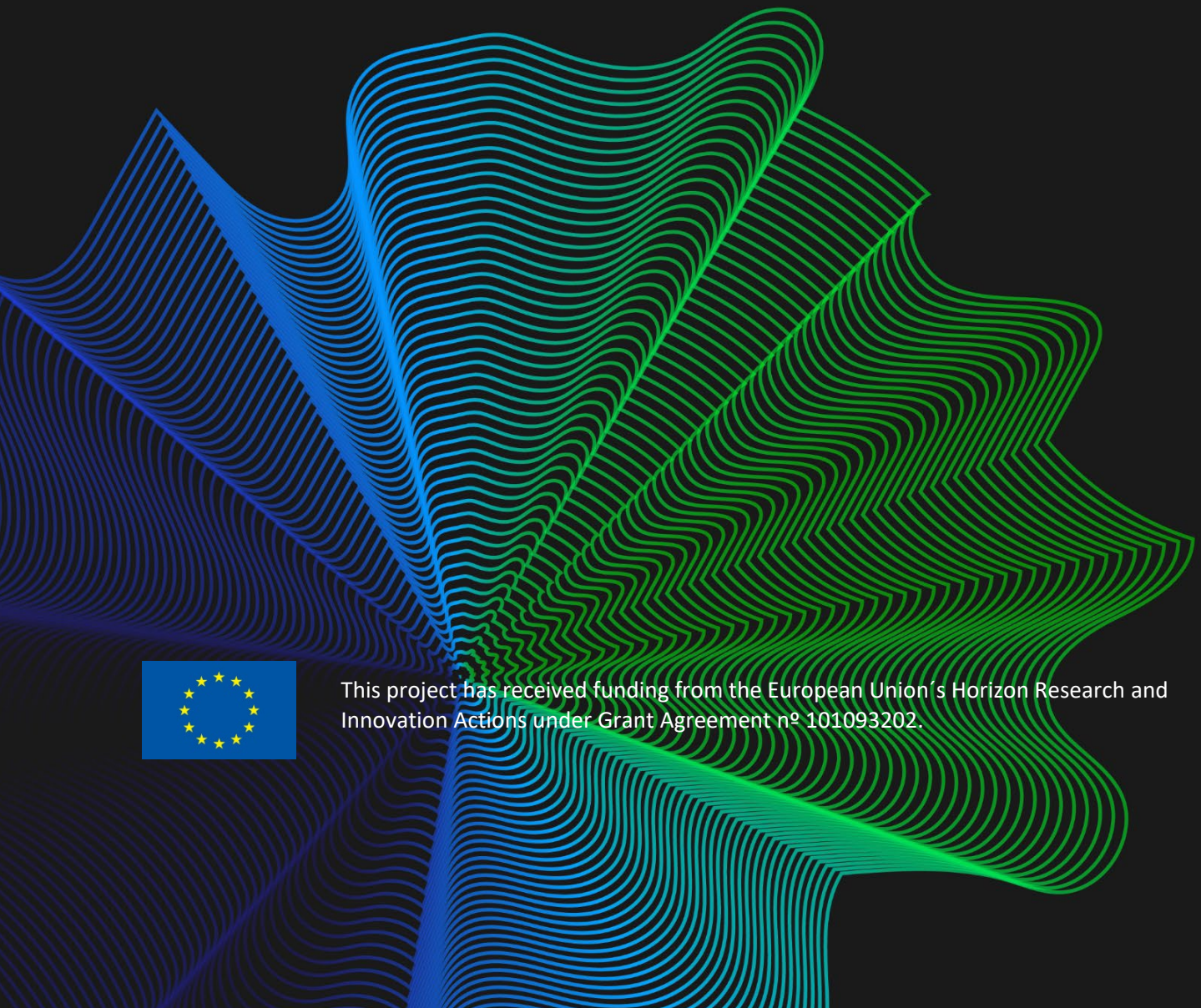
GRAPH MASSIVIZER

D7.1: COMMUNICATION, DISSEMINATION, AND EXPLOITATION STRATEGY



This project has received funding from the European Union´s Horizon Research and Innovation Actions under Grant Agreement nº 101093202.

Title	Document Version
D7.1 Communication, Dissemination, and Exploitation Strategy	1.0

Project Number	Project Acronym	Project Title
101093202	Graph-Massivizer	Graph-Massivizer

Contract Delivery Date	Actual Delivery Date	Deliverable Type* Security*
M3 (March 2023)	M3 (31st March 2023)	R-PU

*Type: *R – Document Report; *DMP – Data Management Plan; *DEM — Demonstrator, pilot, prototype; *Ethics: *Other.
**Dissemination Level: **PU - Public — fully open; **SEN - Sensitive — limited under the conditions of the Grant Agreement; **EU classified —RESTREINT-UE/EU-RESTRICTED, CONFIDENTIEL-UE/EU-CONFIDENTIAL, SECRET-UE/EU-SECRET under Decision 2015/444.

Responsible	Organisation	Contributing WP
Nuria de Lama	IDC	WP7

Authors	Organisation
Nuria de Lama	IDC
Richard Stevens	IDC
Alba Balla	IDC

Abstract
D7.1 provides the strategy and initial operational plan for the communication, dissemination and exploitation activities and the evaluation framework of Graph-Massivizer. The document also reports the preliminary outcomes and activities developed by partners until M3 in those areas.

Keywords
Marketing, dissemination, business models, product positioning, exploitation strategy, planning, benefits, methodology, graph analytics, market research, customers, competition, social media.

REVISION HISTORY

Revision	Date	Description	Author (Organisation)
V0.1	07.03.2023	First version of the document (content structure and guidelines)	Nuria de Lama (IDC)
V0.2	18.03.2023	Draft version of content included for communication, dissemination and exploitation sub-sections	Nuria de Lama (IDC) Richard Stevens (IDC) Alba Balla (IDC)
V0.3	20.03.2023	Introduction, scope, impact booster and partial revision of content	Nuria de Lama (IDC)
V0.4	27.03.2023	Draft version for review, including updated section 3 (exploitation) and updates of D&C activities, proofreading, figures, and acronyms	Nuria de Lama (IDC) Richard Stevens (IDC)
V0.9	30.03.2023	Final refinements, text reduction and inclusion of reviewer feedback	Nuria de Lama (IDC)
V1.0	31.03.2023	Major revision by Project Coordinator	Radu Prodan (AAU)

This project has received funding from the European Union´s Horizon Research and Innovation under Grant Agreement nº 101093202.

More information is available at https://graph-massivizer.eu/.

COPYRIGHT STATEMENT

INDEX

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 PURPOSE AND SCOPE OF THE DOCUMENT

"D7.1: Communication, Dissemination, and Exploitation Strategy" is the first deliverable generated by workpackage WP7, followed by two updated versions at months 12 (M12) and M36. The goal of D7.1 is to provide the strategy of Graph-Massivizer concerning its positioning in the research and innovation (R&I) landscape and market and understanding its added value for its challenges. In addition, D7.1 describes how the project will create impact and the measures, activities and materials developed to maximise it. As such, readers will find here the strategic thoughts about marketing, communication and dissemination, exploitation, and go-to-market approach (that will work hand-in-hand for adequate support). Based on the objectives, an initial operational plan has been drafted. Finally, the document depicts the initial key performance indicators (KPIs) and the monitoring methodology so that corrective measures can be applied if the results do not align with the original expectations.

While the strategy should remain relatively stable, activities to support its implementation (the operational plan) may evolve and adapt as needed during the project. Since the information technology (IT) market evolves quickly, and stakeholders in the innovation landscape may also vary in the next three years, this project will adopt a flexible and agile approach. It will be ready to revise success indicators, channels, messages and activities.

Besides the strategy and operational plans for communication, dissemination and exploitation, this document reports some of the initial activities developed by partners until M3. Subsequent versions (D7.2, D7.3) will provide full reporting of activities and outcomes and the corresponding impact assessment at M12 and M36, respectively, together with revision of the plan, as needed.

Who should be interested in D7.1? This document can be a valuable tool for different audiences, starting with partners in the Graph-Massivizer project (internal communication, full awareness of branding guidelines, coordination in the implementation of strategy and alignment with other WPs) and, of course, external stakeholders that want to understand the dissemination, communication and exploitation measures that will be set up by the project to maximise its impact.

1.2 STRUCTURE OF THE DOCUMENT

D7.1 has three major chapters covering the following contents:

- **Definition and approach of the Graph-Massivizer Impact Booster**. Impact Booster is the name given to the strategic umbrella that will guide and govern the activities related to marketing, communication and dissemination on the one hand and exploitation and market readiness on the other. The chapter describes the concept behind the Impact Booster, its elements, their relationships and how impact will be assessed.
- **Exploitation and Business approach of Graph-Massivizer.** The chapter describes in easy terms the intention to release the project's added value for the current state-of-the-art technologies and market offering. It also includes a description of the potential customers that could benefit from the solutions and the methodology that will allow the project to engage potential validators and other stakeholders needed for a successful exploitation strategy.
- **Communication and Dissemination strategy and operational plan.** The chapter outlines the plans to support the different objectives of the project. We focus primarily on creating awareness, engaging a community of interested parties, positioning the project in the relevant research fields and making it distinguishable from other projects and solutions. The plan also contains activities to understand the potential acceptance of the solution and to gather feedback that allows the project to improve the original products. This chapter also exposes the KPIs as defined by our contract and the relative position of the project concerning them.

