13 Data Space Discovery, Assessment, Recommend and Configure Module (DARC)

13.1 Data Space Discovery, Assessment, Recommend and Configure Module (DARC)

Owner(s): ATC
DOA Task: T5.2 & T5.3
Tier: 1
Nature: Optional/System
Result: K5.2 & K5.3

T5.2: This Task will focus on developing an innovative tool for supporting data consumers and data providers to initiate a complex DLC for data sharing between data spaces. Such data sharing requires discovery and assessment of the technical, business, operational and organisational capabilities (building blocks) of participating data spaces. In this Task, an AI-driven conversational assistant, in the form of an intelligent 'chatbot', will be developed to efficiently elicit information related to the above data spaces' building blocks. Human-centric and user-friendly 'questions-and responses' schemes will guide the users to provide the necessary information. It will work in conjunction with the language utility of T5.1 to support multi-lingual use. Moreover, the intelligent 'chatbot' will analyse the elicited information to assess the technical, business, operational and organisational prerequisites for a complex DCL to execute successfully between participating data spaces automatically where possible. The development of the intelligent 'chatbot' will be based on the humancentric research work performed in WP2 which will initiate the necessary data spaces' assessment framework/model updated in WP5, as necessary. Established 'chatbot' development environments, such as RASA, will be considered. The gender dimension (e.g. anti bias) will, where possible, be factored into these developments.

T5.3: This Task will complement T5.2 and focus on developing an innovative tool for supporting data consumer and data providers to recommend and configure a complex DLC that protects the sovereign rights, including compliance with the EC regulations, over the complex life cycles of data sharing, aggregation and provenance. Such configuration will be based on the assessment performed in T5.2 on the technical, business, operational and organisational aspects of the participating data spaces, which always need to precede it. It will, therefore, respect data consumer ethical and moral values, privacy concerns as well as trade-offs between security, privacy, and data sharing end result. In this Task, a similar AI-driven 'chatbot', will be developed in order to intelligently moderate the efficient configuration of the complex DLC. Human-centric and user-friendly 'questions-and responds' schemes will converse with users to arrive at a recommended configuration of the involved complex DLC. It will work in conjunction with the language utility of T5.1 to support multi-lingual use. It will then interface through the APIs of other WP3-6 modules to perform auto-assisted configuration based on the interaction and recommendations. The gender dimension (eg anti bias) will, where possible, be factored into these developments.

13.1.1 Introduction

During the proposal phase, the task was initially divided into two separate tasks that complemented each other as per the description in the DOA. However, upon further technical investigation during the first months of the project, it became apparent that the modules of T5.2 and T5.3 that were to be delivered will be part of the same technical architecture and will be developed and implemented on the same technological framework. Consequently, it was decided to merge them into a single, unified architecture as well as into a unified implementation plan because they will be provided as a seamless stack of functionality. Hence, the contents of this document consider the full-stack view.

Purpose: 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.

Description: The DARC Module is responsible for allowing data space participants (Data Providers and Data Consumers) to inquire and discover and receive AI-driven assessment, of prerequisites for the successful execution of complex Digital Life Cycles (DLCs) between participating Dataspaces. Then, for DARC to make recommendations of modules and their configuration based on this assessment with the end game to automatically configure modules which are suitably enabled for configuration.

Discover: Through an intuitive UI of an AI-driven conversational agent (Chabot) end-users are able to interact and receive feedback on:

- What kind of Data they need to access and for what purpose, thematic focus, scientific/ research, business or another supported domain.
- Where those specific Data they want to access are stored, i.e. which participating Dataspace.
- How they will access and use those Data.

Assess: Thus, the Chabot provides a comprehensive assessment on what kind of complex DLCs are required to be established between the participating Dataspaces or within a specific Dataspace, if this connection can be fully automated by DS2 modules or not and if there are any limitations that they need to know.

Recommend: The AI-driven conversational agent continues by providing participants a recommended stack of DS2 modules so that they can successfully build their very own IDT. This will also include any related details about Dataspace sovereign rights (legal information, terms and conditions), EC regulations (GDPR etc) as well as security risk mitigation recommendations for their DLC to process and use the Data they want to from selected Dataspace participants. Furthermore, the gender dimension in terms of bias (e.g. issues with LLMs expressing biased assumptions about men and women through stereotypical references) will be addressed, where possible, to ensure bias is excluded from the end-results of the recommendations made by the AI-driven conversational agent.

Configure: To produce a set of configuration instructions and rules on the set-up of the recommended DS2 modules Then, for a limited set of modules (due to the advanced

