclipse Data Space Components: In addition to the IDSA reference architecture, the DSIL incorporates Eclipse Foundation's open-source data space components. These provide a flexible framework for integrating open-source solutions into the data space ecosystem, facilitating interoperability and enabling collaboration between different stakeholders

- Deployment and Infrastructure: DSIL operates on a highly scalable and dynamic infrastructure powered by Kubernetes. This allows for seamless orchestration, management, and scaling of microservices, ensuring that the various components of the data space can be deployed, maintained, and updated with minimal overhead
- Kubernetes-Orchestrated Environment:
    - Scalability: Kubernetes automates the deployment, scaling, and management of containerized applications, enabling the DSIL to manage increased loads and complexity as new participants and assets are added
    - Microservices Architecture: Each data space component—whether it be the IDSA Connector, Broker, DAPS, or other supporting tools—operates as a microservice, ensuring modularity and flexibility. This architecture allows individual services to be scaled independently, leading to more efficient resource management

The entire data space infrastructure is accessible through REST APIs, which provide a unified interface for provisioning, managing, and consuming data assets. This API layer abstracts the complexity of the underlying components, allowing developers and participants to interact with the data space programmatically. They allow participants to provision assets, manage data flows, and consume services with ease. The APIs are designed to be:

- Standards-based: Fully compliant with IDSA standards to ensure that the data exchanged within the space adheres to strict policies around security and data sovereignty
- Extensible: Developers can build custom applications on top of the DSIL infrastructure, leveraging the APIs to create new services or integrate existing systems into the data space ecosystem

Each of the core services—such as the IDSA Connector, Broker, and DAPS—has its own dedicated API endpoint, allowing participants to interact with each service individually or as part of an integrated workflow.

From a user perspective DSIL's experimentation space is designed to support a wide range of data space use cases, including:

- Data Sharing and Sovereignty: The facility allows participants to experiment with secure data sharing solutions, maintaining full control and sovereignty over their data assets
- Cross-Sector Collaboration: By providing an open, interoperable environment, the DSIL enables participants from different sectors (e.g., manufacturing, healthcare, finance) to collaborate and exchange data securely
- Onboarding New Participants: The DSIL facilitates the seamless onboarding of new participants into the data space ecosystem, providing tools and services that simplify the process of integrating new data sources and services

Thus, in conclusion, DSIL serves as a comprehensive, innovation-driven facility designed to support the full lifecycle of data spaces. By integrating a range of reference components, microservices, and tools, the DSIL offers a flexible, scalable, and secure environment for experimenting with and enhancing data spaces.

Key features such as Kubernetes-based orchestration, open-source Eclipse components, and REST API-driven provisioning make the DSIL an ideal space for testing, innovating, and scaling data-driven solutions. As we continue to enhance the facility, we look forward to expanding its capabilities and driving further innovation in the field of data spaces.

11 DS2 ORCHESTRATION MODULE EXAMPLE (ORC)

11.1 DS2 Orchestration Module (ORC)

Owner(s): ICE
DOA Task: T4.1
Tier: Tier 2
Nature: Optional
Result: K4.1

This task will create a federation mechanism to enable different data spaces to interoperate. This task will orchestrate the lifecycle of data from data collection to data exchange, to data disposal/deletion across a federation of distributed data stores. The data lifecycle will include the establishment of a data contract between the data sources, the establishment of trust between entities in the data flow, and the adherence to data sovereignty and security requirements in the resulting federated data set. Other challenges will be addressed, such as accountability for use of purpose, the propagation of new domain-specific data restrictions (such as policy changes) across the federation, and methods for non-repudiable lineage across the lifecycle. Additionally, topics to reduce the latency involved in the transfer of huge amounts of data, such as caching, and data relocation will be investigated. This task will examine current, emerging technologies in this field, such as work being led by European Data Spaces and from the GAIA-X project as a basis for extension.

11.1.1 Introduction

Purpose: To design and then orchestrate at runtime In-Dataspace, Inter-Dataspace, internal, and third-party services which facilitate common data-orientated operations such as transformation of data, checks on data, data updates etc. The orchestrator contains a flexible GUI to design workflows and decision points on these services and run time component to implement the workflow.

Description: Services are added to and then selected from a Service catalog from a participant's local service catalog (In-Dataspace deployment), the DS2 service intermediary catalog (Inner-Dataspace), or other available catalog/service knowledge. These services can be graphically linked together to form a workflow and where decision pathways, decision points, and other operators can be deployed to determine the workflow. Error and exit points should be predetermined with defaults ensuring that failures and error conditions allow flows to be closed automatically. One class of operator is for user defined forms for human input but most often the flows contain backend services. In the context of DS2 the operators will, if necessary, be expanded based on novel usecase peculiarities as will the forms designer. Primarily the design interface is orientated around service interconnectivity, but this will be augmented with a data pre-viewer to help interconnect and understand the results of interconnecting data-orientated services. The orchestrator will be available as a Module and for interparticipant service orchestration will extend the connectors.