2 IMPACT BOOSTER

2.1 MAJOR EXPECTED IMPACT

WP7 defines a series of KPIs associated with communication and dissemination. However, a general overview of the quantitative expectations from all the project areas is needed to understand the value of Graph-Massivizer. The performance achieved by the project for the identified KPIs will be used to attract clients, investors, open-source contributors and academic users. As such, KPIs defined for the technical building blocks and use cases will feed our exploitation strategy and be used for marketing messages. Thus, monitoring the values concerning the original expectations becomes critical to ensure credibility and rigour.

Current graph processing platforms come with various limitations, ranging from high energy consumption, inefficiency, and lack of support for diverse workloads, models, languages, and algebraic frameworks to the difficulty of use for non-experts.

Graph-Massivizer addresses these challenges by delivering an integrated toolkit to support a climate-neutral and sustainable economy based on graph data. The project will develop five open-source software tools for high-performance, scalable, and sustainable graph processing, as well as an enterprise-class commercial version based on the metaphactory knowledge graph platform that tightly integrates the tools in an easy-to-use-and-deploy offering to reach a broader market share. The project will validate the innovative toolkit on four use cases that cover the economic, societal and environmental sustainability pillars to ensure applicability and scalability in real-world scenarios and the feasibility of commercial solutions on top of metaphactory: sustainable green finance, global environment protection foresight, green artificial intelligence (AI) for the sustainable automotive industry, and data centre digital twin for exascale computing.

In this context, major KPIs that highlight the value of the project results cover three dimensions: technical excellence and performance, technical use case results, and business use case impact. These KPIs add to the list of specific indicators to measure dissemination and communication, which are the core of Section 4.3. Table 1 compiles all the KPIs in these categories, added to the list of specific indicators for dissemination and communication, which are the core of Section 4.3. Contribution to societal and environmental impact is characterised by a combination of KPIs, assessed in the following versions of this deliverable.

2.2 WHAT IS THE IMPACT BOOSTER OF GRAPH-MASSIVIZER?

Impact Booster is the name given to the strategy and the operational plan with activities and assets that the project will use to maximise its impact. The term explicitly refers to boosting the project impact starting with the defined KPIs (as illustrated in Table 1), indicating that they will drive the development activities along the project duration. We acknowledge the relationship between different WPs and tasks. Thus, the booster builds upon them in a coordinated and aligned way so that tasks within this WP and others do not work in silos. While WP7 relies on four activities, namely *"T7.1: Dissemination and communication"*, *"T7.2: Exploitation and innovation management"*, *"T7.3: Community building, clustering, and standardisation activities"*, and *"T7.4: Implementation of open access and open-source policies"*, we see T7.3 as part of the dissemination strategy and T7.4 as a pillar of the exploitation and go-to-market approach. Thus, the Impact Booster brings all of them together comprehensively and interconnectedly.

2.3 IMPACT ASSESSMENT

From a methodological point of view, Graph-Massivizer will set up a process where all the KPIs will be monitored and revised as needed, including how they will be measured and the responsible partners to do so (benchmarks for technical KPIs will happen in the corresponding WPs). This process will be run annually for updates on all the indicators so that the consortium has a dashboard with complete and updated information for reference. The list may be expanded with data that may enrich the

process and lead to some final KPIs. For example, getting the commitment of a stakeholder to use and adopt a tool may be complemented by an intermediary – operational – steps like the number of organizations reaching out to showcase the tool, the number of demonstrations run for potential customers, the number of meetings held, etc. Surveys may be used to collect more subjective data for the process. Based on the evolution of the technical work of the project as well as a better vision of the business strategy, starting with this document, Graph-Massivizer will consolidate all the KPIs and provide the first measurements by M12.

Table 1: Graph-Massivizer impact KPIs (excluding communication and dissemination).

Extreme data ingestion, massive graph creation and storage	**KPI-1.1:** Create massive graphs of the scale of 10 billion vertices and 100 billion edges
	KPI-1.2: Achieve a streaming ingestion latency below 500 milliseconds for 95% of the data at 1,000s of new edges per second
Massive graph analytics and reasoning	**KPI-2.1:** Achieve sampling and analytics throughput scale of 500 Giga traversed edges per second
	KPI-2.2: Achieve streaming-graph machine learning and online reasoning performance of 3 million triples per second, retaining 90% of the performance for the last 10 million edges compared to batch algorithms
Graph processing workload modelling with performance and energy guarantees	**KPI-3**: Achieve 80% accurate performance and energy consumption models for two basic graph operations per project use case for at least two heterogeneous multiprocessors, four graph processing units and one field programmable gate array
Sustainable and energy-aware massive graph processing	**KPI-4.1:** Perform sustainability labelling for graph analytics of 100 individual nodes and devices by running benchmarking scenarios and considering current IT and energy-infrastructure data
	KPI-4.2: Demonstrate the possibility of a two-fold improvement in data centre energy efficiency and over 25% lower greenhouse gas emissions over the state-of-the-art basic graph operations processing for more than one use case workload and one general *"what-if"* scenario, considering Tier 1 and national suppliers' and public energy data
	KPI-4.3: Generate predictions and recommendations for sustainability-performance trade-offs, online and offline, with different accuracy, to support stakeholder decision making
Scalable serverless graph analytics over a codesigned secure continuum infrastructure	**KPI-5.1**: Achieve 70% faster graph analytics using codesigned hardware and Graph-Greenifier recommendations over core the commercial AliGraph or other open-source software solutions
	KPI-5.2: Achieve 40% faster runtime deployment compared to AWS CloudFormation
	KPI-5.3: Consume 30% less energy for extract-transform-load storage than Amazon Redshift and open-source software
Continuous development,	**KPI-6.1**: Release five integrated tools validated according to their KPIs (the ones listed above)

integration, and secure operation of the Graph-Massivizer Toolkit	**KPI-6.2**: Validate four relevant use cases according to their KPIs (see below)
Use Case 1 (UC1) **Green and Sustainable Finance**	**UC1-1: Green financial synthetic data multiverse**
	KPI-1.1.1: [1 –5] petabytes of synthetic financial data multiverse, calculated by extrapolating available (e.g., NASDAQ) financial data (4 terabytes per day) for the last 15 years with one to five equivalent synthetic universes
	KPI-1.1.2: 90% energy consumption accountability for extreme data creation
	UC1-2: Greener financial algorithms and better investments
	KPI-1.2.1: Increase return o investment (ROI) by a realistic 2% – 4%, considering mature financial algorithms optimised over the last two decades
	KPI-1.2.2: Increase excess return (alpha) by 1% – 2% with a quick ratio higher than 1.5, reflecting healthy investments with lower risk and higher returns
	UC1-3: Commercialisation and provisioning of the Green Financial Data Multiverse
	KPI-1.3.1: Identify 5 – 10 prospective customers by M36
	KPI-1.3.2: Generate quality synthetic data for at least six relevant securities, including stocks, futures, commodities, bonds, exchangeable trading funds, mutual funds, and currencies
	KPI-1.3.3: Increase ROI by 30% by commercialising the synthetic data two years after the project
	KPI-1.3.4: Generate 20% – 40% cheaper synthetic financial data than actual historical data, with a five-time increase in data volumes with improved quality and more diverse properties
Use Case (UC2) **Global Foresight for Environment Protection**	**UC2-1: Media news knowledge graphs for strategic foresight**
	KPI-2.1.1: Creation of a massive foresight graph of at least a hundred billion edges by ingesting extreme data
	KPI-2.1.2: Latency from event detection to foresight below 24 hours (7 – 30 times improvement)
	UC2-2: Geopolitical and business foresight for environment protection
	KPI-2.2.1: Commitment to using Environment Protection Foresighter by at least one public and one private sector stakeholder
	KPI-2.2.2: At least 25% of relevant insights implemented by expert stakeholders
Use Case 3 (UC3)	**UC3-1: Green manufacturing simulation**

Green AI for Sustainable Automotive Industry	**KPI-3.1**: 5% – 15% improved simulation accuracy, depending on the scale and complexity
	UC3-2: Green AI solutions for manufacturing lines
	KPI-3.2.1: Reach out to hundreds of IT stakeholders from the Gaia-X European Association for Data and Cloud and car manufacturing stakeholders (e.g., Catena-X alliance)
	KPI-3.2.2: 5% – 15% improved quality, time and manufacturing line downtime following standard KPIs for measuring the production process effectiveness
	KPI-3.2.3: Portable operational deployment at 10 BOSCH factories by M36
	KPI-3.2.4: 40% increased ROI by using Graph-Massivizer technology
Use Case 4 (UC4) **Data Centre Digital Twin for Sustainable Exascale Computing**	**UC4-1: Green data centre digital twin at exascale**
	KPI-4.1.1: 100% spatial interaction captured by the massive data centre massive graph
	KPI-4.1.2: 10% power usage effectiveness reduction in data centres
	UC4-2: Sustainable data centre digital Twin at exascale operation
	KPI-4.2.1: over 4,000 active users
	KPI-4.2.2: 10% lower power use efficiency
	KPI-4.2.3: 20% better utilisation, modelling two partitions with 90% critical node identification

3 EXPLOITATION OF GRAPH-MASSIVIZER

Effectively exploiting the Graph-Massivizer results will be ensured through task T7.2 on exploitation and innovation management. This task is divided into three macro activities:

- Design and elaborate an exploitation plan for the Graph-Massivizer Toolkit and its use cases, including market analysis, competitor, business models, and value chains;
- Identify exploitable products and derive business models tailored to their stakeholder needs;
- Prepare the exploitation and sustainability of the results through impact, market and intellectual property rights (IPR) assessments.

