Eurostars Project

SPARQL-ML Machine Learning for SPARQL Query Optimization over Centralized and Distributed RDF Knowledge Graphs

Project Number: 5736 Start Date of the Project: 2024/11/01 Duration: 36 months

Deliverable 6.1 SPARQL-ML Progress Report

SPARQL-ML Progress Report

Dissemination Level	Confidential
Due Date of Deliverable	30.04.2024
Actual Submission date	29.04.2025
Work Package	WP6
Deliverable	D6.1
Type	Report
Approval Status	Actual version
Version	0.5
Number of Pages	22

Abstract: SPARQL-ML takes data management to the next level, bringing high-performance, scalability, and efficient handling of semantic data to the market. Today, there is an increasing need for efficient data distribution and federated query engines to manage large datasets. The main goal of SPARQL-ML, the new research initiative coordinated by eccenca, is to develop two novel and generic solutions for high-performance RDF Knowledge Graph data management: a SPARQL Query planner and optimizer for triplestores, and a SPARQL query planner and optimizer for federation engines.

eureka eurostars

EUROPÄISCHE PARTNERSCHAFT

Kofinanziert von der Europäischen Union

GEFÖRDERT VOM

Bundesministerium für Bildung und Forschung

History

Version	Date	Reason	Revised by
0.1	12.12.2024	Initial Template and Deliverable Structure	Nieves Sande
0.2	16.12.2024	KPIs (OpenLink)	Milos Jovanovik
0.3	01.04.2025		Nieves Sande
0.4	01.04.2025	Updated content on WP1	Milos Jovanovik
0.5	29.04.2025	Last update	Nieves Sande

Author List

Organization	Name	Contact Information
eccenca GmbH	Nieves Sande	Nieves.sande@eccenca.com
OpenLink Software	Milos Jovanovik	mjovanovik@openlinksw.com
Paderborn University	Muhammad Sohail Nisar	msohail@mail.uni-paderborn.de

1. Executive Summary

This document presents the initial report on the Innovation Management and Communication Strategy Work Package (WP6), a pivotal component of the project, acting as its organizational backbone. This work package is designed to ensure the project stays aligned with its objectives, achieves milestones within the established timelines, and consistently upholds high-quality standards.

Following that, it summarizes the major activities and achievements made by each of the participating companies until M6, including progress on achieving technical objectives, the status of the KPIs, progress made for each WP, deviations and risks that occurred or are expected to occur, and how the team plans to mitigate these deviations and risks.

To ensure the project's success, a go/no-go decision will be made after each twelve-month period, based on a thorough evaluation of progress and risk assessment. However, we proactively monitor and analyze risk mitigation and KPI completion every six months. This semi-annual review process allows us to continuously assess the project's status, identify potential challenges early, and implement necessary adjustments to keep the project on track.

The plan outlined in this report will serve as the foundation for monitoring the project's impact and will enable the assessment of how the objectives evolve throughout the project. These evaluations will be documented as separate deliverables, both midway through and at the project's conclusion.

With the finalisation and submission of this deliverable, the Milestone MS8 SPARQL-ML progress has been achieved.

2. Overall Progress of the Project

The SPARQL-ML project focuses on addressing the challenge of optimizing SPARQL query processing within triplestores, a critical component of RDF knowledge graph management.

The collaboration between eccenca GmbH, Openlink, and the University of Paderborn's DICE group in the SPARQL-ML research project is crucial for advancing AI-driven query optimization over large-scale RDF Knowledge Graphs. By combining eccenca's expertise in semantic data management, Openlink's capabilities in high-performance database solutions, and DICE's cutting-edge research in machine learning for knowledge graphs, the project aims to develop novel, standards-compliant tools for optimized SPARQL query execution. This partnership fosters a synergy between academic research and industry applications, ensuring that the final outcome—a set of W3C-compliant tools—effectively enhances centralized and federated RDF data processing. The collaboration will ultimately enable more efficient, scalable, and intelligent data storage solutions, driving innovation in the Semantic Web and linked data ecosystems.

SPARQL-ML started on 01. November 2024, right after the start the consortium partners decided to organize the Kickoff meeting on the 13. November 2024 at the eccenca Headquarters in Leipzig. The Kick off was one day long of presentations. It commenced with a welcome session and a company tour, followed by a project overview presented by the coordination lead. Subsequently, consortium partners engaged in discussions on project management, focusing on key milestones and the definition of key performance indicators (KPIs). In the afternoon, partners delivered presentations on their respective product and research contributions, followed by a session dedicated to technical project planning. The day concluded with a networking social event and a dinner at a renowned local venue.

This meeting provided consortium members with the opportunity to connect in person, strengthening their collaboration and fostering a dynamic and productive working relationship. Notably, the positive and cooperative atmosphere among partners creates a highly favorable foundation for the project's success.