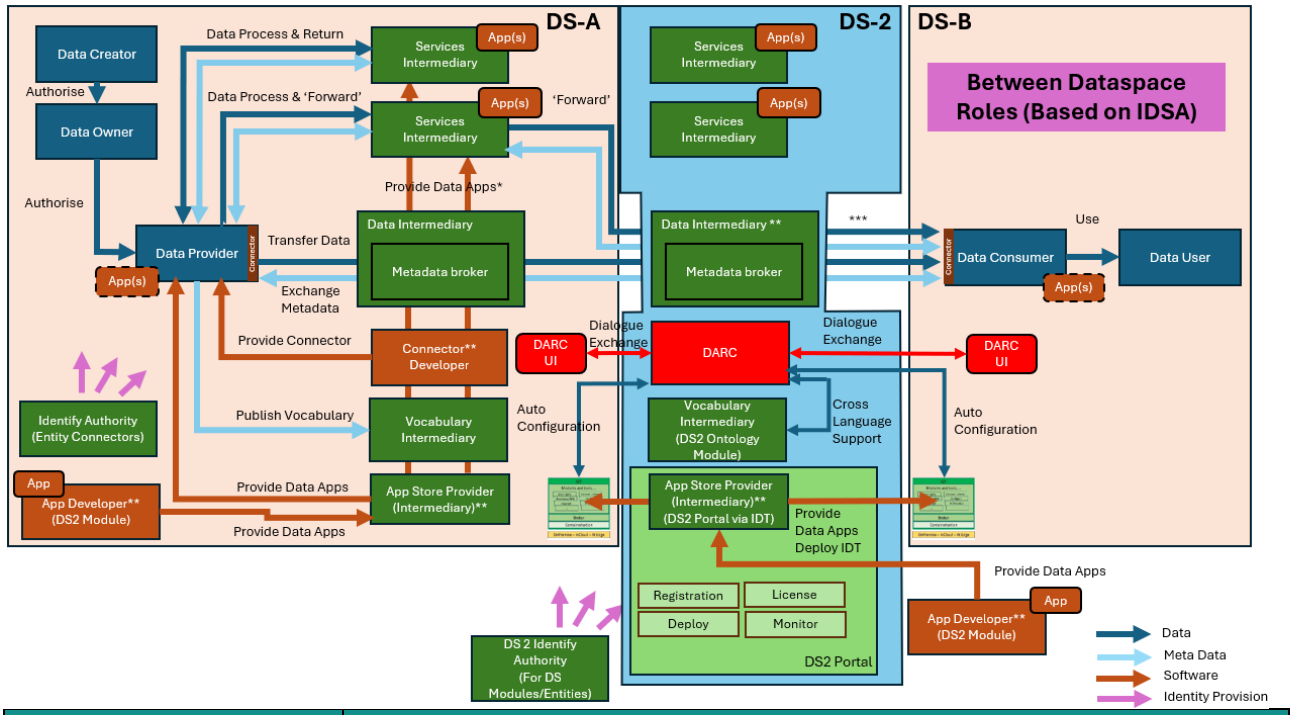
nature of this concept) to interface through their APIs to perform auto-assisted configuration based on the interaction and recommendations. Additionally, the Containerization Module by ICE will be utilized to achieve auto-configuration by using deployment files (Helm chart) to auto-configure other modules that cannot support the endpoint configuration. For example the Data Rights Management (DRM) module will include an automatic configuration endpoint. Another module that can be supported by auto-assisted configuration is the Data Inspector. The Data Inspector can generate notifications and execute actions, such as sending requests or notifications to external tools. However, configuring this module correctly can be complex for users. Therefore, using DARC can be beneficial. Modules that utilize AI, such as the Cultural and Language Model and Data Curator, can also benefit from auto-assisted configuration. AI services provide numerous configuration options, such as encoding and analysing tokens, which can be complex for users to manage. DARC can simplify the configuration process for these modules, making them easier to set up and use. A more detailed list will be created once the configuration options for each module are determined.

Addressing bias: In addressing the gender dimension bias within large language models (LLMs), a multi-faceted approach is crucial. Initially, a thorough analysis and documentation of biases must be conducted, especially through diverse and extensive prompt testing at the baseline evaluation phase. This foundational work helps pinpoint specific areas of gender bias. Following this, techniques such as Prompt Engineering (PE) and In-context Learning (ICL) are employed to modify the model outputs to fit within pre-established boundaries or contexts, effectively reshaping responses. To assess the effectiveness of these interventions, controlled prompts similar to those used in the initial evaluation are retested on the model to gauge any changes or improvements. Furthermore, deploying the model in real-world scenarios to generate content allows for the testing of these guardrails in practical applications, offering insights into their operational efficacy and revealing any persisting biases. This comprehensive strategy ensures that gender biases are not only identified but also systematically addressed, guaranteeing a fair and unbiased performance in real-life applications.

13.1.2 Where this component fits

13.1.2.1 Big Picture

DARC can be used by Participants who themselves could be within the same Dataspaces or different ones as demonstrated in diagram below. DARC itself will be a service intermediary within the DS2 environment (pillar in the middle of the diagram below) whereas different instances of the AI-driven Chat bot UI (DARC UI) will be deployed as a module within the IDTs running in DSA and DSB. Users will interact with the Chat bot UI to get information about how a specific set/stack of DS modules could be configured to meet their requirements and -where possible - automatically configured by DARC. Interaction will also allow data to be discovered from within Dataspace participants/DS Meta Data brokers. DARC modules will interface with the Vocabulary Intermediary (DS2 Culture and Language) Module for multilingual support and assistance with the disambiguation of the ontology of the recommended data.



Where	Status
Within a single Dataspace for use between participants in that Dataspace only	Yes: DARC can be deployed and used by Participants within a single dataspace to initiate data sharing and assess as well as recommend and configure DLCs, ensuring compliance with EC regulations and respecting privacy and security concerns by deploying DARC UI and DARC core within dataspace's boundaries.
Deployed and used by a single participant to enable the participant in either an In-Data space or Inter- Data space scenario	Yes: DARC can be deployed by a single participant to facilitate data sharing in DS2. By installing the DARC UI and DARC core on-premises, the participant can initiate data sharing, assess, recommend, and configure DLCs, all while ensuring compliance with EC regulations and maintaining privacy and security.
Across Dataspaces without Service Intermediary	Yes, DARC can serve without service intermediaries across dataspace, by deploying and configured DARC UI and DARC core within dataspace premises. Participants will utilize DARC for assistance in configuring complex DLCs.
Across Dataspace with Intermediary	Yes, DARC can be used across dataspace with intermediary service. By deploying DARC UI participants will able to utilize DARC for assistance in configuring complex DLCs.
Other Comments	N/A

13.1.2.2 Within a single Dataspace (where applicable)

DARC to investigate the current status of the Dataspace they are involved in and what they will need to do to exchange data with other participates by simply asking questions to the Chat bots and receiving comprehensive feedback and smart recommendations.

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

DARC can be deployed by a single participant to facilitate data sharing in DS2. By installing the DARC UI and DARC core on-premises, the participant can initiate data sharing, assess,

recommend, and configure DLCs, all while ensuring compliance with EC regulations and maintaining privacy and security.