11.1.2 Introduction

11.1.2.1 Introduction

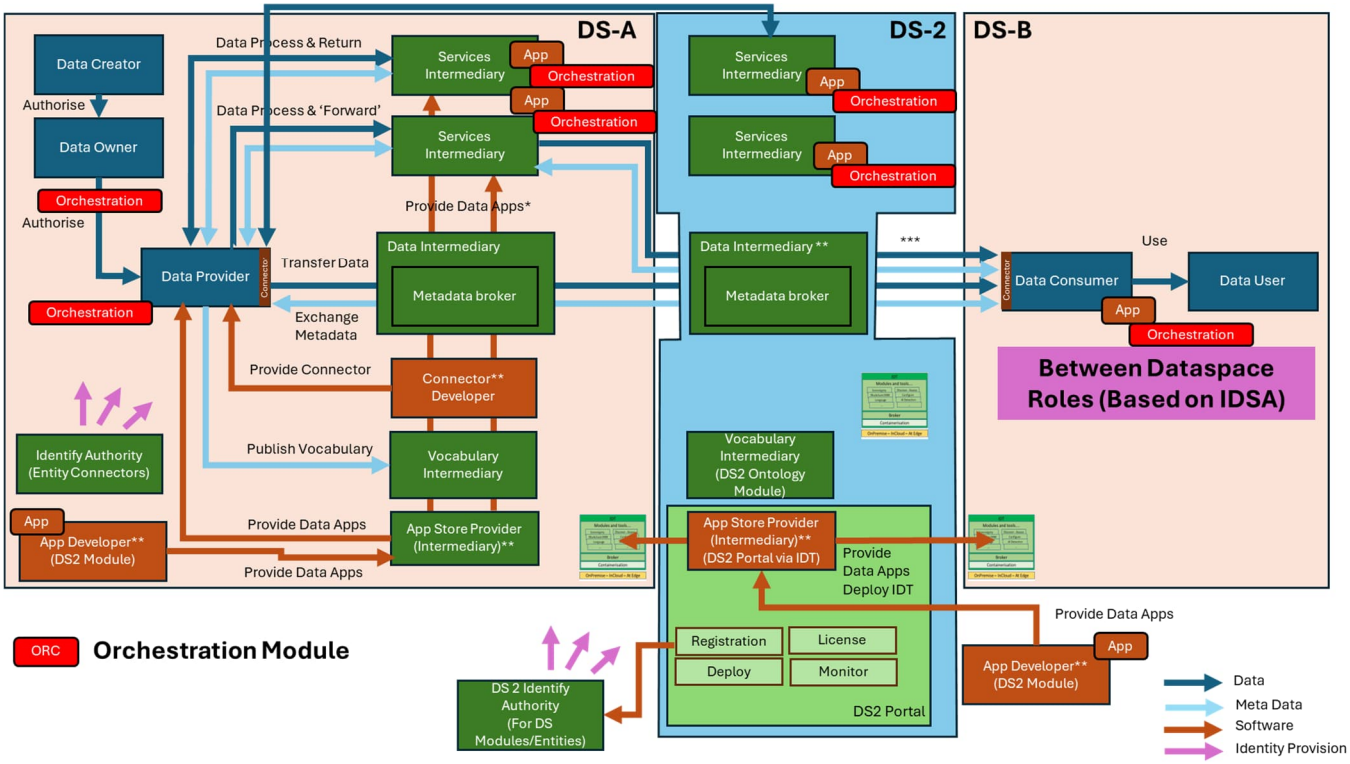


Figure 37. Big Picture

Where	Status
Within a single Dataspace for use between participants in that Dataspace only	Could. This is not piloted/validate in this project but in essence an in-Dataspace version would work similar to the two Across Dataspace scenarios without the need for more complex identity checking
Deployed and used by a single participant to enable the participant in either an In-Data space or Inter- Data space scenario	Yes: NB can still orchestrate external services (eg non-dataspace in other dataspace participants) if user requires and at users (probably necessary) choice can implement DS2 trust mechanisms for service action
Across Dataspaces without Service Intermediary	Yes :Would then mean orchestration Modules "liaise" and interoperate with each other
Across Dataspace with Intermediary	Yes. Potentially could be performed by a Service Intermediary [not covered in DS2
Other Comments	Choice dependent on user needs and they are not mutually exclusive

Table 24: Module placement description

11.1.2.2 Within a single Dataspace (where applicable)

N/A

11.1.2.3 Deployed and used by a single participant (where applicable)

To move data from provider to consumer invariably involves a set of processes. From a Consumer perspective this could involve acquiring via the data in the first place, to navigate it through further data-orientated or movement processes (eg transformation, fusion, filtering), to placing it in their own data store. At the Provider side the processes are similar but in reverse. In this scenario the orchestrator will facilitate this at

either consumer or provider independently since it assists a participant in efficiently creating such a process flow connecting data and other services, as necessary. This thus helps enable their part in the dataspace.