The approach is to begin early and define the market, the types of challenges the stakeholders in those markets experience in processing extreme graphs, and to ensure there is a match between those stakeholders, those needs and the tools and services that Graph-Massivizer is producing. Additionally, the task will ensure that the project partners can identify valuable results, understand the IP issues connected to the results, and exploit those results during the project's lifetime. The following sections explore the context for Graph-Massivizer and the challenges it can expect to find as it moves forward. The chapter addresses the offering and stakeholders expected to use these potential tools and services. It further defines the use cases that will demonstrate this suitability and describes how the project expects to further define the market for these tools.

3.1 CONTEXT AND CHALLENGES IN GRAPH-MASSIVIZER EXPLOITATION

Graph-Massivizer provides solutions in a growing market with apparently excellent exploitation potential. Graphs are a practical approach to data analytics that can analyze complex relationships and patterns between different data points. Traditional data analysis techniques usually focus on analyzing individual data points or sets of data points, whereas the graph-based analytics approach in Graph-Massivizer focuses on the relationships and connections between data points. Graphs are a powerful representation of data that can expose complex relationships between data points. Graph processing enables users to identify patterns and trends in data that might not be apparent through traditional data analysis techniques. Using graph processing, Graph-Massivizer's typical users will better understand the structure and behaviour of complex systems, optimise processes in their tools, and make better-informed decisions based on the insights from the returned information.

Even if graph theory originated in the 1800s, its use for practical, real-world problems did not become popular until the late 20th century, with the development of graph database technologies and social networks. In the early 2000s, companies like Google and Yahoo! began using graph analytics to improve their search algorithms and understand the structure of the web. This hype led to the development of powerful graph processing frameworks like Apache Graph and Apache Spark GraphX. Today, graph processing offers popular methods for analyzing complex systems and networks used in various applications, including social network analysis, fraud detection, recommendation systems, and supply chain analysis. As envisioned and addressed by Graph-Massivizer, massive data sets add significant computational and functional complexity. As the number of vertices and edges increases, so do time and computational resources required to perform graph analytics. Graph-Massivizer is producing tools to assist users, especially in this area.

3.2 GRAPH-MASSIVIZER TOOLKIT

The project will deliver the Graph-Massivizer Toolkit, comprising five open-source software tools and several graph datasets covering the sustainable life cycle of processing extreme data. The tools, focusing on usability (multilingual data ingestion and massive graph creation), automated intelligence (through analytics and reasoning), performance modelling, and environmental sustainability trade-off management (supported by data-driven evidence) spanning high-performance computing (HPC) systems and the computing continuum leverage the emerging serverless computing paradigm and state-of-the-art cybersecurity measures. The Graph-Massivizer Toolkit helps experienced and novice

stakeholders from large and small organizations capitalise on extreme data through massive graph programming and processing. The following diagram outlines the toolset functionalities corresponding to the operation and graph processing layers.

The conceptual layers displayed in Figure 1 create an automated, sustainable toolset for graph operation, processing, and resource management made up of the following components:

- **Graph-Inceptor tool** creates a domain-specific graph from provided input data;
- **Graph-Scrutinizer tool** performs graph enrichment, graph query, and graph analytics to provide an answer to the user;
- **Graph-Optimiser tool** provides a highly efficient implementation of the graph processing required by Graph-Inceptor and Graph-Scrutinizer and predicts the small-scale performance of this solution;
- **Graph-Greenefier tool** scales up the solution from Graph-Optimiser for massive scale and provides sustainability analysis and improvements;
- **Graph-Servelizer tool** deploys the solution for Graph-Greenefier and enables performance and sustainability monitoring.

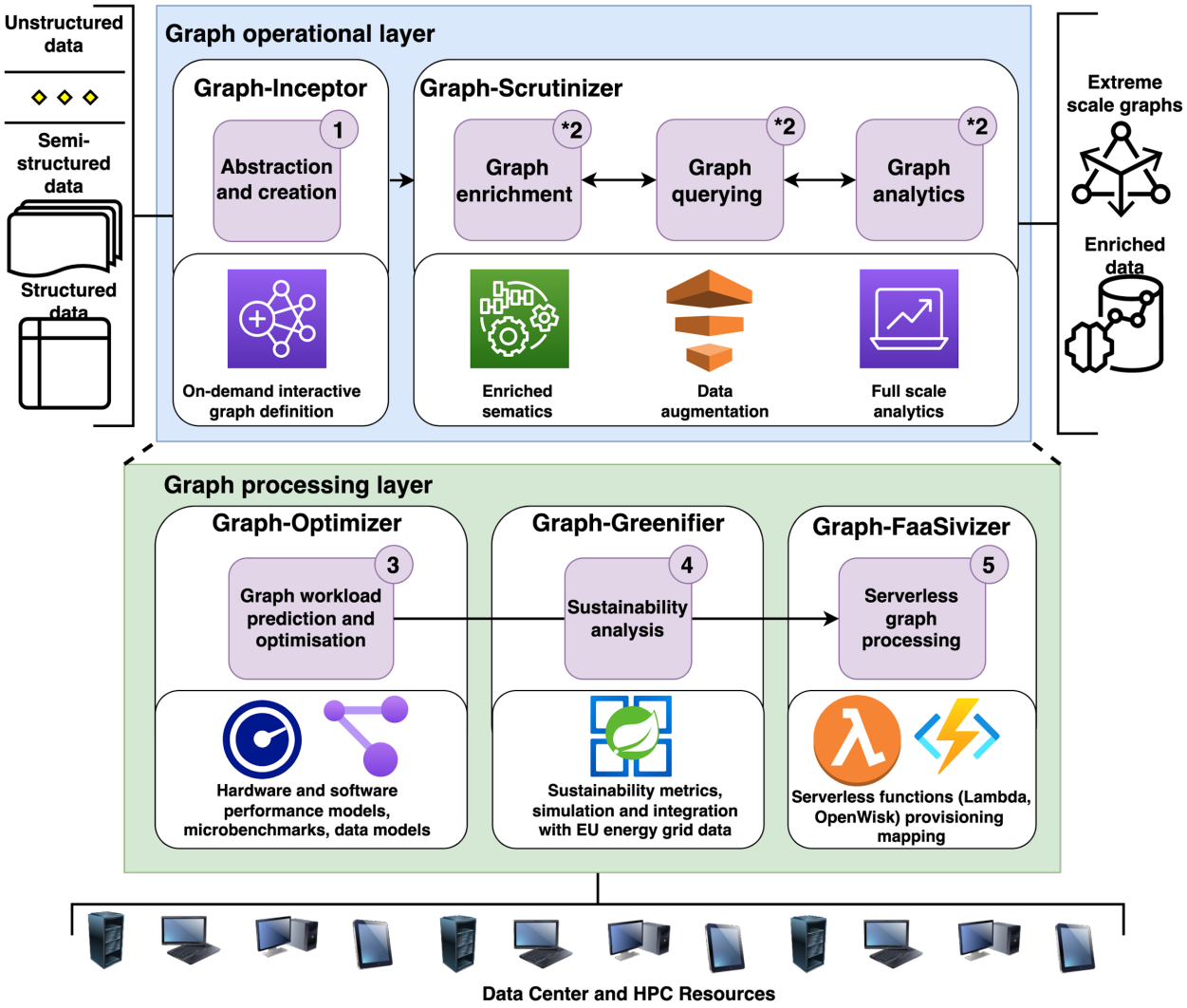


Figure 1: Graph-Massivizer Toolkit conceptual layers.

The hardware infrastructure layer considered by Graph-Massivizer consists of geographically distributed data centres across the Cloud, mid-range Fog, and low-end Edge computing continuum. The users interact with the deployed solution, configured to run across this continuum.

3.3 POTENTIAL USERS OF GRAPH-MASSIVIZER TOOLKIT

Graph-Massivizer has many potential users across industries and domains if applied and tailored to them. Some examples of potential users of graph processing applications include:

- **Social media companies** can use the Graph-Massivizer Toolkit to analyse user relationships and interactions, identify influential users, and detect and prevent (mis)information spreading, fraud, spam, or abusive behaviour.
- **Healthcare providers** can use the Graph-Massivizer Toolkit to analyse patient data, such as electronic health records, and identify patterns of disease spread or treatment effectiveness.
- **Financial services companies** can use the Graph-Massivizer Toolkit to detect and prevent fraud, identify relationships between financial transactions and entities, and analyse customer behaviour to improve targeting and segmentation.
- **Transportation and logistics companies** can use the Graph-Massivizer Toolkit to optimise supply chain networks, identify routes and connections between locations, and track the movement of goods and vehicles.
- **Government agencies** can use the Graph-Massivizer Toolkit to identify and track criminal networks, analyse patterns of social behaviour, and optimise resource and service allocation.
- **Retail companies** can use the Graph-Massivizer Toolkit to analyse customer behaviour and preferences, optimise store layouts and product placement, and identify cross-selling and upselling opportunities.

These are examples of the typical potential users of the Graph-Massivizer Toolkit. Our results will likely apply to many types of data and real-world problems. To demonstrate Graph-Massivizer's value to potential uptakers, it will develop, test, validate and showcase the toolkit on four real-world use cases.