13.1.2.4 Across Dataspaces without intermediary (where applicable)

DARC is far greater for Data Providers and Data Consumers that participate to different Dataspaces. Participants could easily investigate the current status of all -relevant to their needs - Dataspaces and have all the required information presented to them in readable format with comprehensive guidelines of how they should proceed.

13.1.2.5 Across Dataspace with Intermediary (where applicable)

DARC provides substantial benefits to consumers engaged in various dataspaces. Consumers can efficiently explore and assess the status of dataspaces relevant to their specific needs. This will ensure that all pertinent information is presented in a clear and accessible format, accompanied by detailed instructions on how to navigate and optimize their interactions within these dataspaces.

13.1.3 Component Definition

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

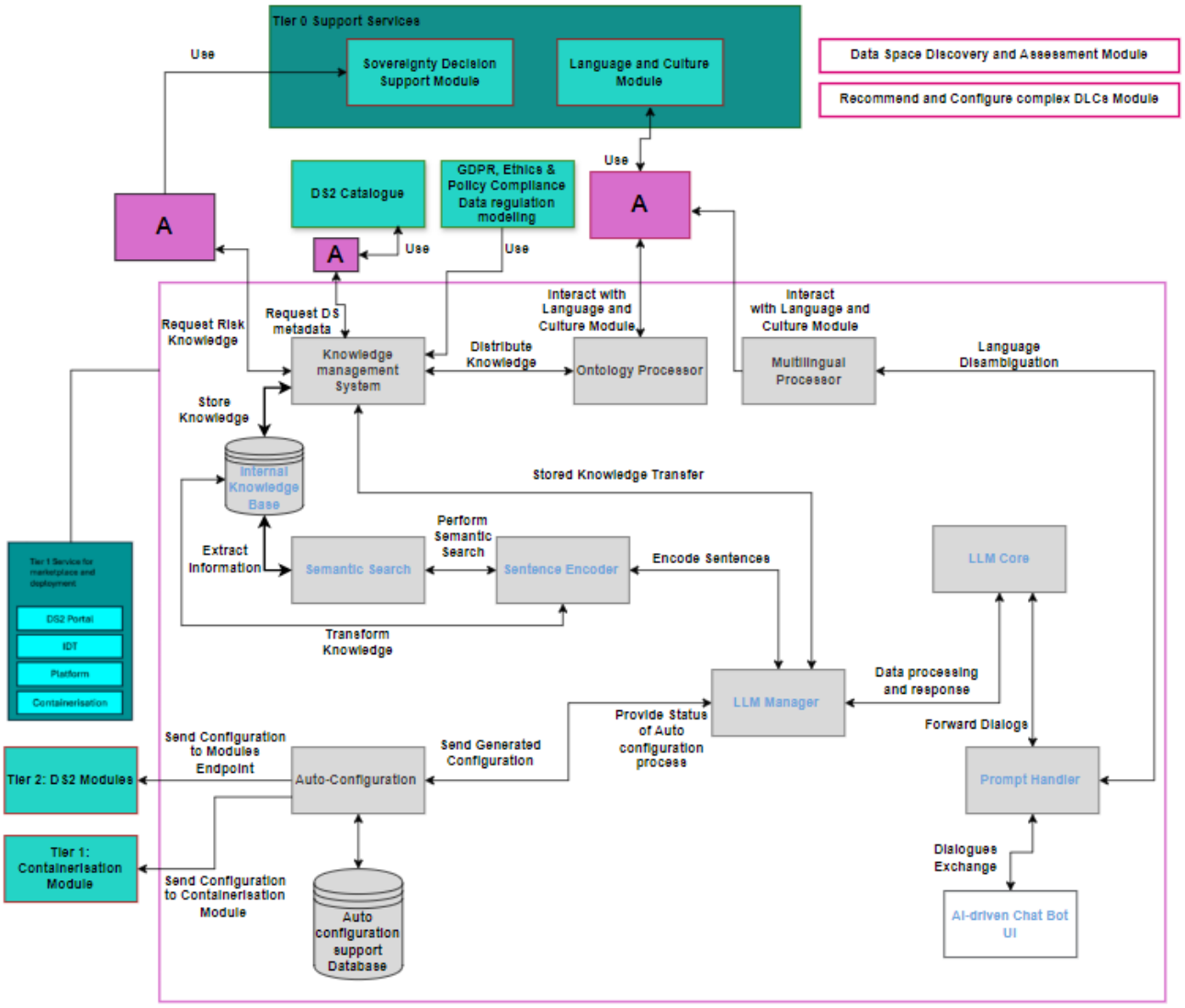


Figure 1: Schema for DARC

The subcomponents of DARC are as follows:

- **AI-driven Chat Bot UI:** The User interface subcomponent through which end-users interact with DARC. It allows end-users to provide inputs and receive outputs, facilitating a smooth and intuitive interaction DARC. This subcomponent directly interfaces with end-users for dialogue exchange.
- **Prompt Handler:** The Prompt Handler facilitates the exchange of dialogue between the Large Language Model and end-users. It performs initial checks, correcting grammar errors and verifying supported languages. If a language is not supported, the module engages the Ontology/Multilingual Processor for language disambiguation before forwarding the prompt to the LLM Core. This ensures that interactions are both clear and linguistically compatible.
- **LLM Manager:** The LLM Manager subcomponent oversees and coordinates the operations of the language model. It ensures seamless integration with other DARC subcomponents and manages the flow of data to and from the LLM core. It acts as an manager and is responsible for all DARC's module operations that interact with the LLM. This subcomponent facilitates interaction with the LLM Core, managing

dialogues, transferring crucial information from the Knowledge Base (using the sentence encoder and semantic search subcomponents), and providing it to the LLM Core to generate outcomes based on user requests. Furthermore, it engages with the Auto Configuration subcomponent to forward the configuration file generated by the LLM Core and reports back to the LLM Core the status of auto configuration to update the user with the status.