Prior to this meeting, the project coordinator (ECC) took the initiative to launch the project website, further detailed in the communication activities section of this document. The website, along with the successful conclusion of the kickoff meeting, provided consortium partners with a strong foundation to commence project activities. The specific tasks carried out within each work package are outlined in detail in the following section. Notably, both project management and technical contributions from all consortium members have been exceptional, facilitating smooth progress and seamless collaboration.

3. Progress for each WP

Between M1 and the start of M6 of the SPARQL-ML project, as shown in Figure 1, activities were focused solely on WP1 – Requirements Elicitation & Conceptual Architecture and WP6 – Innovation Management and Communication Strategy. Additionally, WP2 commenced in April 2025, contributing just one month of work during this timeframe.

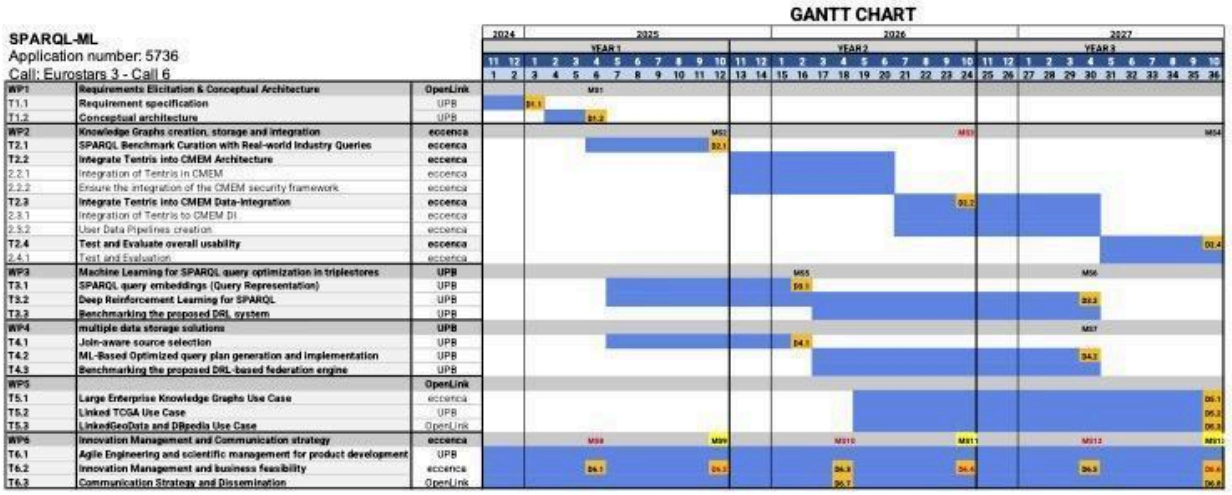

Figure 1. Gantt Chart of SPARQL-ML

WP1 (Lead: ALL)

This work package, led by OpenLink and involving all partners, has two main objectives. First, it aims to define the precise functional requirements for system components based on identified use cases. This includes an analysis of current developments and a comprehensive review of the state of the art. In the second phase, the consortium designs a technical architecture that aligns with the established requirements, which serves as the foundation for subsequent work packages.

D1.1 “Requirement specification” has been finished on time, it reports the functional requirements for the system components, based on the identified use cases. It contains the results of task T1.1 “Requirement specification”, as part of which the consortium conducted an in-depth review of the current state of the art and articulated the envisioned state post-project completion. This analysis was then used as the foundation for the requirements elicitation process, and the detailed requirements specification, which have been outlined in the deliverable.

The second task, T1.2 “Conceptual architecture” has been achieved. During this task and having the technical requirements as a starting point, the partners worked on and produced an overall system architecture for the project. This system architecture contains and describes the components and services of the system, their roles and interconnection, along with an overall system information flow, which details the main expected behaviors. All of these results have been summarized in D1.2 “Conceptual architecture”.

As a result, Milestone MS1 (M6) “Completion of requirements and specification for the system architecture” has been achieved.

Also all the activities from WP1 have been started and completed. Therefore there will be no other further developments in WP1.

WP2 Knowledge Graphs creation, storage and integration (Lead: ecc)

In Work Package 2, the primary focus is on achieving seamless integration between the triple store Tentris and the eccenca Corporate Memory data management platform. The overarching goal is to establish a robust connection that enhances the functionality and efficiency of the Corporate Memory system, particularly on query runtime execution and federated query processing.

The following tasks have been initiated:

- **Tentris integration with eccenca Corporate Memory as a backend storage system**: This task is underway. During the integration process, compatibility issues with SPARQL 1.1 were identified. These have been reported to the Tentris team and are currently being addressed.

- **Knowledge Graph Exploration with Tentris (as a backend of Corporate Memory)**: Progress on this task has started but was impacted by the compatibility issues discovered in the integration task above, which are currently under resolution.

The following tasks have not yet started:

- **Knowledge Graphs instantiation and management with Tentris (as a backend of Corporate Memory)**

- **Content Annotation Manager Tool (CAM) integration**

Additionally, Task T2.1 – *SPARQL Benchmark Curation with Real-world Industry Queries* – has also been initiated.