UC1: Green and Sustainable Finance[1] is a solution to overcome the limitations of traditional financial market data and data providers. These providers often offer limited financial data volume, reduced accessibility, high costs, and limited historical relevance. Graph-Massivizer offers a fast and semi-automated process for creating realistic and affordable 'synthetic extreme financial data sets that are unlimited in size and accessibility. These extreme financial data sets (synthetic data multiverse) are leveraged to enhance AI-based green investment and trading simulations. Using synthetic data can eliminate critical biases such as prior knowledge, over-fitting, and indirect contamination resulting from historical financial data scarcity and limitations. UC1 has two main objectives:

- **Generate energy-efficient synthetic financial data (multiverse)** validated using standard (green) financial investment and trading algorithms against real historical financial data sets;
- **Use the Financial Data Multiverse** to improve green AI-enhanced financial algorithms with reduced bias, risk, and higher performance.

UC1 intends to significantly improve the performance and risk reduction of selected financial algorithms tested on the synthetic data multiverse produced by the Graph-Massivizer platform. This proof-of-concept will enable consortium members to commercialize synthetic data, allowing European investment professionals to improve testing and de-risking their financial algorithmic models, with a strong focus on green investment. The outcomes are crucial in meeting the demands of the financial industry for vast and reliable datasets while addressing the pressing issue of environmental sustainability. It can potentially transform how financial market data is generated and used.

[1] https://ec.europa.eu/environment/enveco/sustainable_finance/pdf/studies/Defining%20Green%20in%20green%20finance%20-%20final%20report%20published%20on%20eu%20website.pdf

UC2: Global Foresight for Environment Protection focuses on geopolitical and business aspects of sustainable environmental goals, including climate action, responsible production and consumption patterns, clean water and sanitation, and clean and affordable energy. The foresight comprehends insights on future trends and scenarios to guide decision-making in developing better policies.

- **Build** a contextual graph through data from the Common Crawl, Linked Open Data Cloud, and global media news provides unique insights into the convergence of mass media's three societal systems (economy, politics, and science);
- **Develop** complex data and operations to extract causal templates, understand future events, and predict consequences;
- **Extract** geopolitical and business foresight concerning global warming and the environment are essential for timely decisions, maximising impact, and mitigating negative scenarios.

UC2 will deploy, test and validate all of the tools in the Graph-Massivizer Toolkit.

UC3: Green AI for Sustainable Automotive Industry will leverage new graph-based encoding that captures several value-chain stages to predict their outcome better and detect anomalies. The automotive industry needs better and quicker analysis to prevent defect propagation and unnecessary waste and enable a more sustainable, circular, and climate-neutral automotive industry. In this context, Graph-Massivizer will:

- **Combine graph-based ML methods with digital twins** providing new insights and boosting the efficiency and scalability of the diagnosis beyond that of more expensive alternatives- **Predict *"best"* production configurations**;
- **Express** an automotive manufacturing welding machine digital twin in **OWL 2 ontology**;
- **Use the manufacturing graph generator** with extremely controlled scaling of a manufacturing graph in multiple dimensions (number of sensors, production lines, processes).

The insights gained will help to optimise manufacturing operations and improve the operational quality of the resulting products**.**

UC4: Data Centre Digital Twin for Sustainable Exascale Computing will improve energy **consumption** by improving heat dissipation by managing spatiotemporal dependencies between the cooling equipment, computing nodes, and computational workloads. Additionally, the use case will attempt to improve **system utilisation** by creating digital twins of performance data linked to **power usage effectiveness and greenhouse gas emissions.** It will:

- **Develop a digital twin of system parameters** to capture the spatiotemporal-ontological dependencies among computation, computing nodes, and cooling equipment and predict the impact of the spatial power distribution on cooling efficiency and cost;
- **Represent and validate an open data model** of the Marconi 1000 and EuroHPC Leonardo supercomputers at exascale.

UC4 will deploy, test and validate all of the tools in the Graph-Massivizer toolkit.

The consortium partners have identified the toolkit as the project's primary exploitable result, and its use, demonstration and validation in the use cases will be a central vehicle for exploitation. Users from the use cases and all the partners developing, testing and validating the tools and creating new knowledge will each have a different expectation as to how to extract value from their participation in the project, from their input to the project and from the products or services that they could exploit after the end of the project. The following section describes how we intend to perform individual exploitation planning.

3.4 PLAN FOR EXPLOITATION ACTIVITIES

3.4.1 ACTIVITY 1: GENERIC JOINT EXPLOITATION PLAN

The first step in the process to ensure adequate exploitation of the project results is to define a **generic joint exploitation plan** for the products and services that might be available for exploitation after the project and to define a **high-level business plan** for different possible and comparable exploitation paths to choose the most favourable result. We are not putting the results into any specific partners' reality. Instead, we perform a generic business analysis to check if someone wanted to *"sell"* the project results, what they would sell, and how they would do this to maximise the potential offered by the **Graph-Massivizer Toolkit**. To start defining a generic exploitation plan for the Graph-Massivizer Toolkit and its use cases, we analysed their requirements and examined potential solutions to meet these needs. The use case-driven analysis takes place in joint meetings with participants of task T2.1 and will be released in the deliverable *"D2.1: Graph-Massivizer Requirements Elicitation and First Architecture Design"*. Further, based on the use case requirements and interviews with stakeholders from industry, research labs, and academia, we will assess the functionalities expected to be included in the toolkit. Therefore, following M6, we will have a clear definition of the actual components – i.e., the tools and services that make up the Graph-Massivizer Toolkit. Examining this toolkit as the primary exploitable result, we will explore competing business cases based on competing market scenarios. We have begun to discuss existing business models, debate, and determine additional business models to exploit Graph-Massivizer results. In an initial workshop, we discussed the types of business models used or known to the group and stakeholders from the task T2.1 consultations. These are common business models in other business and data analytics domains. Still, they are not definitive as we do not know precisely how the toolset will look in the deployable configuration. They include:

- **Subscription license model**: This is a common business model used by data analytics software providers. The customers pay a monthly or annual subscription fee to access the software features;
- **Perpetual license model**: The customers pay a one-time fee to purchase the software license and can use it indefinitely. However, they may need to pay for upgrades and support;
- **Freemium model**: The partners expecting to exploit would provide part of the software modules for free, but some premium features would only be available to paying customers. This model could help attract more users and upsell premium features to them;
- **Consulting and professional services model**: Some data analytics software providers offer consulting and professional services to help customers implement and use the software effectively. They generate revenue by charging fees for these services;
- **Partnership model**: Data analytics software providers may partner with companies to integrate their software into other products or services. They can generate revenue by charging licensing fees or receiving a percentage of the revenue generated by their software;
- **Open-source model**: Some data analytics software providers offer their software under an open-source license, allowing users to use and modify it for free. They can generate revenue by offering support, consulting, or additional features to users who need them;
- **Hybrid Open Source/Licensing mode**l: some parts of the Graph-Massivizer Toolkit will be open-source, thus ready to use and modify, and other parts of the Graph-Massivizer toolkit (domain-dependent) will be under a type of licensing to be determined.

3.4.2 ACTIVITY 2: DETAILED JOINT AND INDIVIDUAL EXPLOITATION PLANS

We must answer the questions of *"who"* will exploit results in a joint exploitation model and *"how"* this group would do it. Partners will assess potential configurations for exploitation and possible commitments to jointly exploit the results. Potential preliminary suggestions from the group include:

- **Licensing**: Licensing the results to a third party for a fee can be a practical approach where partners (like the universities in Graph-Massivizer) do not have the possibility or do not want to commercialize the results themselves.

- **Spin-off companies**: As with licensing, universities or industry researchers might create a spin-off company to commercialize the results from Graph-Massivizer. It is a mechanism to maintain independent control over the exploitation of results while benefiting from any profits generated.
- **Joint ventures**: A joint venture could include Graph-Massivizer partners (or a subset of those) partnering with a private company or organization to exploit the results together. They could be helpful when the Graph-Massivizer partner might not have the resources or expertise to commercialize the results independently.
- **Patents**: Graph-Massivizer partners may register patents to protect the results and prevent others from exploiting them without permission. Patents can be helpful when the results have significant commercial potential, which might improve the chances of success of the other mechanisms described above.
- **Open access**: Graph-Massivizer partners may consider making the results freely available to others for non-commercial use, useful when the results are not expected to generate significant profits but can still be helpful to others. It can be a way to develop the *"dependency"* of users on Graph-Massivizer tools and lead to licensing and joint venture opportunities as dependent organisations move towards commercialisation of their products and services.

Although a selection process is neither imminent nor urgent, the nature and scope of Task 7.2 are to prepare the consortium well to make such choices. The choice of method for the joint exploitation of results in Graph-Massivizer will depend on the nature of the results, the goals of the researchers, and the resources available to them. The business model prepared by each partner will be personalised and include **market analysis**, **competitor landscaping**, **business model canvassing**, and **value chain analysis**. Each of the partners will be assisted in identifying the particular results they intend to integrate and understand exactly which market segments they are addressing, the tools with which they will be competing and the impact they will have on the value chain.

Furthermore, it is understood that not all partners will have *"commercial"* interests in exploiting Graph-Massivizer. For example, a university may be interested in exploiting knowledge to create advanced training and educational programs. Commercialisation will also have associated costs, such as resources to translate research into teaching curricula and benefits, such as increased enrolment in the university. Thus, all partners will be guided to perform a **Strengths, Weaknesses, Opportunities and Threats** analysis to understand what value they achieve and then perform the same **Cost/Benefit** and **exploitation canvas exercises**.