- **LLM Core:** The primary language model that generates outputs based on the processed inputs and configurations. The LLM Core is the heart of DARC, providing the natural language processing capabilities that drive DARC's recommendation and configuration capabilities. This subcomponent actively interacts with the LLM Manager, requesting additional details about user queries and receiving crucial information from the knowledge base to ensure dynamic and contextually aware interactions. Additionally, it works with the Prompt Handler, which forwards the tailored responses back to the user, completing the communication loop effectively.
- **Sentence Encoder:** The Sentence Encoder, an AI model, converts dialogues and knowledge base content into embeddings (a structured numerical format that makes complex linguistic data more understandable for the LLM and other AI components). This process supports LLM operations by enabling the Semantic Search subcomponent to extract and decode crucial information from these embeddings, transforming it back into text for the LLM Core. This integration enhances DARC's ability to deliver precise and contextually relevant responses to user queries.
- **Semantic Search:** This subcomponent enhances search functionality by using context-aware algorithms to analyse transformed data from the Knowledge Base and inputs derived from user prompts. It processes this information and relays it to the Sentence Encoder subcomponent , which converts the embeddings back into textual data. This allows the LLM to further process and understand the content effectively.
- **Internal Knowledge Base:** This is a central repository that stores all essential data and knowledge, such as Regulations and DS2 module documentation, configuration. Serving as the backbone of the application, it provides a robust database that DARC subcomponents can access through Knowledge Management. This ensures that DARC has a solid informational foundation for making informed decisions and configurations. The Knowledge Base works in conjunction with the Knowledge Management subcomponent , which continuously populates it with critical operational details, such as module configurations and compliance with EC regulations. It also interfaces with Semantic Search to extract necessary information for the LLM.
- **Knowledge Management System:** The Knowledge Management System subcomponent organises, stores, and retrieves knowledge within DARC. It facilitates efficient information management by structuring and indexing data such as, Dataspace regulations, DS2 modules documentation and configuration information, for easy access and distribution of said data within DARC subcomponents. The Knowledge Management subcomponent supports the internal Knowledge Base and interacts with other DARC subcomponents, namely LLM Manager and Ontology Processor, to provide relevant information as needed. This subcomponent engages with the Data Sovereign Module API, the Culture and Language Module API, and Dataspace Meta-data Broker ensuring that the Knowledge Base is continually updated with crucial information necessary for DARC's operations. The Knowledge Management System also interacts with the internal knowledge base by incorporating previous information into the internal knowledge base for enhanced decision-making and operational efficiency

- **Ontology Processor:** This subcomponent plays a crucial role in managing ontology information within DARC. It is tasked with processing, storing, and updating ontology data in the Knowledge Base. Additionally, it retrieves and analyses the ontology of data already present in the Knowledge Base. By ensuring that ontology information is accurately maintained and accessible, this subcomponent enhances the DARC's ability to interpret and utilize data effectively, fostering a deeper understanding of the relationships and structures within the stored information.
- **Multilingual Processor:** This subcomponent provides robust cross-language support, enhancing DARC's global functionality. It processes outputs from the Culture and Language subcomponent, enabling the LLM to comprehend and interact with previously unknown languages. Additionally, it interacts with the Prompt Handler subcomponent to ensure that multilingual interactions are seamlessly integrated, allowing users retrieve information about datasets that are in other languages. This capability significantly improves accessibility and user experience across diverse linguistic backgrounds.
- **Auto-Configuration:** This subcomponent plays a crucial role in automating the setup process for DS2 modules. It begins by receiving a configuration file that has been generated and forwarded by LLM Manager, originally derived from the LLM core. The subcomponent systematically examines this file to identify and rectify any potential errors or inconsistencies. Once it confirms that the configuration file is error-free, it proceeds to implement these settings directly on the DS2 modules' endpoint. After its successful application, it completes the process by sending a detailed status report back to the LLM through the LLM Manager subcomponent, ensuring seamless communication and integration within DARC.
- **Auto configuration Support Database:** This subcomponent maintains essential information such as module endpoints and configuration rules to facilitate the functionality of the auto-configuration subcomponent.
- **Services and APIs:**
    - **Catalog Module**: Catalog Module Meta data broker will send the requested Dataspaces meta-data to DARC's Knowledge Management subcomponent for discovering data in the participating Dataspaces.
    - **Tier 2 DS Modules/Service:** Auto-configuration service. for Tier 2 DS Modules via the Auto-Configuration Module.
        -
    - **Tier 0 Support Service Stack:**
        - **Sovereignty Decision Support** Module Risk Knowledge Base: API will be driven by DARC to provide information about potential security risks involved during the configuration, assessment, and definition of a participant's DLC.
        - **Culture and Language Module**. Ontology and Multi-language Facility Module API – will provide DARC with multilingual support and ontology disambiguation, enhancing DARC's ability to interpret and process diverse linguistic data accurately and effectively across different languages.
    - **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 were used

- **Tier 1 Containerization Module**: The Containerization module utilizes Helm Chart standard base templates to describe a DS2 module, with Helm charts containing the necessary configuration variables for the module. By leveraging the DARC auto-configuration module, the required configuration variables for the Helm chart can be generated and passed to the Containerization module, facilitating the deployment of the DS2 module.

13.1.4 Recommend Configuration and Autoconfiguration of modules

DARC will enhance user support by offering recommendations and automatically configuring DS2 modules. This functionality not only streamlines the setup process but also optimizes module performance based on user needs and system requirements. By integrating these intelligent features, DARC aims to simplify the user experience and improve operational efficiency.