WP6 Innovation Management and Communication Strategy (Lead: ecc)

The Innovation Management and Communication Strategy Work Package (WP6) serves as the foundation of the project, playing a crucial role in ensuring its success. The primary objective of this work package is to guarantee that the project stays on track, meets all milestones within the specified timelines, and adheres to the established quality standards.

Progress within the Innovation Management and Communication Work Package (WP6) has been highly positive. Specifically, in relation to **T6.1 – Agile Engineering and Scientific Management for Product Development** – the internal coordination of iterative and agile engineering methodologies, including requirements management, has been successfully implemented. The technical approach centers on leveraging Machine Learning (ML), specifically Deep Reinforcement Learning (DRL), to enhance query execution efficiency by determining optimal or near-optimal join orders in query plans. Inspired by the success of DRL in optimizing relational database queries, this methodology holds significant potential to improve triplestore performance. The scientific method consists of the following steps:

1. Query Representation: Transforming SPARQL queries into vector embeddings to serve as inputs for the DRL agent. This involves selecting and extracting features from queries to capture their essence numerically, facilitating machine understanding and processing.

2. Environment and State Modeling: Integrating attributes from the Knowledge Graph (KG) such as size and the cardinality of query variables with the embeddings. This enriched representation forms the state space in which the DRL agent operates, providing a detailed observation of the current query environment.

3. Action and Reward Definition: Actions in this context involve selecting joins from the unoptimized query plan. The reward is inversely related to the query execution time, encouraging the agent to find faster execution paths.

4. DRL Agent Training and Feedback Loop: Utilizing a custom gym environment and Python bindings for seamless integration with the triplestore, the DRL agent iteratively improves its query optimization strategy based on feedback from executed actions.

5. Benchmarking and Evaluation: Comparing the optimized query execution times against those from state-of-the-art triplestores using real-world and synthetic SPARQL benchmarks.

The project's technical development is handled via GitHub, a powerful platform for version control and collaborative software development. GitHub is instrumental in managing the project's code, enabling team members to contribute, review, and test code in a structured and organized manner. Moreover, GitHub is seamlessly integrated with the project's official website, ensuring that technical updates, repositories, and progress are easily accessible to both internal and external stakeholders. This connection enhances the visibility of our work and promotes transparency, allowing collaborators and interested parties to track the latest developments in real-time.

Together with the shared cloud for management activities storage and GitHub provides the foundational infrastructure for collaboration, supporting the efficient execution of the SPARQL-ML project and fostering transparency, coordination, and innovation.

T6.2 – Innovation Management and Business Feasibility – is also progressing well. eccenca has taken the lead in overall project management, establishing a solid framework for coordination, including tools and procedures such as the project repository, a joint software development environment, mailing lists, and a meeting calendar. Some of the internal infrastructure described in detail below has been described in the Communication and Dissemination strategy that will be included in D.6.7 Intermediate Report on Performed and Planned Dissemination Activities which will be submitted by M18.

Internal Infrastructure

The SPARQL-ML project's primary platform for collaborative work and project management is a shared cloud storage space, which plays a pivotal role in our daily operations. This cloud-based workspace is overseen and regularly updated by the Project Coordination team, serving as the central hub for all project-related documents, resources, and vital information. The structure of the shared folder is organized to ensure easy access and retrieval of files, allowing all team members to efficiently contribute to the project. It also supports robust version control, ensuring smooth updates, tracking, and management of documents throughout the project lifecycle. This centralized storage is crucial for maintaining effective communication and collaboration, ensuring all team members are aligned with the most current developments. In addition to the shared folder, the project's technical development is handled via GitHub, which has been already described before together with the scientific management activities.

Furthermore, Overleaf is employed as our preferred platform for the creation, editing, and collaborative sharing of project deliverables. With its real-time editing capabilities and version control, Overleaf ensures that all team members have access to the

most up-to-date versions of documents, enabling effective collaboration on deliverable preparation.

For immediate communication, we rely on Skype as our primary tool for live discussions. Skype's instant messaging functionality allows team members to quickly resolve questions or issues. Additionally, it facilitates the scheduling and conducting of online consortium meetings, enabling efficient coordination across different time zones and geographic locations.

The project partners have agreed to meet once a month, on the second Wednesday of each month, to review recent developments in both technical and management areas. If needed, additional calls will be arranged to address specific issues. The consortium agreement has been successfully finalized and signed within the designated time frame. To ensure efficient coordination, the project partners have shared calendar invitations for these recurring meetings, which will continue until the end of the project. As a result, all meetings have already been scheduled in advance.

By integrating these tools, we ensure a highly efficient, organized, and collaborative workflow, which is essential for the successful progress of the SPARQL-ML project. Moreover, both Risk Management and KPI Management (in the next section) processes have been effectively initiated.