11.1.2.4 Across Dataspaces without Service Intermediary (where applicable)

Whilst individual data provider/consumer orchestrations can add value, there may be further opportunities to link process engines already established at either end. For example, a provider may have a set of processes which then fill a staging database and then when those processes are done trigger a orchestration engine at the consumer end to begin a process which extract the data and populates their own database. This allows control by both parties yet still allow for sharing processing operations. The processes (services) need to be selected from each participants' service catalogs, so the participant/services need to be known in advance. Note that the process between participants in the same dataspace or differing dataspaces is the same but will be reliant on the use of the DS2 IDT connector.

11.1.2.5 Across Dataspace with Dataspace Intermediary (where applicable)

A dataspace service intermediary can be used host the service catalog of consumer and provider which allows orchestration to be made without especially knowing the participant in advance. The process will operate similar to the use of a metadata broker.

11.1.3 Component Definition

The figure below represents the actors, internal structure, primary sub-components, primary DS2 Module interfaces, and primary other interfaces of the Module.

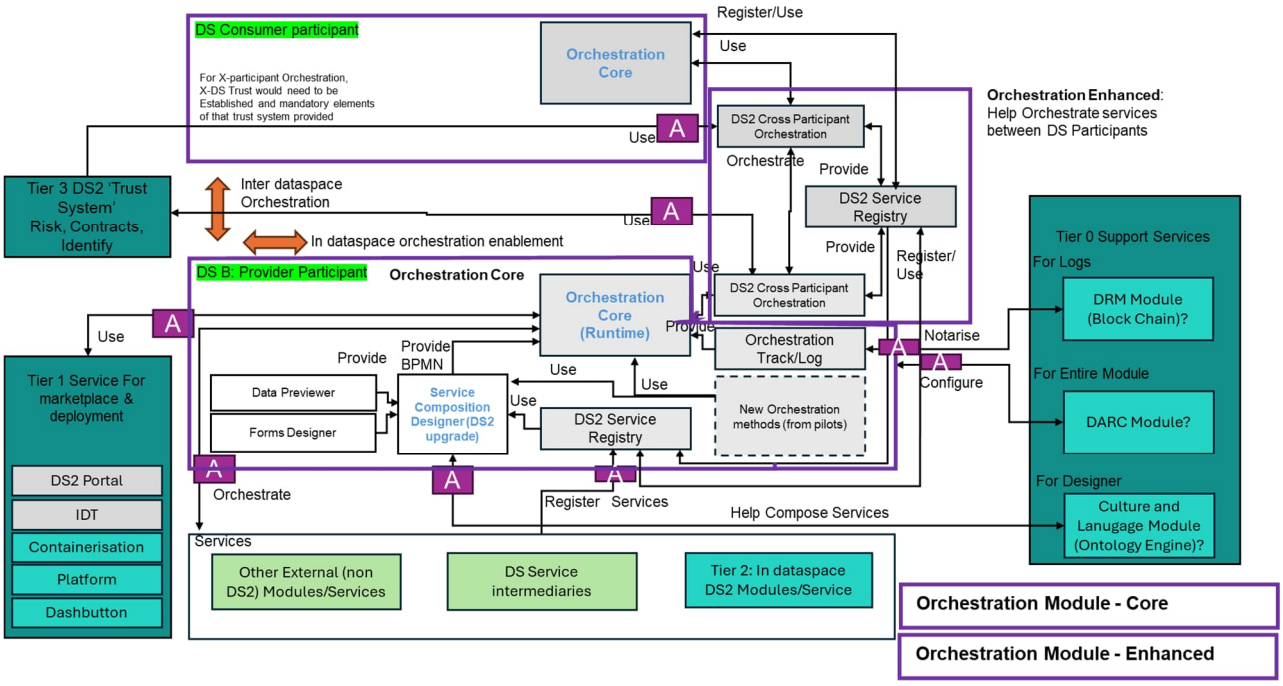


Figure 38. Schema for the Module

This Module has the following subcomponent and other functions:

Orchestration Module – Core:

- Orchestration Core: This is the runtime heart of orchestration which conducts a process (workflow) triggering other services and via a BPMN Module from the Service Composition designer repository. For the tier 1 standard connections (Portal etc) it can be perceived as the entry point. If new orchestration design methods are needed it will use them. Runtime events are connected to the logging components and for inter-participant dataflows it will interface with the DS2 Cross Participant

Orchestration subcomponent. This is currently ICE background and will see little development except for a DS2 compliant UI.