For DARC to effectively recommend and auto-configure DS2 modules, it is essential that module owners should provide DARC module owners with a bare minimum of documentation. This documentation should include configuration templates for each module, detailed descriptions of each configuration parameter, and a step-by-step guide to the configuration flow. Additionally, it should outline all possible values, the interdependencies between steps, and any potential risks involved. This thorough documentation will enable DARC to provide tailored configurations that meet specific user requirements while ensuring smooth and reliable operation.

The configuration templates will be processed and transformed for use by the LLM. This will enable the LLM to assist users effectively through a question-and-answer format, facilitating both manual and automatic configuration of the modules. This approach ensures that users receive tailored guidance and support, enhancing the setup process and overall usability of the modules.

13.1.5 Technical Foundations and Background

ATC brings extensive experience and expertise to the table in the development of conversational agents, leveraging cutting-edge technologies to create intuitive and efficient chatbot solutions. Additionally, ATC possess a deep understanding of advanced data retrieval techniques, including robust algorithms like RAG (Retrieval-Augmented Generation), enhancing the effectiveness and accuracy of information retrieval processes Retrieval-Augmented Generation (RAG) is the process of optimizing the output of a Large Language Model, so it references an authoritative knowledge base outside of its training data sources before generating a response. This RAG technique will be instrumental in achieving DARC functionalities, enabling the system to dynamically access and integrate up-to-date, relevant data from diverse sources, thus ensuring more informed and contextually accurate interactions within the DARC framework.

Subcomponent/Component	Owner	License
Milvus	Milvus	Apache 2.0
Neo4j	Neo4j	Apache 2.0
Faiss	Meta	MIT

13.1.6 Interaction of the Component

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

With Module/Feature	Receives From/Gives To	What
Catalogue	Receive From	Metadata contains of Dataspace.
GDPR, Ethics & Policy Compliance Data regulation modelling	Receive From	Defines Regulations that must be followed
Tier 2 Modules	Receive From	Auto configuration Endpoints
Tier 2	Give To	Configuration settings of the tool
DSS Risk Knowledge Base	Receive From	Information about Risks
DSS Risk Knowledge Base	Give To	Query to retrieve information about risks
Language and Culture module XYZ API	Give To	Non-English prompt that user typed
Language and Culture module XYZ API	Receive From	Ontology Language Disambiguation details
Containerization module	Give To	Configuration Helm Chart for the deployment

13.1.7 Technical Risks

Risk	Description	Contingency Plan
Configuration of modules	Configuration of modules could be problematic if there is not much information.	Make sure tool owners provide detailed information about tools
Configuration of modules	The configuration of modules requires them to support Endpoint configuration (based on DOA). However, many modules may not have this capability.	Generate Helm configuration files and supply them to the Containerization module, enabling deployment with the appropriate configuration.
Cross Language Support	Support for multiple languages may be constrained by the challenges of ontological translation and the restricted number of languages available.	Explore alternative methods to enhance translation quality and improve the effectiveness of ontology translation.
Dataspace Internal Regulations	The internal regulations of a dataspace are not clearly defined regarding when DARC should retrieve that information.	After discussions with the catalogue owner, we agreed to include an additional field that describes restrictions on the catalogue, allowing this information to be provided to DARC.

13.1.8 Security

Security Issue	Description	Need
Inter-participant trust	There is no specific DS security risk	N/A
In-Dataspace	There is no specific DS security risk	N/A

13.1.9 Data Governance

Data Governance Issue	Description	Need
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.

13.1.10 Requirements and Functionality

This module will be used in the following usecases:

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

Their 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.

In respect to all use cases, the DS2 DARC module is classified as an Optional Module but also System one. It is “System” in that it is not part of a user scenario in that it can simply be optionally used to configure other modules which are used within any use case. As such it does not appear explicitly in any user scenarios which are connected with datasharing vs configuring for datasharing.

WHERE	WHAT	WHY	Run/Design Time	Priority
	Use Case 1: City Scape			
N/A				
	Use Case 2: Green Deal			
N/A				
	Use Case 2: Precision Agriculture			
N/A				

13.1.11 Workflows

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

13.1.11.1 Data Discovery and Regulations

This feature allows users to easily ask the AI agent questions about data and its regulations within dataspaces. Users can receive quick and clear information directly from the AI. It simplifies access to data guidelines and rules.

The main steps/functionalities are as follows:

- Prompt
- Refine Prompt
- Ontology and Language disambiguation
- Use knowledge base
- Knowledge augmentation of the query
- Return answer to user