Risk Management

We have written a risk mitigation strategy plan that outlines various technology and management risks that could potentially affect the progress and success of the project, along with the corresponding strategies to address them. The coordinator with the WP leaders will be responsible for undertaking risk assessment and control at any time during the project especially at the end of each six month period. Risk assessment allows project management to:

- Identify risks and areas of uncertainty in the entire project plan.
- Analyze and evaluate how those risks and areas of uncertainty can impact the performance of the
- Project, either in duration, cost, or meeting the users' requirements.
- Prioritize risks, to establish which risks should be given gravest consideration (because of
- Potentially heavy impact), which should have regular management attention, and which are
- Sufficiently minor to avoid detailed management attention.

And risk control will allow the project manager to:

- Take whatever actions are possible in advance to avoid or mitigate the effects of risks. It is better
- To spend some money on mitigation rather than include a costly contingency plan.
- For all those risks which are deemed to be significant, there is a contingency plan in place before it happens.
- Evaluate, monitor, and track the effects of the risks identified and manage them to a successful conclusion.

As part of the project, these risks are divided into Technology risks and Management risks and they span multiple areas. Therefore, each identified risk is categorized based on its likelihood and potential impact, and tailored mitigation approaches have been devised to prevent or minimize these risks. The table in Annex 1 presents a comprehensive breakdown of these risks and their respective mitigation strategies.

Regarding the last Task in this Work Package: **T6.3 Communication Strategy and Dissemination**, the project partners started with the Work package with the publication of a Website.

Website and Logo

The SPARQL-ML project maintains a public-facing website, accessible to all internet users, at https://sparql-ml.eu/. The website is primarily managed and regularly updated by eccenca, with content contributions from all project consortium partners.

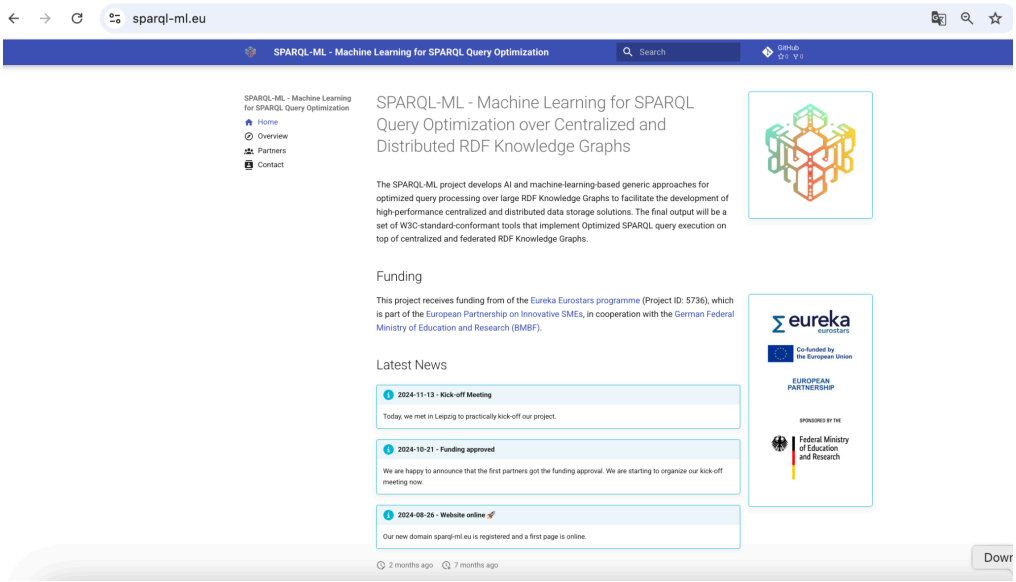

Figure 2. SPARQL-ML Landing Page

The website is organized into four main sections, each with relevant subsections to provide comprehensive information about the project:

- **Homepage**: The front page of the website, offering an overview and easy navigation to key sections. The SPARQL-ML website presents a clear and professional structure, starting with a Landing page that outlines the project's scope and goals, featuring the project logo and funding information. It highlights the financial support received from the Eureka Eurostars Programme (Project ID: 5736) and the German Federal Ministry of Education and Research (BMBF), with the logos of all funding institutions displayed prominently. The page also includes a section for the latest news, featuring updates such as the kickoff meeting.

- **Overview**: This section provides a detailed introduction to the SPARQL-ML project, including an overview of the project's objectives and its technical architecture. Overview, provides the public with an in-depth introduction to the project, explaining its objectives and technical architecture.

- **Partners**: A dedicated section listing all project partners, with direct links to their respective websites and online resources. This third page showcases the project's partners, displaying their logos along with brief summaries of their roles in the project.

- **Contact**: This section includes contact information, providing email addresses for each project area, allowing visitors to get in touch with the respective responsible team members of the project coordinator.

Additionally, the website features a search engine, enabling users to quickly find specific information. It also includes a direct link to the project's external GitHub repository, offering easy access to technical resources and code.

This well-organized structure ensures that visitors can easily access relevant project information, stay informed about progress, and engage with the technical community through GitHub.

The coordinator has also designed a project logo image file, which can be downloaded from our shared folder and every partner has the necessary rights and permissions to properly make use of it for dissemination activities.