A first version of the resulting joint and individual exploitation plans will be released at M12. The plans will not be expected at a readiness level that is robust enough to approach investors but will guide the partner towards macroscopic vehicles for exploitation that may require extensive time and resource investments, such as, for example, setting up a spin-off company.

3.4.3 ACTIVITY 3: SUCCESS STORIES

We will develop **five hypothetical success stories** before M16 to demonstrate how the Graph-Massivizer Toolkit and adjacent project results could hypothetically be exploited. These stories are expected to be used as a model to help the specific class of partners (technology partners, educational institutions, analysts and commercial enterprises) understand the end-to-end process for exploiting results by seeing an example of how someone like them could have exploited results in a similar situation. Each success story will contain a realistic illustration of the sector, product or service description, collaborators, business models adopted, costs and revenues (or savings or process/quality improvement or new enrolments etc.), achieved financial results and a description of the impact and its calculation. We will liaise with marketing specialists from the communications teams to make an example of a commercial-grade marketing pitch to describe how the specific use case success story could describe the result(s) in 200 words or less and convince someone to use or buy the results or to fund on-going research or enrol in a technical workshop or course.

3.4.4 ACTIVITY 4: EDUCATIONAL RESOURCES

The task will hold an **IPR seminar** before M24 to help partners understand the concept of IP and related issues; the seminar will help partners identify the other partners with some claim to the IPR for any given result. We will supply descriptions, suggest methods and templates manage the relationship between the *"owners"*, and provide mechanisms to protect IP resulting from the project to make sure partners can exploit it in the future (this may include patenting, copyrights, trademarks, or perhaps the opposite publication, open-source and open access).

We will also organize an additional **seminar on potential funding opportunities,** to be carried out before M26, to ensure partners understand how to calculate the resources they may need to complete effective exploitation, including resources to perform future development, promotion and maintenance of the results exploited. It will describe where this support, such as government grants, venture capital, or private investors, can be found.

Finally, we will organize a **legal and regulatory considerations webinar or seminar** to be carried out before M30 to help partners foresee the legal and regulatory environment for the project's results, including any ethical, privacy, or data protection concerns that may arise.

With the information, suggestions, and knowledge provided by the consortium, IDC will prepare an online Survey that partners will complete. The survey will resemble the innovation radar prepared by the European Commission. However, it will already contain most of the information regarding the exploitable results owners, impact potential, IP terms and conditions for use and exactly how it is represented in the Graph-Massivizer toolbox as it will draw from our Exploitation catalogue and descriptions of the individual exploitation plans containing an individual business plan highlighting financial and operational planning with a five-year horizon. The outcome of this task will also include a detailed plan (if the partners decide to do this) for joint exploitation considering the elements described in the paragraphs above. If this is not the case, this document will explain why joint exploitation is not possible or desirable. The Individual and Joint Exploitation plans will be included in Deliverable D7.3 – Final communication, dissemination, and exploitation report at M36.

4 DISSEMINATION AND COMMUNICATION IN GRAPH-MASSIVIZER

4.1 DISSEMINATION AND COMMUNICATION STRATEGY

The communication and dissemination (D&C) strategy is based on **three fundamental pillars**:

- **Understand the outcomes**, added value and benefits of the products and solutions generated by Graph-Massivizer and **comprehensively describe them and adapt messages** to the different constituencies and channels
- Create a clear **positioning of Graph-Massivizer in the research and innovation landscape**, build the right alliances, partnerships and clustering, create awareness and make the project well-known. Generate the necessary materials and set up multiple channels to reach the desired audience (in terms of diversity, representativeness and volume, e.g., critical mass)
- **Engage stakeholders needed for an effective impact** along the project duration (requirements completeness, validation of concepts and releases, potential extension of artefacts, intermediaries, potential customers) by looking at the market preparedness of the solution.

These pillars will allow the project to have clear and understandable messages that are translated conveniently based on different target audiences, make the project well-known in those fields and channels of interest (generate attraction and visibility in a very competitive and evolving IT landscape where everyone is overloaded with a massive amount of information) and identify, select and engage the right stakeholders in the suitable milestones and phases of the project to refine the project concepts or the corresponding implementations.

From a **timing point of view**, the strategy revolves around four different phases, defined based on the evolution of the project and the subsequent availability of results:

- **Phase 1 (M1 – M12):** Create awareness about objectives and expected results for all stakeholders;
- **Phase 2 (M13 – M24):** Increase potential impact by reaching out to target stakeholder groups;
- **Phase 3 (M25 – M33):** Collect results by leveraging the awareness raised by the project;
- **Phase 4 (M34 – open end)**: Valorise results by demonstrating the Graph-Massivizer products to selected audiences, including joint ventures and commercialisation beyond the project.

Coordination and alignment with other parts of the project are critical to the success of the dissemination strategy. Particularly intense collaboration is expected with the exploitation prospects, explaining why impact (KPIs along different project dimensions) and exploitation and market strategy have been presented before the D&C strategy and plans.

4.1.1 OBJECTIVES

The primary aim of the C&D strategy is to enhance the effectiveness of all the project activities. This objective has two components. Firstly, it aims to support all WPs to achieve their goals regarding engaging and involving the community. Secondly, it strives to spread awareness about the benefits and services generated by the project as widely as possible among the industry and society. Specific objectives for communication and dissemination in Graph-Massivizer will guide the planning of activities in the operational plan.

4.1.1.1 COMMUNICATION OBJECTIVES

- Provide a clear, appealing image of Graph-Massivizer and its results in easily accessible terms for the audience, highlighting sustainability and environmental aspects as especially valuable elements of the project;
- Achieve high awareness and attraction through a website and social media channels;
- Engage with the public audience and guide the Graph-Massivizer toolkit towards a product appealing to public consumption at special public events;

- Engage with citizens to create an active community of interest and integrate citizen science knowledge in the requirements and validation considered by the project activities;
- Generate general-purpose materials as presentation cards for the project and its partners.

4.1.1.2 DISSEMINATION OBJECTIVES

- Adapt messages and channels to the needs and requirements of different constituencies to achieve higher penetration and impact, especially for critical stakeholders to the adoption of Graph-Massivizer results;
- Prepare communication material adapted to the target groups for dissemination and exploitation of results to create awareness of the project among the stakeholders;
- Establish liaisons with other projects and initiatives for knowledge and innovation transfer;
- Achieve results recognition among research communities, industrial and other standardisation working groups, potential customers, and policymaking institutions;
- Support the project's exploitation strategy, which may require a specific focus on validation, trials, business demos, penetration and activities targeted at special markets and countries.

4.1.2 GRAPH-MASSIVIZER POSITIONING

Our positioning strategy will include work in **two axes: research and innovation ecosystem, and business and market positioning**. While the offering is the same, the positioning work in the two environments will allow us to define more customised messages and get better results. Some companies operate in both, but the reality is that the departments/profiles and the objectives they pursue are usually different, as seen from the distinction between Research/ Academia and Industry audiences defined in section 4.1.3. While the first group is interested in lower technology readiness levels, the second one will concentrate on higher ones, competitiveness in the market, value and benefits of the minimum value product, business models, performance, scalability, support, and quality of documentation or licenses. Thus, depending on the context, major actors, influencers, competitors and *"customers"* (understanding this group as the ones *"buying"* what we offer) to consider are different.

4.1.2.1 RESEARCH AND INNOVATION LANDSCAPE

The main goal of Graph-Massivizer in the current innovation context is to make the project easily identifiable among a wide range of research actions and initiatives, detect major partners and projects where synergies could be exploited and use relevant initiatives as potential ambassadors of our messages and solutions (dissemination multipliers) or as vehicles to engage and attract potential parties for validation and endorsement (looking at market preparedness). Section 4.1.4 extends this part as a collaboration strategy.

4.1.2.2 BUSINESS POSITIONING

Showcasing the results of Graph-Massivizer in trade and industrial fairs can be a different game. Here, the most important thing is to demonstrate added value concerning the current market offering. This activity requires a good understanding of the market, including both supply (market trends, competitors, strengths and weaknesses of our solutions compared to other benchmarks) and demand (degree of coverage of requirements and customer needs by Graph-Massivizer tools, potential market disruptions by creating new and emerging needs as a result of the application and adoption of the proposed technologies, etc.). This content is subject to the exploitation activities offered by the project in task T7.2, as anticipated in Section 3).

Despite the high potential of graph analytics to be a game changer in the market, most companies are neither aware nor skilled to benefit from graph analysis yet, with many research centres/academia publishing and disseminating research work at a too abstract level to be easily accessed by industry,

and small and medium enterprises (SME) in particular. According to IDC[2], market saturation is less than 5%, with banking, financial and energy services taking the lead, but a more aggressive investment is expected in a three to five-year timeframe. Due to the limited offering of graph analytics tools, the profile of existing adopters is that of large enterprises that build their own. When a pronounced gap exists, it gets covered quickly. Companies are moving fast, and the market is getting dynamic with increasing investments in the field.