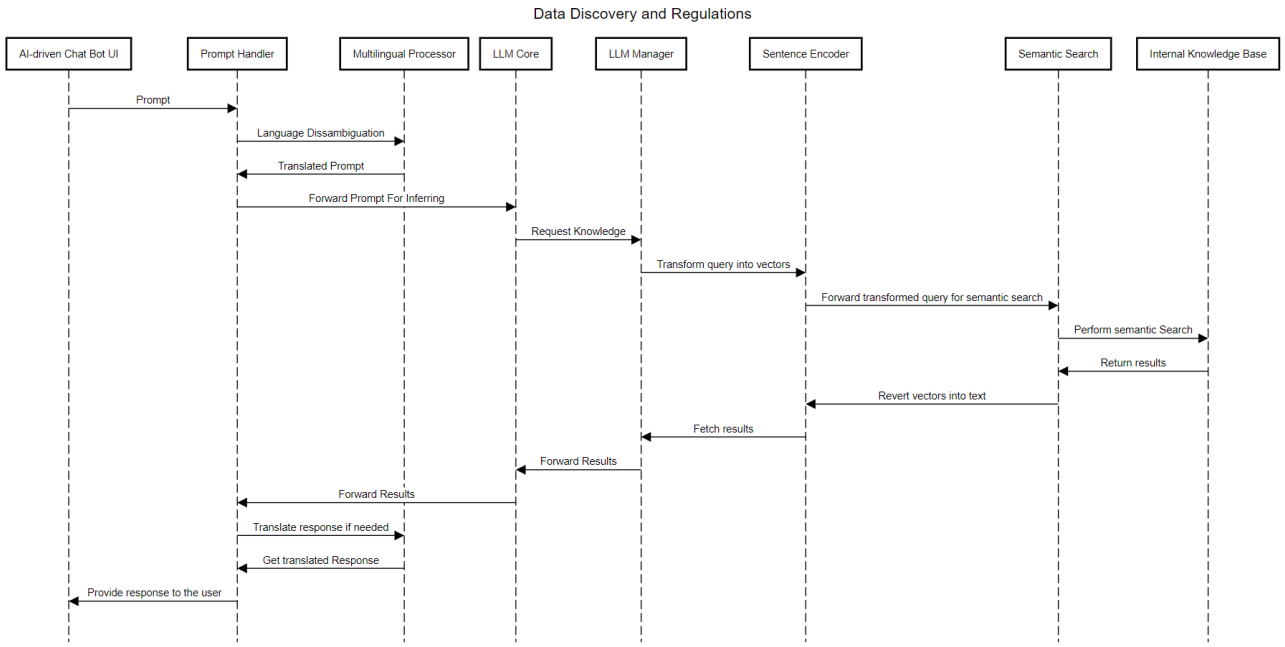


Figure 2: Data discovery and regulations sequence diagram

13.1.11.2 Discovery And Assessment

This module helps users discover and assess the capabilities of various data spaces for sharing. An AI-driven chatbot guides users through providing necessary information and analyses it to evaluate technical, business, operational, and organizational requirements. This simplifies initiating complex data linkage and sharing.

The main steps/functionalities are as follows:

- Prompt
- Refine Prompt
- Ontology and Language disambiguation
- Use knowledge base
- Knowledge augmentation of the query
- Return answer to user

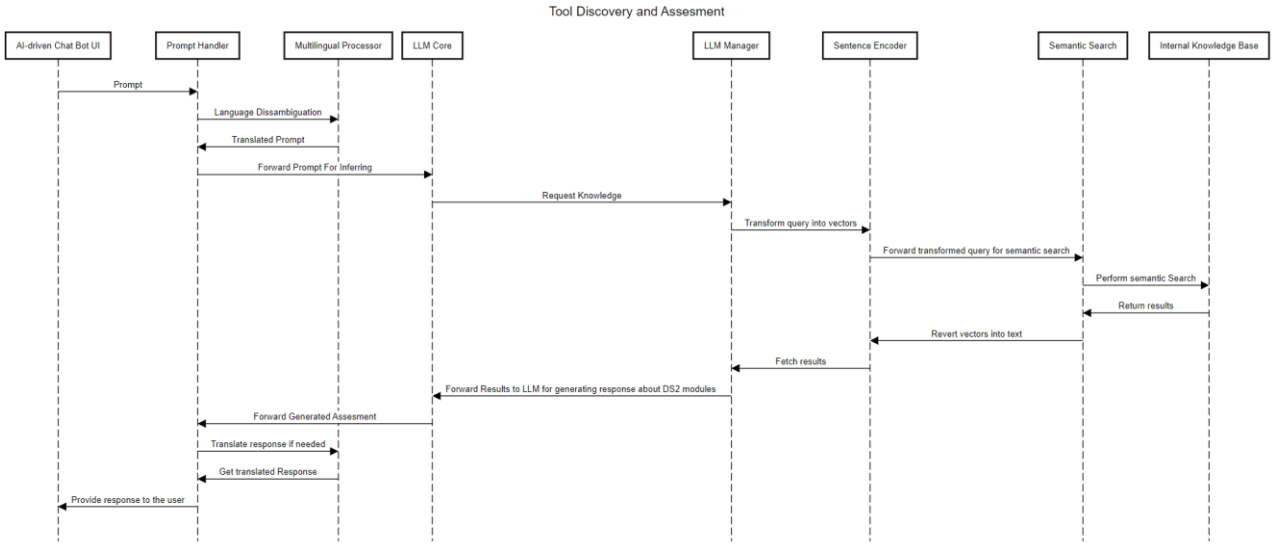


Figure 3: Tool discovery and assessment sequence diagram

13.1.11.3 Recommend and Assist DLC

This module helps users recommend and configure complex data life cycles (DLCs) for data sharing. An AI-driven chatbot guides users through the process, ensuring compliance with regulations and balancing privacy, security, and data sharing needs. It simplifies the configuration of DLCs by providing clear, user-friendly interactions.

The main steps/functionalities are as follows:

- Prompt
- Refine Prompt
- Ontology and Language disambiguation
- Use knowledge base
- Knowledge augmentation of the query
- Return answer to user