Figure 3. SPARQL-ML Logo

The SPARQL-ML project will enhance its visibility and impact through a strategic dissemination and communication plan. A first draft of this plan has been already written, a second draft will be submitted by M18 in D6.7. This first draft is focusing on publishing at least one high-quality research papers in reputable journals and conferences, with early-stage and mid-to-late-stage publications to share findings and conclusions. The project aims to participate in 3–6 conferences or workshops, including ISWC, ESWC, and industry-focused events, ensuring active contributions through presentations and discussions. A key initiative is organizing the Text2SPARQL workshop at ESWC where the project can be presented. Cost-effective strategies include leveraging virtual events, co-authoring papers with partners, and seeking institutional support. This approach ensures broad scientific, technical, and commercial outreach.

Deliverables and Milestones Management

WP#	D#	Lead	Title	Month	DONE?	Date
WP1	D1.1	OpenLink	Requirement specification	3	YES	31.01.2025
WP1	D1.2	OpenLink	Conceptual architecture	6	YES	31.04.2025
WP6	D6.1	ECC	SPARQL-ML Progress Report #1	6	YES	31.04.2025

Table 1. SPARQL-ML list of Deliverables for the last period

During the first six months of the project, only three deliverables were scheduled for submission. The first, D1.1 Requirements and Specifications, was successfully completed and submitted on time, as outlined in the activities of Work Package 1 (WP1). The second deliverable, D1.2 Conceptual Architecture, represented the next key task within WP1 and has been thoroughly detailed in its respective section of this document. This deliverable was also completed and submitted on schedule.

The third deliverable, D6.1 Progress Report, is this very document. As demonstrated here, it has been successfully prepared and submitted within the required timeframe.

WP#	MS#	Lead	Title	Month	GO/NO GO?	Achieved?
WP1	MS1	Openlink	Completion of requirements and specification for the system architecture	6	NO	YES
WP6	MS8	eccenca	SPARQL-ML progress	6	YES	YES

Table 2. SPARQL-ML list of Milestones for the last period

During the first six months of the project, two milestones were scheduled for completion. Milestone 1 (MS1): Completion of Requirements and Specifications for the System Architecture was successfully achieved by Month 6. Additionally, Milestone 8 (MS8): SPARQL-ML Progress and First NOGO Milestone was also successfully met within the designated time frame.

KPI Management

Related to the KPI Management, all key performance indicators (KPIs) have been successfully described. The project's Key Performance Indicators (KPIs) serve as essential reference points for tracking progress and ensuring alignment with project goals. While these KPIs provide a foundational framework, they will be continuously reviewed and refined throughout the project to adapt to evolving requirements and insights.

By the end of the sixth month of the project, the primary work packages actively contributing to its progress have been **WP1** and **WP6**. WP1, focused on Requirements Elicitation and Conceptual Architecture, has been instrumental in defining the functional and technical requirements, as well as developing the initial system architecture to guide subsequent phases. Simultaneously, WP6, which centers on Innovation Management and Communication Strategy, has played a crucial role in ensuring that the project remains aligned with its objectives, milestones are met within the planned timelines, and effective communication is maintained among all stakeholders. The contributions of these work packages have laid a solid foundation for the project's development, enabling a structured and coordinated approach to achieving its goals.

For those Work Packages that have been started we have agreed and fixed following KPIs:

WP1 (Lead: ALL)

KPI 1.1: 100% requirements have been gathered (done)

KPI 1.2: 1 conceptual architecture has been produced (done?)

WP6 (Lead: ecc)

KPI 6.1: All planned deliverables for the period have been submitted

KPI 6.2: All planned milestones for the period have been successfully achieved

KPI 6.3: KPI strategy has been updated and validated

KPI 6.4: Communication dissemination strategy has been updated

KPI 6.5: Risk and mitigation table has been updated and validated

WP2 (Lead: ecc)

KPI2.1: 100% successful scalability tests for federations from 2 to 10 endpoints.

KPI2.2: 80% success in integration tests with seamless Knowledge Graph construction

WP#	KPI	#	KPI Description	Achieved?	Month	MS
WP1	1.1	100%	Requirements have been gathered	YES	6	MS1
WP1	1.2	1	Conceptual architecture has been produced	YES	6	MS1
WP6	6.1	100%	Deliverables have been finalised	YES	6	MS8
WP6	6.2	100%	Milestones have been successfully achieved	YES	6	MS8
WP6	6.3	100%	KPI strategy has been updated and validated	YES	6	MS8
WP6	6.4	100%	Communication dissemination strategy has been updated	YES	6	MS8
WP6	6.5	100%	Risk and mitigation table has been updated and validated	YES	6	MS8
WP2	2.1	100%	Successful scalability tests for federations from 2 to 10 endpoints	to be achieved by	12	MS9
WP2	2.2	80%	Success in integration tests with seamless Knowledge Graph construction	to be achieved by	12	MS9

Table 3. KPIs for the last period

The table outlines various work packages (WP), key performance indicators (KPIs), and milestones (MS) for a project. It tracks the progress of each work package by listing specific KPIs, detailing what has been accomplished, and showing the respective milestones and months when these objectives were completed.