- DS2 Service Registry:
  - In-participant: This is a local registry of all services which a participant may potentially use in a workflow, composed together in the designer, and executed in the runtime. Registration can be automatic in the case of IDT installed services. It will also expose services in the inter-participant DS2 registry. The subcomponent exists now but will be rebuilt in the context of DS2 and IDT.
  - Between Participant: 95% the same functionality but can function similarly to a metadata broker to host services from multiple participants which can be shared in a controlled way to the In-participant registry to allow participant-participant service interactions
- Service Composition Designer (DS2 Upgrade): This is the main UI for the Orchestration Designer based on existing ICE background. It allows a user to select or drag various elements from a toolbox (services/APIs from the DS2 Service Registry, methods), which can be placed on a canvas where they can then begin to start designing their orchestration by dragging and connecting various elements together. The saved BPMN2.0 notation model will then be used by the runtime orchestration core. The DS2 upgrade will be mainly for UI and inclusion of New Data Previewer and Forms designer blocks
- Forms designer: Many orchestrations have a need for user input and whilst some might come from other systems this can be complex when only limited information. The forms designer will allow the easy inclusion of simple form in any Service composition and also ensure that it respect the data flow as well as service needs
- Data Previewer: This is also a new subcomponent which will be rendered via the Service composition designer. Currently services are connected but when designing it is useful to know at design time what might be the inputs and the expect result. In a data orientated project this is especially useful, and this utility will allow some rendering of data to help show flow operations between building blocks before they are deployed
- New Orchestration methods (from pilots): Many methods – eg choice boxes, selections are already implemented in the orchestrator, but it is possible that the pilot might suggest further ones that could be interesting to implement – although at this stage of the analysis it seems there is not. The new methods will be exposed in the orchestrator runtime & designer
- Orchestration track/log: Currently this is rudimentary and especially in the trustworthy context of dataspace an major overhaul is necessary to extract more granular logging information at runtime.
- Services and API: These are the services that can be orchestrated, and the API block is the interface to:
  - Other External (non DS2) Modules/Services:
  - DS Service Intermediaries:
  - Tier 2 In dataspace DS Modules/Service:
- Tier 0 Support Service Stack:
  - DRM and API: For further exploration, but if room to implement and a match of requirements to feature the blockchain part of the DRM Module to enhance logging
  - DARC & API: As with DRM but in this case to use DARC to help configuration of the Module
  - Culture and Language Module and API: As with DRM but in this case to use this Modules ontology engine to help auto-link services
- Tier 1 Service Stack for Marketplace and deployment and API: The full stack will be implemented as generically described elsewhere in this document. Exceptions: The Platform will only be needed for inter-participant service orchestrations if used

Inter-Participant orientated:

- DS2 Cross Participant Orchestration: This is a new runtime Module which will act as a bridge between the orchestration within each participant through interconnections to the Inter-Participant Service Registry and the Orchestration core at each participant
- DS2 Service Registry: As described above
- Tier 3 Trust Stack and API: For interparticipant service the Module will use relevant parts of the DS2 trust stack – see diagram below

The component for cross participant flows will fit into the IDT/Connector architecture as follows:

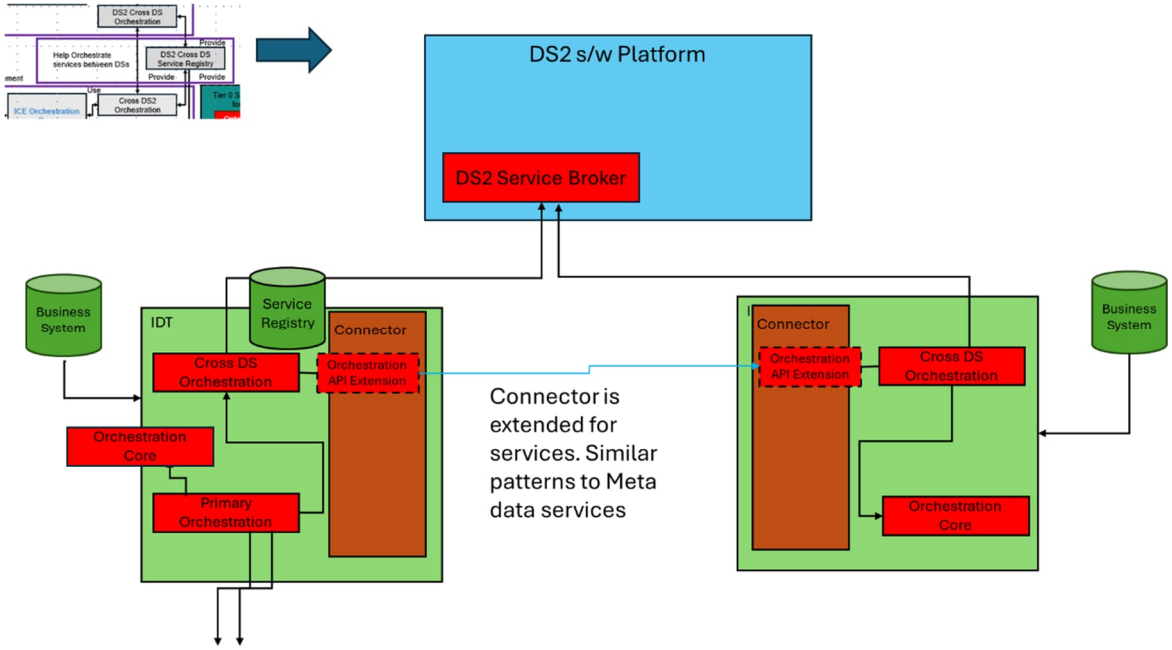


Figure 39. IDT Fit

11.1.4 Technical Foundations and Background

The foundations of the Orchestration Designer make use of an open-source BPMN modellers framework BPMN-IO, which is a rendering toolkit and web modeller for BPMN. It allows easy creation of BPMN2.0 diagrams using a web-based modelling library, which can be extended to add the functionality for DS2. The toolbox inside the Process List Manager offers DS2 Modules (specifically those that can be run as a service) that are available from the portal. It contains all the standard BPMN modelling elements, such as Parallel or Exclusive Gateways, Processes, and many others that are needed to create a BPMN diagram. The BPMN2.0 Rendering Service can render the diagram whereas the Orchestration Designer can save, retrieve, and annotate a BPMN process as well as deploy it for execution. The execution functionalities are made possible by accessing the public interfaces from the called services. Other relevant ones are Web technologies such as Angular2, NodeJS, JavaScript ES6 and HTML5. The participant-participant connector expects to upgrade the EDC Connector used as a basis within IDT.

Subcomponent/Component	Owner	License
Wasp	ICE	Apache 2.0

11.1.5 Interaction of the Component

The following table specifies the primary input/output controls/data to blocks which are not part of the Module

T	Receives From/ Gives To	What
Portal	Gives To	When published on the portal information (technical, how to's etc etc) will be provided according to the general model
Tier 3	Receives From	Authenticity of participant information in participant-participant scenario
Tier 3	Gives To	Participant identity
Services	Receives From	Service End Point information of Other external Modules/Services. DS2 Service Intermediaries, DS2 Modules/Services
Services	Gives To	Run time triggering of end point APIs
DRM	Gives To	Log/Control information for notarisation
DRM	Receives From	Log/Control information from notarisation
DARC	Receives From	Configuration information
DARC	Receives From	Interactive requests for configuration information and information of how to auto configure
Culture and Language	Gives To	Name and descriptive information from services and their bindings
Culture and Language	Receives From	Ontological alternative suggestions of the given information

Table 25: Primary input/output controls/data to blocks which are not part of the Module

11.1.6 Technical Risks

Risk	Description	Contingency Plan
Changes to Data	Should the data change in form or function then the data model will need to be updated accordingly	Make sure the component is aware of changes to data. It will aim to raise a warning if the data has changed in format or schema

Table 26: Technical risks

11.1.7 Security

Security Issue	Description	Need
Inter-participant trust	Potentially services could be intercepted	Strong relationship with identity system since the service requests may operate outside a connector
In-Dataspace	Where the orchestrator is deployed solely in at a participant there is no specific DS Security risk	N/A

Table 27: Security issues

11.1.8 Data Governance

Data Governance Issue	Description	Need
Data Previewer	By-design the data previewer subcomponent will show samples of the data (eg before and after transformation) to help compose workflows	A mechanism is needed so approved (or open) data samples can be shown
Handling of business process	Business processes may represent confidential information	Secure data transfer of the business process information and of the users instantiating the orchestration
Handling of personal data	This component is not set up to deal with the monitoring of personal data	User/Provider should ensure personal data transferred is transferred according to relevant regulations

Table 28: Data governance issues

11.1.9 Requirements and Functionality

This Module will be used in the following usecases:

City Scape	Y
Green Deal	Y
Agriculture	Y
Inter-Sector	TBD

The requirements and functions/extensions to achieve them relative to this Module, specifically extracted from the use case are as per the table below noting that in many cases further discussion might need to take place between pilot partner and Module partner to determine if a fit or the scope of the precise fit:

WHERE	WHAT	WHY	Run/Design Time	Priority
	Use Case 1: City Scape			
Section 2.2 UC1.1	N/A		R & D	M
Section 2.2 UC1.1	N/A		R & D	M
Section 2.2 UC1.1	N/A		R & D	M
Section 2.2 UC1.1	"Sharing and gathering data from multiple sources and sectors"	To orchestrate the data from the sources to one location	R & D	M
	Use Case 2: Green Deal			
Section 2.2 UC2.1	"Relevant data sources to be obtained from both data spaces within the use case. "	To orchestrate the data from the sources to one location	R & D	M
Section 2.2 UC2.2	"Relevant data sources to be obtained from both data spaces within the use case. "	To orchestrate the data from the sources to one location	R & D	M
Section 2.2 UC2.3	N/A		R & D	M
Section 2.2 UC2.4	N/A		R & D	M
Section 2.2 UC2.5	N/A		R & D	M

	Use Case 3: Agriculture			
Section 1.1.4	"Data Integration and Accessibility"	"Facilitating seamless integration"	R & D	M
Section 2.2 UC3.1	The section 2.1 is quite generic but in general involves the movement of data between different parties and processes	To achieve efficiency	R & D	M
Section 2.2 UC3.2	The engagement of orchestration is similar to UC3.1 except the domain is different – from fruit sourcing to forecasting/crop management	To achieve efficiency	R & D	M
Section 2.2 UC3.3	This is also the same except it is based on Crop Productivity	To achieve efficiency	R & D	M

Table 29: Module use in use cases

11.1.10 Workflows

The following sub-sections describe the sequence diagrams of the Module.