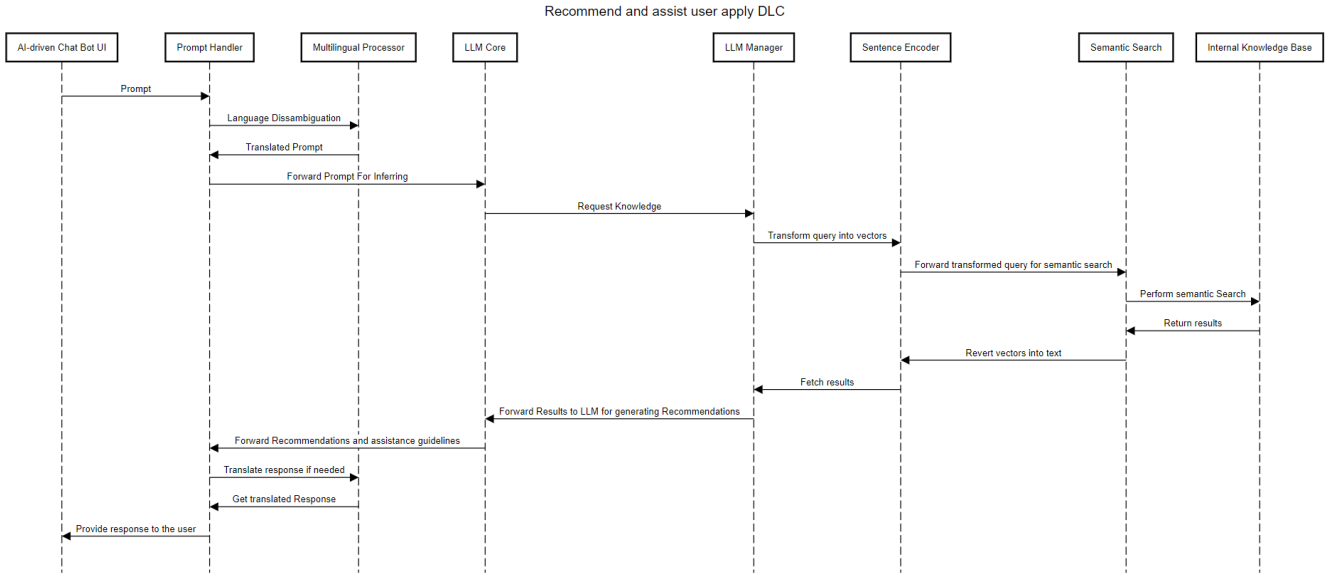


Figure 4: Recommend and Assist DLC sequence diagram

13.1.11.4 Configure DLC

This module helps users configure complex data life cycles (DLCs) for data sharing. An AI-driven chatbot guides users through the configuration process, ensuring compliance with regulations and balancing privacy, security, and data sharing needs. It simplifies the DLC setup with clear, user-friendly interactions.

The main steps/functionalities are as follows:

- Prompt
- Refine Prompt
- Ontology and Language disambiguation
- Use knowledge base
- Knowledge augmentation of the query
- Generate configuration
- Forward configuration to auto configuration module
- Apply configuration.

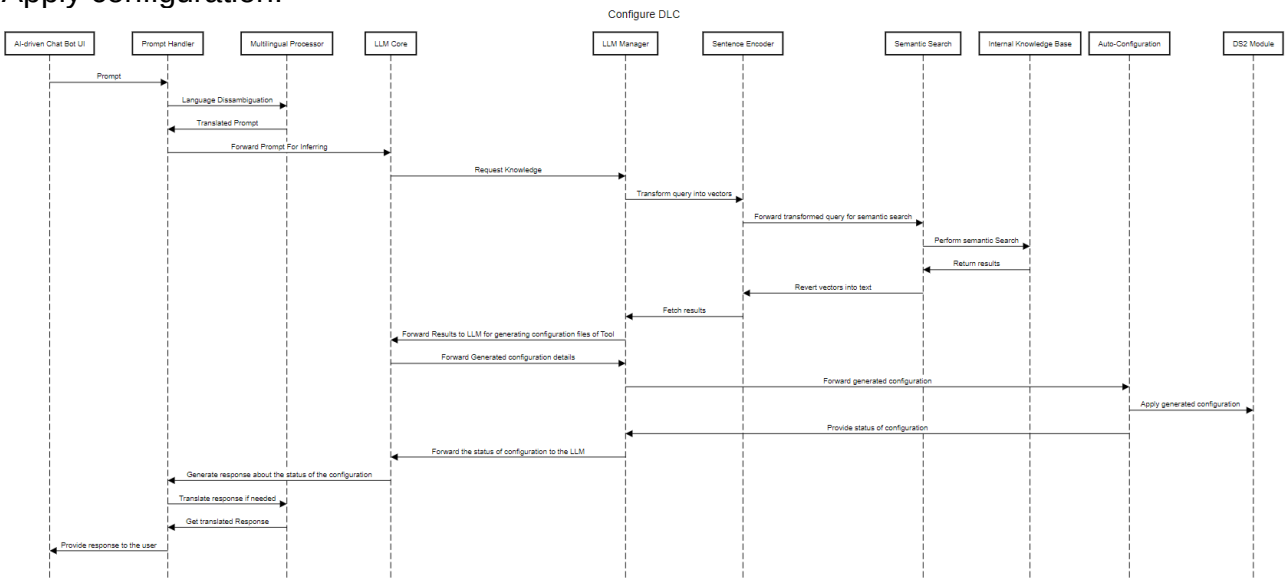


Figure 5: Configure DLC sequence diagram

13.1.11.5 Knowledge Base Population and Information Flow

The flowchart below illustrates the process by which DARC's Knowledge Base acquires information, stores it within the internal knowledge base, and utilizes it effectively.

- Request Data from corresponding API
- Fetch Data
- Normalize Data
- Store data to internal knowledge base