Here is a description of the columns and the data in the table:

- WP#: The Work Package number, which indicates different phases or areas of the project. For example, WP1 and WP6 are two different work packages.
- Title: A brief description of the KPI or task within each work package. These provide specific tasks or objectives to be achieved. For example, "Requirements have been gathered" or "Deliverables have been finalised."
- #: The percentage of achievement or numeric identifier for each KPI. For example, "100%" indicates the KPI has been fully achieved.

- KPI Description: A more detailed description of the specific task or objective for each KPI. It explains what was to be accomplished. For example, "Conceptual architecture has been produced" or "Risk and mitigation table has been updated and validated."
- Achieved?: A "YES" indicates the KPI has been successfully completed within the required timeframe.
- Month: The month in which the achievement was completed. For all the listed entries, this is "6", indicating that the deliverables were achieved by the sixth month of the project.
- MS: The milestone associated with each KPI. It indicates which milestone the achievement is related to. For instance, "MS1" for the first work package and "MS8" for the milestones associated with WP6.

For the upcoming work packages, we have already prepared a draft of potential KPIs that can be incorporated to enhance our overall KPI strategy. These preliminary indicators are designed to align with project objectives, ensuring comprehensive performance tracking and continuous improvement. By refining and expanding our KPI framework, we aim to establish clear benchmarks that will support efficient monitoring and decision-making throughout the project's next phases.

WP3 (Lead: UPB)

KPI3.1: A sufficient number of queries should have corresponding embeddings.

KPI3.2: We aim to build a fully trained DRL model that generates optimized query plans

WP4 (Lead: UPB)

KPI4.1: We aim for a significant decrease in query execution time over existing methods.

KPI4.2: The goal is to have a better query runtime.

WP5 (Lead: Openlink)

KPI5.1: 100% Evaluation of the tensor-based query engine in combination with traditional triplestores on real use case applications has been successful

KPI5.2: The solution should handle datasets than those used in state-of-the-art evaluations without performance degradation.

KPI5.3: We aim to improve the query execution time w.r.t state of the art

4. KPIs Evaluation and GO/NO GO decision

All key performance indicators (KPIs) have been successfully met within the designated timeline. Key deliverables, including the completion of reports, validation of communication and dissemination strategies, and updates to the risk and mitigation table, have been finalized as planned. The project has remained consistently aligned with its objectives, achieving all milestone requirements by month six. Ongoing monitoring and refinement of the KPI strategy continue to ensure that the work package stays on track, supporting the overall success of the project. The active collaboration among consortium partners has been instrumental in maintaining high-quality outputs throughout the work package's execution.

Following a comprehensive evaluation of the defined KPIs at the six-month mark, all indicators have been successfully achieved within the expected timeframe, with no significant technical or management risks identified at this stage.

The GO/NO-GO decision is scheduled for the end of the first project year. However, based on the current progress in both development and management, it is anticipated that the project will proceed as planned. This decision will be reassessed before year-end, with the coordinator engaging with the project officer to provide an update and confirm the next steps.

5. Deviations and Risks

During the six-month period of the project, there were no reported deviations from the planned objectives, timelines, or deliverables. All tasks progressed as scheduled, and no unforeseen risks or challenges emerged that could have impacted the project's overall trajectory. The consortium effectively managed the implementation of planned activities, ensuring that both technical and managerial aspects remained on track. Regular monitoring and evaluation confirmed that the project proceeded smoothly, maintaining alignment with initial expectations and strategic goals.