11.1.10.1 Register New Service (Local and Intermediary)

This feature provides the capability to register a new Service API in the local service registry to then be able to use it in a new flow. This also allows to create a new remote service registry and synchronize the local and remote services by updating the local one with the remote services and publishing the local services to the remote registry. The figures below show the sequence diagram of this feature.

The main steps/functionalities are as follows:

- Fill in New Service Details
- Crete and Store New Service
- Fill in New Remote Registry
- Create and Store New Registry
- Sync Local and Remote Registries

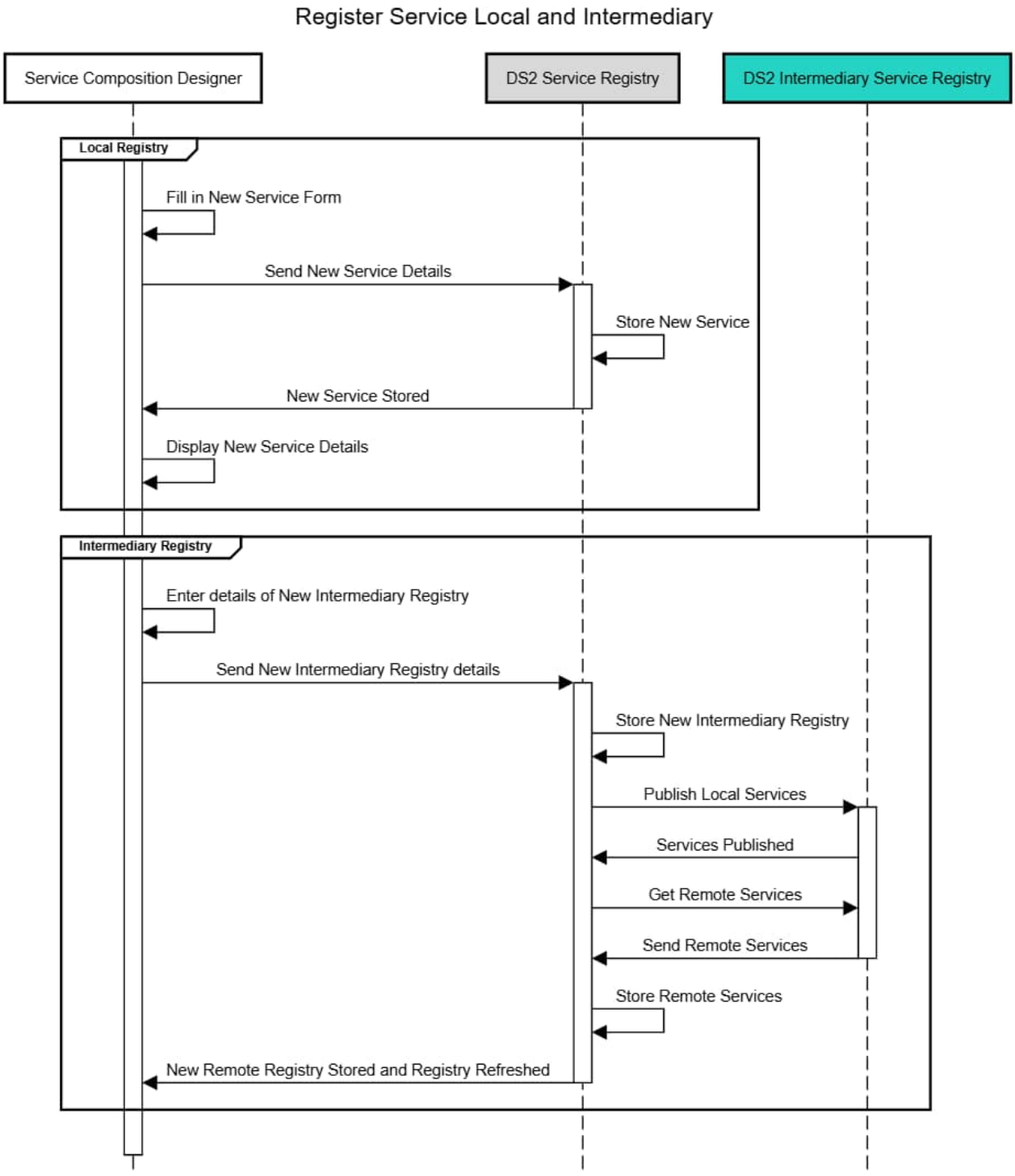


Figure 40. Register New Service Local and Intermediary Schema for the Module

11.1.10.2 Design and Run New Flow (Process)

This feature provides the capability to create a new Flow or BPMN Process by composing a number of services from the Local or Remote Service Registry. This allows to also use the Form Designer and the Data Previsualisation tools. Figure 1 shows the sequence diagram of this feature.