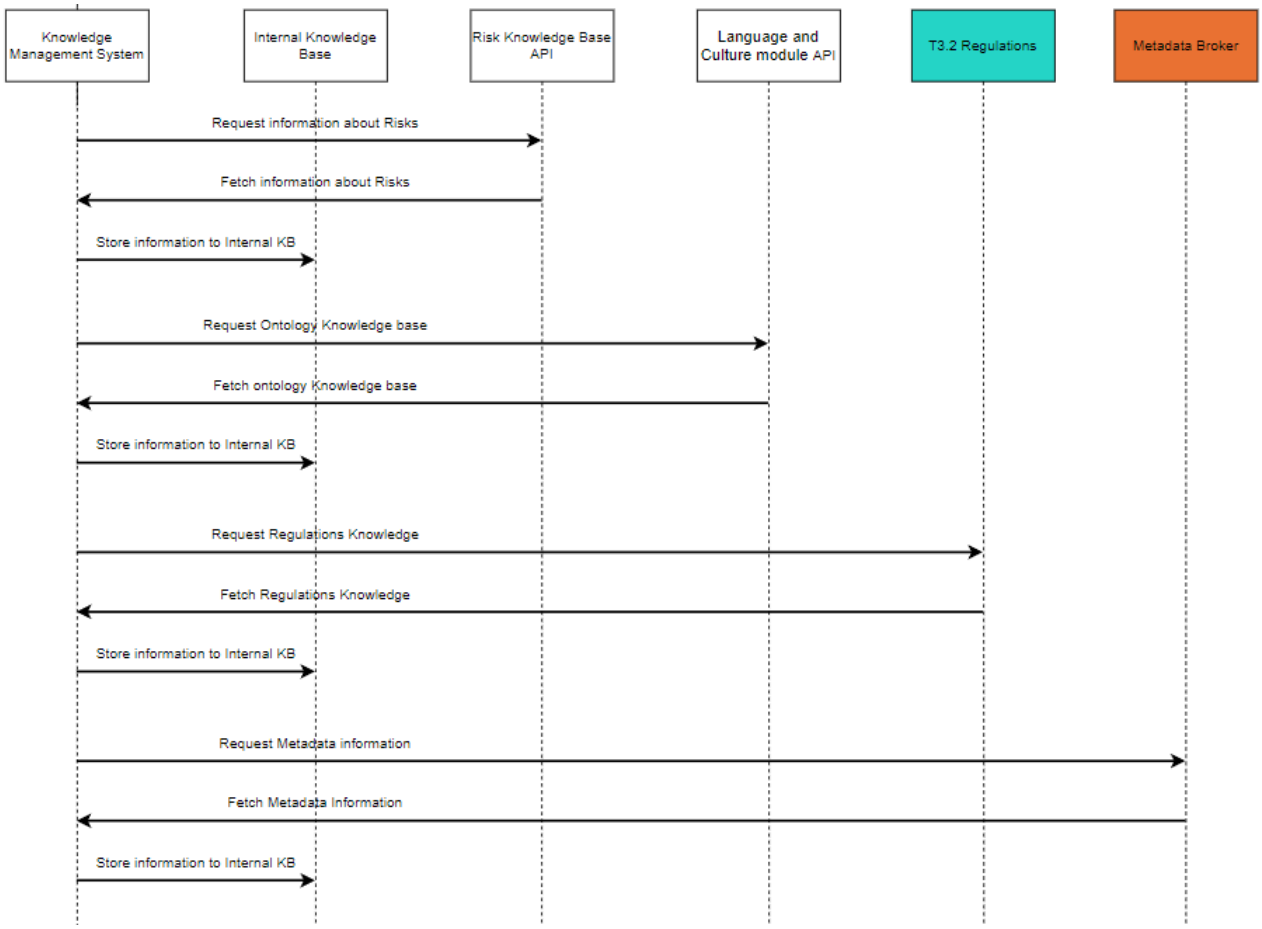


Figure 6: Knowledge Base population and information flow diagram

13.1.12 Role, Resourcing, and Milestones

Sub-component	Main Activity	M18	M24	M30	M36
Ontology Processor	Convert the outcomes of the language and culture module into data formats compatible with Large Language Models				
Multilingual Processor	The language and culture module does not offer direct language translation; instead, it focuses on 'disambiguation' of language through an ontology-based approach. We need to explore and develop methods to facilitate cross-language support using this type of disambiguation.				
Knowledge Management System	Development of a data “orchestration” subcomponent designed to manage, distribute, and transform knowledge across DARC components as needed.				
Internal Knowledge Base	Configure deploy knowledge store				
Semantic Search	Development of component that utilizing context-aware algorithms to analyse data from the Knowledge Base and user inputs, processing this information for effective understanding and retrieval by the LLM				
Sentence Encoder	Development of AI model that converts (and reverts) texts into vector representations				
Prompt Handler	Development of a subcomponent to manage and refine user prompts for optimal clarity and language compatibility				
LLM Manager	Development of a subcomponent that orchestrates the LLM’s operations, coordinating data flow and module interactions. Phase 1: Development of functionality for interaction with internal knowledge Base				
LLM Manager	Development of a subcomponent that orchestrates the LLM’s operations, coordinating data flow and module interactions. Phase 2: Development of functionality for interaction with auto configuration Module				
Auto-Configuration	Development of a subcomponent to automate and streamline the configuration of DS2 modules, ensuring error-free implementation. Phase 1 Gather knowledge about Tool’s configuration, processes and store to internal knowledge Base				
Auto-Configuration	Development of a subcomponent to automate and streamline the configuration of DS2 modules, ensuring error-free implementation. Phase 2 Develop functionality to process generate configurations, error corrections etc.				
Auto-Configuration	Development of a subcomponent to automate and streamline the configuration of DS2 modules, ensuring error-free implementation. Phase 3: Develop Functionality that forward configurations to DS2 modules, parsing error				

	messages and update LLM about the progress or any correction that might be needed.				
LLM	Development of Chatbot UI, discover functionality, bias mitigation				
LLM	Fine-tune and train LLM for assess and recommend functionality				
LLM	Fine-tune and train LLM for assisted and auto configuration functionality				
Table Total/DOA Task Total/Resilience	**Comments:**				

13.1.13 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
Ontology and Multilingual support	How the Culture and Language module can be used by DARC to achieve multilingual support	Strong interaction with component owner during architecture phase and continue this in further design / implementations	INTU
Which modules	It is unlikely there will be resource provision for the configuration of all modules, thus the aim will be to take 3 optional modules a prime candidates and ones from different uses cases	Assess with module owners best feasible candidates as a module subset	ATT
Auto configuration	Guidelines and API definitions for tool owners on creating documentation and informational materials to support the automatic configuration learning phase.	Research on existing tools	All tool owners which are from a subset agreed
Dataspace internal regulation	Need to be defined how the internal regulation of dataspace will be accessible by DARC to assist in the assessment and recommendation of DLC	Schedule meetings with UC owners	Use Case owners
Catalog	The precise mechanisms of exchange between DARC and the CAT module needs to be determined	Schedule meetings with CAT owner	ATC, VTT