ANNEX 1: Risk Management Table

#	RISK	Risk Level	Definition	Mitigation strategy	Impact
			Technology Risks		
Risk 1	**Expressivity of SPARQL**	Moderate to High	While ML approaches for relational databases demonstrate significant performance improvements, the expressive nature of SPARQL could potentially undermine the overall performance gains.	Further fine-tuning of the proposed approach may be necessary to achieve the desired performance. In this regard, the University of Paderborn has already begun exploring this topic to identify the key features of SPARQL queries that require special attention in the proposed research method. We are confident that the proposed solution will push the boundaries of current advancements and enable us to achieve performance results that are orders of magnitude better than existing solutions.	If the expressivity of SPARQL is not effectively addressed, it could lead to performance delays, requiring additional time and resources to optimize queries and adjust the approach, which may push back project milestones. More time and expertise may be needed to explore and resolve these challenges, potentially diverting attention from other critical tasks. If left unaddressed, this risk could also affect scalability, limiting the expected performance improvements from the ML approach. In the long term, failure to resolve these issues could impact the system's ability to handle large-scale datasets, jeopardizing the viability and success of the approach.
Risk 2	**Triple store availability and crashes**	High	Data storage availability is a significant challenge for RDF knowledge graph providers. In our case, this issue is even more critical, as we anticipate handling multiple parallel requests from end users.	We will stress-test the data storage using our IGUANA framework (https://github.com/AKSW/IGUANA) to identify the crash points of the storages. Based on this analysis, we will develop mechanisms within the underlying triplestores to handle user requests without reaching these crash points. Additionally, we may leverage replicas of the highly loaded data storage to distribute the computation and improve overall performance.	By stress-testing the data storage using the IGUANA framework, we will identify potential crash points, allowing us to develop mechanisms within the triplestores to handle user requests without reaching these critical limits. This proactive approach will help ensure the system remains stable under high load, preventing service interruptions. Furthermore, by leveraging replicas of the highly loaded data storage to distribute computation, we can enhance performance and scalability, ensuring the system can efficiently handle increasing user demands and large datasets without compromising reliability.
Risk 3	**Scalability**	Moderate to High	We aim to develop generic solutions that lead to scalable triplestores. It is possible that the proposed approaches perform better on small datasets and may not scale well when it comes to large knowledge graphs. Therefore, it is important to test the scalability of the proposed solution.	We have defined both initial and final deliverables for each of the main technical tools. The approach involves developing each tool in the initial deliverable, with a focus on scalability in the final deliverable. Additionally, we aim to base all technical developments on scalable frameworks and implementations. To achieve this, we have carefully selected two of our partner's triplestores and various use-case datasets, ranging from small to very large, to test scalable runtime performance.	The impact of this approach is that by clearly defining both initial and final deliverables for each of the main technical tools, we ensure that the solutions are designed with scalability in mind from the outset. The focus on scalability in the final deliverable allows us to develop robust and future-proof solutions. By selecting two different triplestore systems and utilizing various datasets — ranging from small to very large — we can thoroughly test the scalability of the performance. This enables us to efficiently identify potential scalability issues and address them accordingly, optimizing system performance and ensuring the development of a solution that remains efficient even as data volume and user load increase.

				By utilizing two distinct triplestores and multiple real-world datasets, we can efficiently identify the root cause of any potential scalability issues and address them accordingly.	
Risk 4	**Components compatibility and integrability**	Moderate	Different components of the SPARQL-ML will be developed by different members of the consortium. It is possible that they are not easily compatible and integrable to each other.	We will clearly define the input and output parameters for all interfaces provided by the various components. All interfaces will adhere to W3C standards to ensure that the developed components are generic and can be easily integrated with other components and products. Additionally, we will organize a series of meetings and hackathons to ensure smooth integration and resolve any issues that may arise.	The impact of this approach is that by clearly defining the input and output parameters for all interfaces and adhering to W3C standards, we ensure that the developed components are modular, generic, and easily integrable with other components and products. This will enhance the interoperability of the system, allowing for smooth integration with external systems and future scalability. Organizing a series of meetings and hackathons will further facilitate collaboration, enabling the identification and resolution of integration issues early on. As a result, this approach will minimize potential roadblocks during the integration phase, ensuring timely project progress and a more cohesive final product.
Risk 5	**Misunderstanding or failure to meet user requirements**	Moderate	It is possible that a task is not clearly specified by the users or it is not correctly implemented by the consortium member.	We will prepare questionnaires and hold in-person discussions to clarify the exact system specifications with the end users. Users will be involved throughout the project to ensure their needs are addressed. In general, an agile software development approach will help mitigate risks #4 and #5.	The impact of this approach is that by preparing questionnaires and holding in-person discussions, we ensure that the system specifications are clearly understood and aligned with the end users' needs from the outset. Continuous user involvement throughout the project helps in addressing any concerns early, reducing the likelihood of miscommunications or unmet expectations. Adopting an agile software development approach allows for iterative feedback and flexibility, which will enable timely adjustments to the system based on user input. This will minimize risks related to unclear requirements and improper implementation, ensuring that the project stays on track and delivers a solution that meets user needs effectively.
Risk 6	**Testability**	High	Most products are hard to test, or no test plans being made.	We have predefined the main standard key performance indicators for each component which are already used in the literature	The impact of this approach is that by predefining the main standard key performance indicators (KPIs) for each component, we ensure that the project's success is measured using widely accepted and established benchmarks from the literature. This allows for easier comparison with existing solutions and ensures that the components meet recognized performance standards. By using standardized KPIs, we also enhance the credibility of the results, making it easier to evaluate the effectiveness of the system and identify areas for improvement. Additionally, these predefined KPIs can guide development and testing processes, helping to keep the project aligned with industry standards and expectations.