At the time of writing this deliverable, several important players in graph processing offering various graph databases, graph analytics frameworks, and graph processing libraries exist:

- **Neo4j** is one of the leading graph database providers, offering a highly scalable and performant graph database solution that many organizations in various industries use. This graph database is designed for handling large-scale graphs, providing a highly scalable and performant graph database solution that can handle billions of nodes and edges. It also provides a query language called Cypher, which can be used to perform graph queries.
- **TigerGraph** is another leading graph database provider. It offers distributed graph database designed for real-time graph processing, providing a highly scalable and performant graph database solution that can handle trillions of nodes and edges. It also provides a query language called GSQL, which can be used to perform graph queries.
- **DataStax** offers a graph database solution called DSE Graph, built on top of the Apache Cassandra distributed database providing advanced graph traversal and analytics capabilities.
- **Amazon Web Services** offers various graph processing services, including Amazon Neptune (a fully managed graph database service) and Amazon Managed Apache Neptune (a fully managed Apache TinkerPop graph database service). Amazon Neptune is a fully managed graph database service that handles large-scale graphs. It provides a highly available and durable graph database solution that can handle billions of nodes and edges. It also supports the open-source graph query language Apache TinkerPop Gremlin.
- **Microsoft** offers Azure Cosmos DB, a globally distributed, multi-model database service that supports graph databases. It provides a highly available and scalable graph database solution that can handle billions of nodes and edges. It also supports the open-source graph query language Apache TinkerPop Gremlin.
- **Google** offers Cloud Spanner, a globally distributed relational database that supports graph queries, and Google Cloud Bigtable, a NoSQL database for graph processing.
- **Apache Software Foundation** provides several open-source graph processing frameworks:
    - Apache Giraph built on top of open-source Apache Hadoop provides a scalable and fault-tolerant way to process large-scale graphs using the bulk synchronous parallel model;
    - Apache Spark GraphX built on top of open-source Apache Spark provides a distributed engine that supports various algorithms for the efficient processing of large-scale graphs;
    - Apache TinkerPop is a graph computing framework for graph databases and analytics.

The ambition of Graph-Massivizer is to be a frontrunner in energy efficiency and sustainable graph processing on heterogeneous computing infrastructures, adapting to the computing continuum. In particular, *Graph-Massivizer targets 70% more efficient analytics than AliGraph, and 30% improved energy awareness for extract-transform-load storage operations than Amazon Redshift. Furthermore, it aims to demonstrate a possible two-fold improvement in data centre energy efficiency and over 25% lower greenhouse gas emissions for basic graph operations.*

[2] IDC TechBrief: Graph Databases (June 2021)

The updated version of the Graph-Massivizer exploitation plan in M12 will provide a detailed market analysis reflecting the evolution of these players and the supply and demand dynamics. C&D will evolve messages based on the achievements and performance of the project as well as the changing conditions of the market or the constituencies we aim to attract.

4.1.3 TARGET AUDIENCES

The general audiences identified by Graph-Massivizer for dissemination include the following **categories of stakeholders** that will imply different messaging and channels:

- **Research and academia**: Scientists and researchers working on the different domains tackled by the project. Their main interest is in the technologies, the progress beyond the state-of-the-art and the potential reusability of the artefacts and tools generated by the project. Here it is important to use scientific language and communicate through specialised channels. As part of this audience, we include the overall R&I project landscape, further detailed in the next section.
- **Business and industry**: IT and non-IT people may be interested in the project results. In this case, the scientific evidence is less important, provided a sound concept and solutions have a solid background. While technical profiles will be more attracted to functional and non-functional aspects coped with excellence and support, management-level profiles will pay attention to hard and soft requirements, infrastructures, costs, maintenance, IPR and potential benefits (translated, if possible, as an ROI). In particular, SMEs may benefit from reusing open-source software in their solutions, extending the products or getting engaged in an ecosystem that could be technically and business-wise attractive.
- **General audience**: general public can be very heterogeneous, and messaging will depend on the nature of the groups and communities, but in all cases, will be characterised by the need for less formal language, with technicalities translated into more understandable terms where the benefits get highlighted easily without the need of having deep technical knowledge. In the case of Graph-Massivizer, some specific groups that may be considered as part of this category are policymakers, funding bodies and the press.

However, the refinement of messages and specific activities will be affected by a higher-granularity segmentation of stakeholders according to their interest in different project parts and the potential benefits they may get from the proposed solutions. For example, within the industry, the range of stakeholders will be quite wide, including specific users in the vertical sectors addressed by the different project use cases, as depicted in Section 3.3. Table 2 goes deeper into the benefits for different stakeholders.

4.1.4 COLLABORATION LANDSCAPE

As anticipated in Section 4.1.2, the project intends to develop alliances and partnerships and use existing channels with relevant projects that could help stay tuned about last technology advances and maybe collaborate with some of them in specific areas. In addition, these connections will be used to validate our results and engage external parties in the project's different processes for this purpose (e.g., three validation campaigns, five focus groups, and five external beta testing groups). Project awareness and dissemination will also be channelled through some of the instruments facilitated by these initiatives. Besides the involvement of specific individuals in the validation groups, we expect to drive forward agreements with some initiatives that may help in technology transfer as potential distributors of the solutions or as playgrounds for experimentation. We propose here a general overview of the relevant projects for clustering and networking purposes and characterization of the current evolving landscape of initiatives under consideration by Graph-Massivizer (operating essentially in the data, AI and computing continuum spheres).

Table 2: Target stakeholders for dissemination and main benefits.

Knowledge service providers • Faster and cheaper creation and processing of massive graphs • Easier management of graph analytical processes with no third-party intermediaries • Significantly increased opportunities for extensive graph data processing • Seamless, sustainable management of growing cross-sectorial graph data	**Data analysts** • Time and cost-effective processing of extreme data as massive graphs on heterogeneous infrastructure (e.g., on-premises, Cloud) • Seamless deployment and use of graph analytics on the computing continuum
FinTech Industry 4.0 • Mining, building, storing, querying, and analysing massive graphs for sustainable green finance and green automotive • Opportunities for inventing novel sustainable solutions	**DevOps integrators** • Increased energy awareness and transparent total cost of ownership • Sustainable analytics at exascale
Resource providers • Increased energy awareness and transparent total cost of ownership • Sustainable analytics at exascale	**Entrepreneurs** • New business opportunities related to innovative data services and applications • Seamless integration of constantly growing volumes and complexity of graph data
Policymakers • Faster and more effective decision-making based on cross-sectorial graph data	**Society** • Provide sustainable and eco-friendly success examples for green finance, environment protection, green manufacturing, and green computing industries • Involve citizens in beta-testing and crowdsourcing

4.1.4.1 PROJECT CLUSTERING

The focus of our work in terms of project networking and clustering will be on those projects funded under the topics illustrated in Figure 2. Of particular relevance are those addressing extreme data mining and analytics.

Figure 2: Portfolio of relevant projects for clustering.

NEARDATA (Extreme Near-Data Processing Platform) proposes a novel platform for data mining large and dispersed unstructured data sets that can be deployed in the Cloud and at the Edge (HPC, Internet of Things (IoT) devices), leveraging advanced AI technologies and offers a novel cybersecurity layer for trusted data computation. It features use cases in genomics, metabolomics and surgery.

SYCLOPS (Scaling extreme analYtics with Cross-architecture acceLeration based on OPen Standards) focuses on the AI acceleration market, particularly aiming to avoid vendor lock-in. Proposed AI and

data mining solutions address extremely large and diverse data, and AI acceleration follows open standards enabling a more innovation-driven ecosystem for Europe. This goal will be realised thanks to the convergence of the two industrial trends:

- standardisation and adoption of RISC-V, a free, open instruction set architecture for AI and analytics acceleration;
- emergence and growth of SYCL as a cross-vendor, cross-architecture, data-parallel programming model for all types of accelerators, including RISC-V.

EXA4MIND (EXtreme Analytics for MINing Data spaces) delivers an extreme data platform that combines large-scale data storage systems and powerful computing infrastructures by implementing novel automated data management and adequate data staging. The project aims to make big data from heterogeneous storage systems and database management systems available uniformly for processing on European pre-exascale supercomputers, cloud and AI-processing systems, bringing together the concepts of computing continuum with data spaces. It will be driven by cases in molecular dynamics, autonomous driving, smart agri-/viticulture and health/social big data.

EFRA (Extreme Food Risk Analytics) applies extreme data mining, aggregation and analytics to major scientific, economic and societal challenges associated with food safety and quality.

EMERALDS (Extreme-scale Urban Mobility Data Analytics as a Service) creates an urban data-oriented mobility analytics-as-a-service toolset to untap the potential of extreme urban mobility data.

EXTRACT (A distributed data-mining software platform for extreme data across the compute continuum) improves the lifecycle of extreme data mining workflows, significantly enhancing performance, energy efficiency, scalability and security while fulfilling the extreme data characteristics. The platform will be validated in two use cases: a personalised evacuation route service integrating many heterogeneous data sources and transient astrophysics with a square kilometre array pathfinder.

Main synergies rely on the ability to understand how to process and do analytics on extreme data that vary significantly from one case to another and, in particular, addressing the computing continuum as the underlying infrastructure that most cases will be based on due to the needs to combine centrally located computation with real-time at the edge. None of them addresses extreme graph processing and analytics, which could be an added value and a contribution from Graph-Massivizer to the ecosystem. Projects in the other boxes -even if funded under different topics may bring similarities or interesting links with our project. In addition, some partners participating in several projects will help to create the necessary hooks, as already depicted in some activities reported in the following sections.

4.1.4.2 LANDSCAPE OF EUROPEAN INITIATIVES

When looking at more general European partnerships to increase awareness about Graph-Massivizer and find interesting roots for validation and potential adoption, four areas come into the picture:

Data Analytics and AI. Main organizations that will be reached out to include AI, Data and Robotics Association (ADRA)[3] (as a contact point for activities associated with this partnership) and contributing entities (Ellis, EurAI, Claire, euRobotics). Big Data Value Association (BDVA) will be a priority because of the focus on data-related topics and the high involvement of some Graph-Massivizer partners (IDC, SINTEF, CINECA). We will check the opportunities behind AI4Europe[4] acting as the AI on-demand platform, where some assets could be helpful or share some through their contribution gateway.