The main steps/functionalities are as follows:

- Get List of Services

- Design Flow
- Design Form
- Run a Data Preview
- Store Flow
- Run Flow

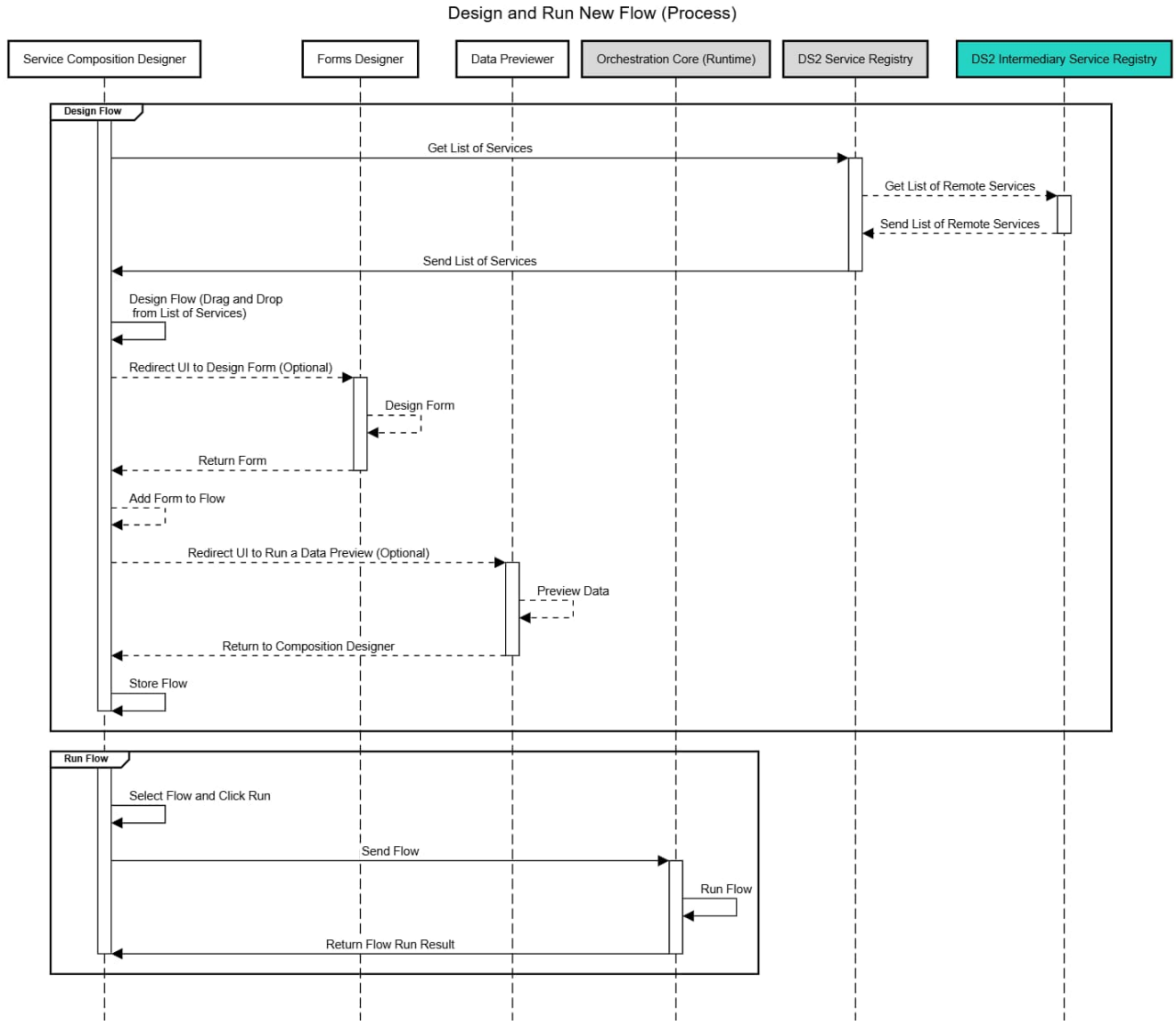


Figure 41. Design and Run Flow (Process)

11.1.10.3 Interaction with the IDT Connector

Regarding the interaction with the Connector the flow is exactly the same as the previous ones. The Connector API will be pre-registered in the Orchestration local service registry and there will be some pre-defined flows to use it. In order to design and run new flows, the sequence diagram in Section 11.1.10.2 can be used.

11.1.10.4 Participant to Participant Interaction

Regarding the interaction between participants, this will depend on the services exposed in the Intermediary Service Registry. A participant will be able to design a flow using services from the other participant. For this, again, the sequence diagram in Section 11.1.10.2 can be used.