Risk 7	**Unexpected events**	High	Even the best planned project can be tarnished by unexpected events.	Each deliverable will be reviewed internally after the submission for the desired functionalities. In case of unforeseen issues, we will do an overall assessment of how serious each problem is, then work on the most serious problem first.	The impact of this approach is that by conducting internal reviews of each deliverable after submission, we ensure that any issues related to the desired functionalities are identified early. This allows us to quickly assess the severity of unforeseen problems and prioritize resolving the most critical ones first. By addressing the most serious issues promptly, we can minimize delays, reduce the impact on the overall project timeline, and maintain focus on delivering key functionalities. Additionally, this process enhances the quality of the project, as it ensures continuous improvement and refinement throughout development, ultimately leading to a more reliable and functional final product.
Management Risks					
Risk 8	The inability of some partners to commit their duties (e.g., lack of personnel with the right expertise, lack of components, underestimation of complexity, covid/war issues etc)	Moderate	Activities require the collaboration of some partners to complete the task. This is reasonable since the consortium is well balanced and fully committed in terms of knowledge and skills. Furthermore, all partners have committed to tasks which are of strategic interest for their business or research activities and, as such, have a vested interest in successfully completing these tasks	The coordinator will resolve any commitment issues on a case-by-case basis, reallocating effort to other partners, if necessary, in an extreme case.	Delays in achieving the results or partial task completeness. This approach may result in delays because reallocating efforts to other partners often involves redistributing tasks and responsibilities, which can disrupt the established workflow and coordination within the team.
Risk 9	The task required more personnel or may be delayed	Moderate	The work plan of the project has been carefully planned, and the partners have a lot of experience in coordinating/running research projects.	In case of delay, the WP leader, with the Coordinator, will define mitigation actions and will re plan the activities and shorten those tasks with enough leeway in favor with the ones that have been affected by the delay.	This process can lead to delays in achieving the overall results as it requires adjusting the project timeline, redistributing responsibilities, and potentially altering the scope of certain tasks to compensate for lost time. While some tasks may be shortened or rescheduled, it can also result in partial completion of other tasks as the project team works to balance the workload and minimize further setbacks. These adjustments may cause a shift in focus and priority, potentially delaying the delivery of fully completed tasks or impacting the overall project timeline.

Risk 10	Milestone Delay or Quality Deficiency Risk	Moderate	One of the milestones is not met, or the outcomes do not meet the required quality.	Workpackage Lead and Coordinator decide the gravity of the problem and act as outlined. A minor delay may be addressed by a discussion with the involved partners to find a solution. More serious delays may require reallocation of resources or tasks to prioritize the bottleneck. A minor quality problem will be dealt with through a minor revision performed by the involved partners. A more serious quality problem will require Coordinator to initiate discussion.	Similarly, while minor quality problems may be resolved through quick revisions by the involved partners, more serious quality issues may take additional time to address, as they could involve a more thorough investigation and intervention by the Coordinator. These efforts, although necessary for resolving problems, can ultimately slow down the progress of the project, causing delays in achieving the intended results.
Risk 11	The task required more personnel or may be delayed	Moderate	The work plan of the project has been carefully planned, and the partners have a lot of experience in coordinating/running research projects.	In case of delay, the WP leader, with the Coordinator, will define mitigation actions and will re-plan the activities and shorten those tasks with enough leeway in favor with the ones that have been affected by the delay.	This reallocation of effort may create imbalances, with some activities being rushed or altered, potentially affecting their outcomes. Additionally, the process of re-planning and re-adjusting the schedule may cause temporary disruptions, leading to further delays in the overall delivery of results. While the intention is to minimize the impact, these adjustments could still slow progress and delay the achievement of final objectives.
Risk 12	Departure of key partner	High	The departure of a key partner can significantly impact project continuity, strategic direction, and stakeholder relationships. This risk may lead to a loss of critical expertise, disruptions in decision-making, delays in project timelines, and potential financial or reputational consequences.	In case of departure of a key partner, there will be a quickly communication with all stakeholders to ensure transparency about the situation and reassure them that steps are being taken to minimize disruptions. An evaluation about the specific areas where the departure has the most significant impact, such as the loss of expertise, disrupted project timelines, or potential gaps in decision-making. Identifying these areas allows for targeted actions to address the immediate needs. We will try to find out companies to replace the partner. If this is not possible we believe we will still be able to produce excellent results using the remaining structures for complex large-scale simulations	While this approach aims to minimize the risks, the impact of a key partner’s departure could still cause delays, resource reallocation challenges, and temporary disruption to the project’s progress.
Risk 13	Departure of key experts	Moderate	The departure of key experts can result in a loss of critical knowledge, skills, and experience essential to the project's success.	They will be replaced by a different expert, preferably from same partner, otherwise from a different partner. The expertise of partners overlaps decreasing impact for many topics.	Overall, while the overlapping expertise reduces the risk of major disruptions, this approach may still lead to minor delays and temporary adjustments in the project’s flow, especially during the transition period as the new expert integrates into the project.

Risk 14	IPR disputes	High	Intellectual Property Rights (IPR) disputes arise when there is a conflict over the ownership, usage, or infringement of intellectual property, such as patents, trademarks, copyrights, or trade secrets.	To mitigate this risk, we have clear agreements and clear rules on IPR from the beginning. We will conduct thorough due diligence, and ensure compliance. The dispute settlements are tackled in the Consortium Agreement.	This approach significantly mitigates the risk of IPR disputes, but the potential impact includes minor delays or increased operational effort in case disputes arise and require formal resolution processes.