Data Spaces. Besides data collection, analytics or storage, since a few years ago, there is a momentum around data sharing, with a myriad of actions intended to foster data-sharing practices keeping data sovereignty. The OPEN DEI project was the frontrunner initiative in publishing the *"Design principles for data spaces"*. Since then, several books, white papers and regulations pushed by the EC have been

[3] https://adr-association.eu/

[4] https://www.ai4europe.eu/

part of the playground. There is an interesting landscape of initiatives and projects specifically targeting the domain of data spaces. Graph-Massivizer could find exploitation roots with its tools considered as part of the so-called *"technical building blocks"*. Our strategy will focus on connecting with the Data Space Business Alliance and the operational arm of the data spaces part of the Digital Europe Program, the Data Space Support Centre (DSSC). Fortunately, some leading players are common in both, notably BDVA, GAIA-X, IDSA and FIWARE. GAIA-X may also focus more on standards and technical specifications (notice that one of our KPIs relates precisely to this entity). Thanks to the involvement of some partners of Graph-Massivizer in this field, access not only to those initiatives but also to some of the vertical data spaces (Great, Dates, EU Data Sp4ce) and technical projects contributing with added value technologies and platforms (Pistis, Upcast, etc.) is guaranteed.

Experimentation and deployment (*"test before you invest"*). One of the most interesting aspects will be capitalizing on existing infrastructures supported by public money to run part of our validation, which will provide feedback on our work leading to potential validation of the technical and business concepts. In addition, they could serve to attract users and adopters to the Graph-Massivizer tools. Initial targets to validate concrete experimentation approaches include the network of data-driven experimentation hubs created by BDVA (known as i-Spaces) and the flagship project EUH4D, which provides access to a federation catalogue of data, services and infrastructures. It is not by chance that hubs in the federation are labelled as digital innovation hubs (DIH). Our analysis will include a more comprehensive overview of opportunities within e-DIHs and testing and experimentation facilities.

Computing Continuum. The infrastructure aspect and challenges associated with seamless operation in the Computing Continuum is one of the areas of interest for Graph-Massivizer, as depicted in the analysis of *"sister"* projects above. We will target associations and initiatives that cover different angles of this picture: ETP4HPC and EuroHPC for concrete aspects of high-performance computing, while Alliance for Industrial Data, Edge and Cloud, the associated Important Projects of Common European Interest and GAIA-X (that covers not only the data spaces layer but also the underlying infrastructure as a network of federated computing nodes) plus possibly the SIMPL implementation will help us to address opportunities, partners and technologies in the Cloud-to-Edge-to-IoT area.

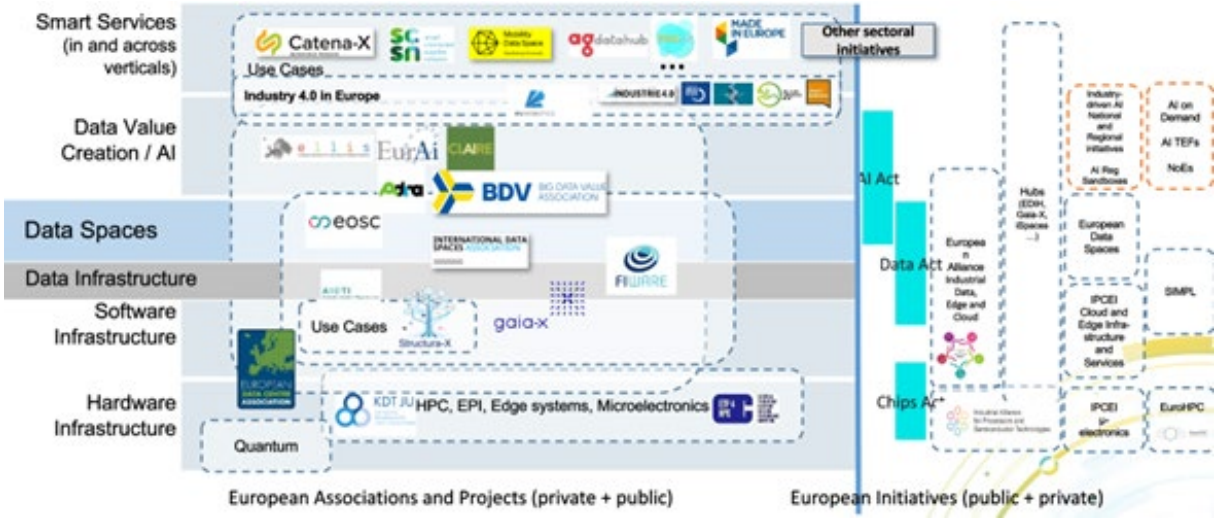


Figure 3: Sample of strategic digital/data initiatives in Europe (source: BDVA).

4.2 COMMUNICATION AND DISSEMINATION OPERATIONAL PLAN

The plan incorporates various communication tools like social media channels, newsletters, podcasts, videos, and printed materials. The project aims to elevate its C&D activities by actively participating in events, forming partnerships with other initiatives, and establishing press partnerships. Gradual

implementation of channels and other tools will be carried out through the launch, consolidation, and sustainability phases over the three years project lifetime.

4.2.1 METHODOLOGY

Communication and dissemination activities will shift their focus over time.

- **In the short term** (phase 1), we will insist on making the project known and establishing all the channels needed to reach our audiences and a critical mass for these channels;
- **In the medium term**, during the consolidation phase, we will focus on communicating about the ongoing activities and keeping the audience engaged;
- **In the long term**, during the sustainability phase, we will highlight in particular the impacts and results of the project, and we will emphasise the contribution of the Graph-Massivizer project in having a positive impact on EU society and economy, creating or supporting paths for exploitation.

Table 3 summarises the phases' duration, priorities and key milestones while each phase has its objectives and target group and performs the activities using the most suitable channels.

Table 3: Detailed D&C operational plan: phases, priorities and milestones.

Phase	Priorities	Milestones
Phase 1 (M1 – M12): Create awareness about the project objectives and expected results for all stakeholders.	• Establishment of tools • Establishment of all procedures • Focus on the use cases • Achievement of critical mass for the channels • Development of a community	• Launch of the website • Set up of social media account and achievement of a critical mass of followers • Publication of first blog items • Organisation of first events
Phase 2 (M13 – M24): Increase potential impact by reaching out to target stakeholder groups	• Focus on advertising use case results and services • Campaign for each use case	• Availability of several press partnerships and alliances, leveraged based on the need • Critical mass of material on the website (blog posts, podcasts, etc.) • Participation in various key events
Phase 3 (M25 – M33): Collect results by leveraging the generated awareness raised by the project	• Focus on disseminating information in best practices, lessons learned, impact and results	• Critical mass of material on the website (blog posts, podcasts, etc.) • Participation in various key events
Phase 4 (M34 – open end): Valorise Graph-Massivizer products to selected audiences, including joint ventures and commercialisation activities beyond the project	• Focus on disseminating information in best practices, lessons learned, impact and results	• Participation in various key events

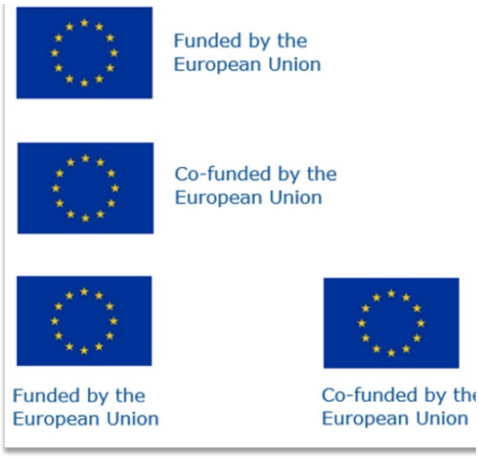


Figure 4: **EU emblem.**

Figure 5: **Graph-Massivizer project logo (versions).**

4.2.2 GRAPH-MASSIVIZER BRANDING

The Graph-Massivizer project branding was developed during the project's first month. The Graph-Massivizer logo was created using the Space Age typeface. A sans serif typeface that enhances the values of innovation, technology and modernity. The second sans serif typeface, Byom, refers to technological elements used for *"Massivizer"*. The letter *"A"* for the graph represents the data ecosystem. This concept is reinforced by the different circles surrounding it. Finally, the gradient is intended to reflect the technological and digital scope of the project. In addition, the green colour conveys the project's fundamental sustainability and energy pillars. Annex I includes a sample of the branding guidelines generated for coherence using visual elements during the project.

As per the Grant Agreement, all communication and dissemination activities related to Graph-Massivizer must acknowledge EU support by displaying the EU emblem (Figure 4), the project logo (Figure 5), and a clear statement indicating that the project has received funding from the EU. This requirement applies to various activities, including media relations, conferences, seminars, and information material such as brochures, leaflets, posters, and presentations, whether in electronic form or traditional media, as well as dissemination activities and any infrastructure, equipment, vehicles, supplies, or significant results funded by the grant.

4.2.2.1 Project Templates

In the first phase of the Graph-Massivizer project, several templates were created for internal and external use, as detailed in the Project Handbook[5] (Figure 6). The purpose of these templates is to provide partners with a structured framework for creating project documents and ensuring consistency in the overall visual identity of the project. By utilizing these templates, partners can effectively communicate the project while complying with EU funding visibility requirements and following specific visual identity guidelines.