11.1.11 Role, Resourcing, and Milestones

Sub-component	Main Activity	M18	M24	M30	M36
Orchestration Core (Runtime)	Largely upgrade to look feel				
DS2 Service Registry	Ensuring support of Modules and data orientated services				
Service Composition Designer (DS2 Upgrade):	Upgrade for new types of services (Modules)				
Forms designer upgrade	Upgrade to respect form approaches perceived in data vs service orientated environment				
Data Previewer	Full Module development				
New Orchestration methods orientated to data flow (from pilots)	Indeterminate until user flows are known				
Orchestration track/log	Significant upgrade to also match data space orientated log expectations				
Blockchain & Blockchain API:	Understanding and implement DS2 block chain Module				
DARC & DACR API:	X – Indeterminate – depends on selection of test Module				
Ontology Engine & Ontology API:	Further time to explore if a working ontology Module could advance ICE orchestration				
Dashbutton	Integration				
Tier 3 Trust Stack & Tier 3 API:	This will be key and new work and new understanding on how to innovate/develop connectors for cross-participant service orchestration				
DS2 Registry Service:	Creation of a communal registry record				
DS2 Cross DS Orchestration	Cross party orchestration by connecting orchestrators at each participant and by utilising IDT				
Table Total/DOA Task Total/Resilience	Comments: Blockchain, Ontology are sacrificial. Anticipated no new methods are needed				

Table 30: Role, resourcing and milestones

11.1.12 Open Issues

The following table summarise open issues/uncertainties that need to be resolved during the next stages or implementation.

Issue	Description	Next Steps	Lead or Related Component
Support Module interfaces	Final design and at least some Implemented of block chain/DARC/Ontology Modules. These are not key/blockers but if there is an opportunity it could be useful for the Module	Wait for final design/implementations	WP3: Block Chain & WP5: DARC and Ontology
Connector	From theoryàresearchàdesignàimplementation of upgrades necessary to chosen connector	Further connector research	None
Usecase	Precision is necessary on the flows to see if any additional forms, service methods are need	Discuss with usecases	ICE and all usecase

Table 31: Open issues for the module

12 DEVELOPMENT CONTROL

Whilst the focus of the body of this document is on the basis for the architecture, the overall and individual module architectures and related functional aspects, such as the visual identity, for DS2 to achieve its goals will also need a strongly managed development process which is detailed on the GitHub (https://github.com/ds2-eu) and is based on software development and U project best practices. This is mandatory for all development partners to ensure.

Annex A on the GitHub covers:

- Management Responsibility
- Development Control and Infrastructure
- Source Code Management
- Progress Tracking
- Software Testing
- Pipelines
- Software and Accompaniements Delivery

13 CONCLUSION

First part of the deliverable D2.2. provides the key outcomes of the work done with use case partners, focusing on findings and recommendations. It emphasizes the establishment of user requirements, baselines, and the alignment between user needs and DS2 technical modules. This part of the document particularly highlights how these requirements are mapped to ensure optimal functionality across multiple sectors, such as smart cities, precision agriculture, and the Green Deal. The document also presents the first version of use case 4, with an initial idea of data sharing and the technological setup of the environment. It also illustrates the successful definition of industry-specific and technology-specific KPIs across use cases, ensuring that both technical and operational goals are met. These KPIs are integral in evaluating the success of the DS2 project in promoting sustainability, optimizing resource use, and enhancing cross-sector data collaboration.

Finally, this section of the document provides a transition between the requirements mapped out in this section and the technical architecture to be explored in the second part of the document. This ensures that the DS2 system will address identified user requirements while providing a robust and scalable data-sharing infrastructure. A general link between requirements and the selection of the DS2 modules is provided, and a detailed description of the modules can be found in the second part of this document.

The second part of this deliverable has provided the specification for all Modules to be constructed and delivered by DS2, including their fit into the DS2 'Big Picture', their primary subcomponents, and internal connections as well as external interfaces showing their interactions with high-level sequence diagrams. Each also specifies base technologies and their licenses, risk, and security concerns as well as data governance issues. Each also shows resource planning and a delivery timetable. This will thus provide a solid starting basis for developments in WP3-6 which will enable developers to plan and for planning and monitoring by the various leadership roles.

Equally, architectures are not static and will evolve due to both internal and external influences and discover especially in an research and innovation project. As such, the architecture will be revisited overtime, updated, and openly published online which will encourage others to explore the DS2 technical capability and direction.

With the completion of this document the project now moves broadly from the imaging, initial research, and paperwork phase into a deeper technical phase of deeper research and practical implementation also accompanied with WP3 delivery at M12 re "Data Governance and Methodologies - Knowledge Initialisation". This implementation will have a 6 monthly pace with regular WP, and Technical WP leads reviews. At M18, the user's setups should have been completed and from then onwards there will be a process of user validation and feedback until the remainder of the project.

ANNEX: MODULE DESCRIPTIONS

Process Orchestration is an example of one module. All other modules follow the same format and content, but for reasons of size and ease of use, all module architecture documents are reachable on the GitHub repository: https://github.com/ds2-eu/architecture

During the remaining time of the project, these documents will be adapted to a more dynamic Git/HTML approach and additional information added, as detailed elsewhere in the document.