The PowerPoint template for project presentations (Figure 7) features various slides that partners can use in any public or internal presentation. This template is designed with branding guidelines, project reference numbers, and logos. Partners can access it through the WP7 channel in the project's ownCloud repository.

Another template for project deliverables was created and shared among partners during the first phase of the Graph-Massivizer project. The Microsoft Word template adheres to the project's visual identity guidelines and includes the project's logo and reference number (Figure 8). The purpose of this template is to ensure that all project deliverables have a consistent and recognizable layout. The project's document template is available in the WP7 channel of the project's ownCloud repository.

[5] https://graph-massivizer.eu/resources/

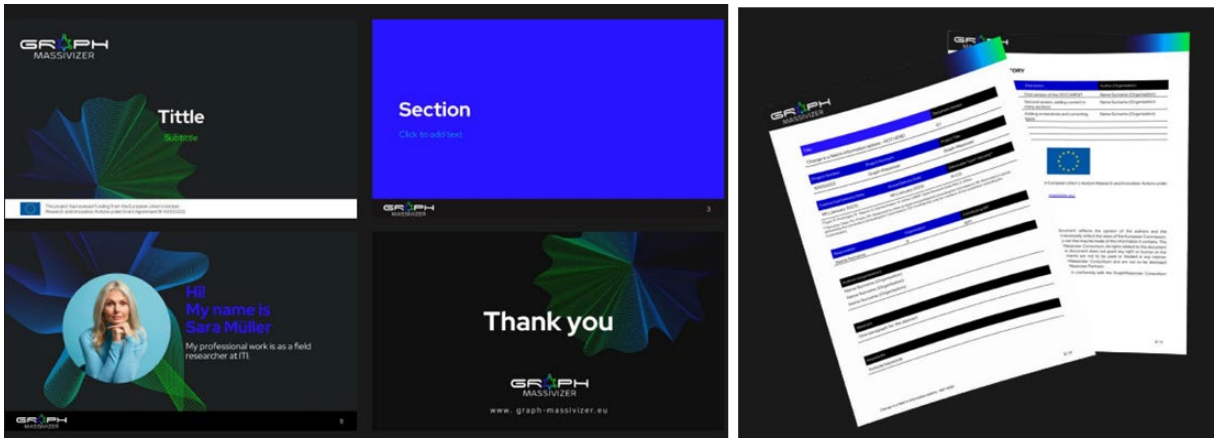


Figure 6: Project templates.

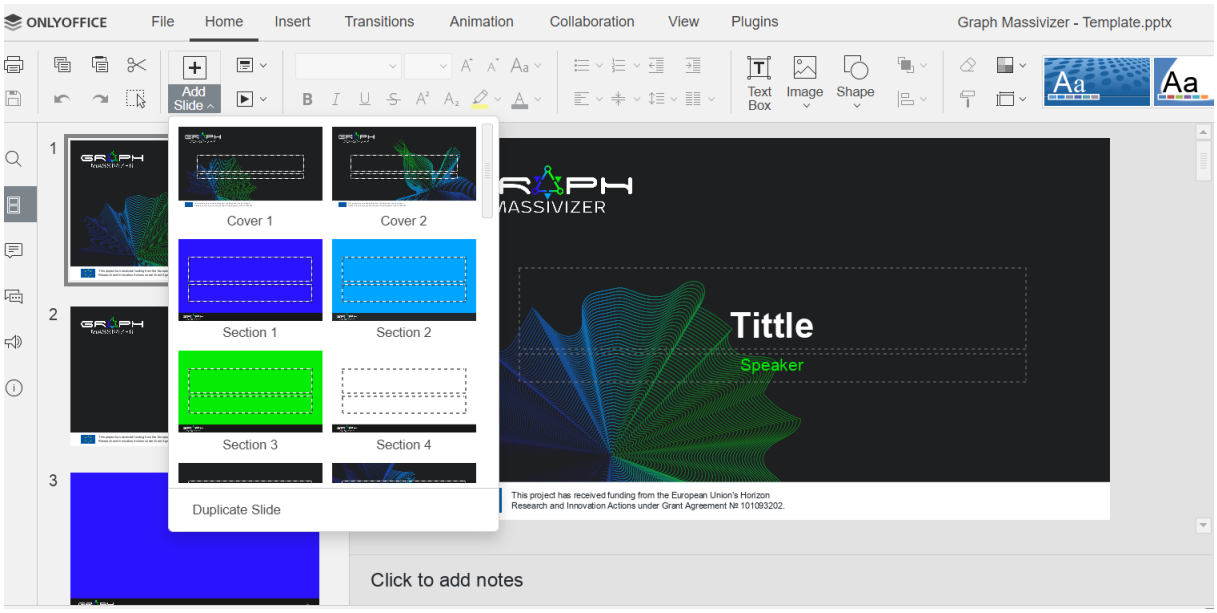


Figure 7: PowerPoint presentation template.

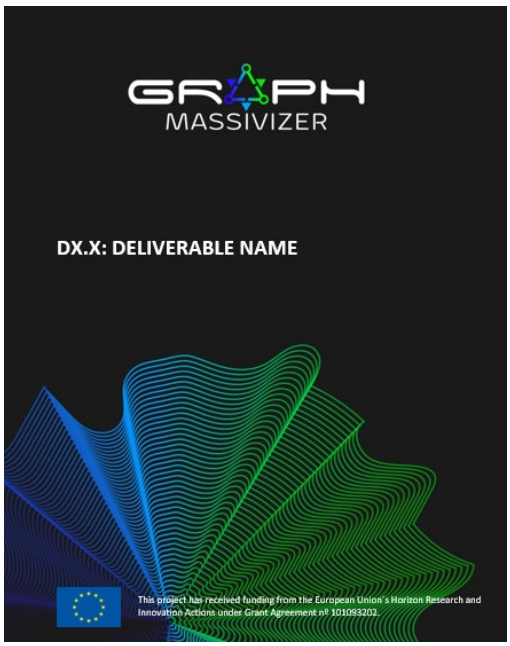


Figure 8: Word deliverable template.

4.2.3 GRAPH-MASSIVIZER CHANNELS

In the initial three months, the primary project communication channels were created. Graph-Massivizer employs a comprehensive strategy for communication resources, utilizing various digital tools. WP7 will enhance these channels in the upcoming months, considering the strategic objectives and striving for a significant impact.

4.2.3.1 GRAPH-MASSIVIZER WEBSITE

The Graph-Massivizer website[6] is a crucial communication and dissemination channel for the project and serves as the primary digital platform for the project activities, news, events, and promotional material. The website is designed to engage all target audiences and stakeholders and will be periodically updated and refined to meet specific requirements and functionalities. The homepage provides an overview of the project and links to social media channels. The website features all consortium members, including a contact card and information on their ecosystem and services. A section is dedicated to project use cases managed by the members, and a specific section is presented related to the project software tools. The *"News & Events"* sections contain relevant news and blog articles related to the project. The information about project resources and other relevant materials is gathered in the *"Resources"* section.

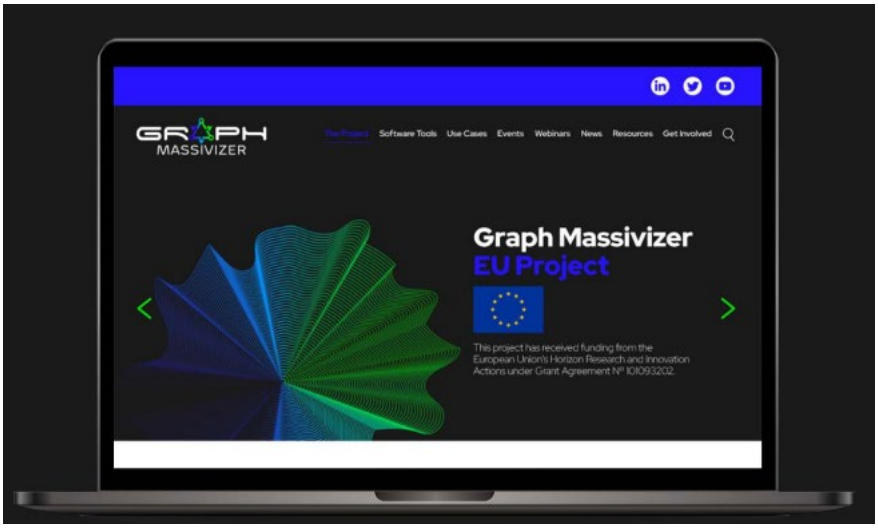


Figure 9: Graph-Massivizer website.

4.2.3.2 GRAPH-MASSIVIZER SOCIAL MEDIA

The Graph-Massivizer project has created a LinkedIn account[7] to communicate information about the project's activities and opportunities for stakeholders. This platform will also be used to explain the project's role and target other initiatives. Additionally, the project will advertise events and expand its database through LinkedIn. A project's Twitter account[8] has also been created to engage with the community and share news and outcomes. User engagement activities such as Twitter polls and the launch of hashtags will be carried out to establish a dialogue with stakeholders beyond the R&I community. The project has also set up a YouTube channel[9] to host all videos produced during the project, including promotional videos and recordings of webinars and events.

[6] https://graph-massivizer.eu/

[7] https://www.linkedin.com/company/graph-massivizer-project/

[8] https://twitter.com/graphmassivizer

[9] https://www.youtube.com/@graphmassivizer

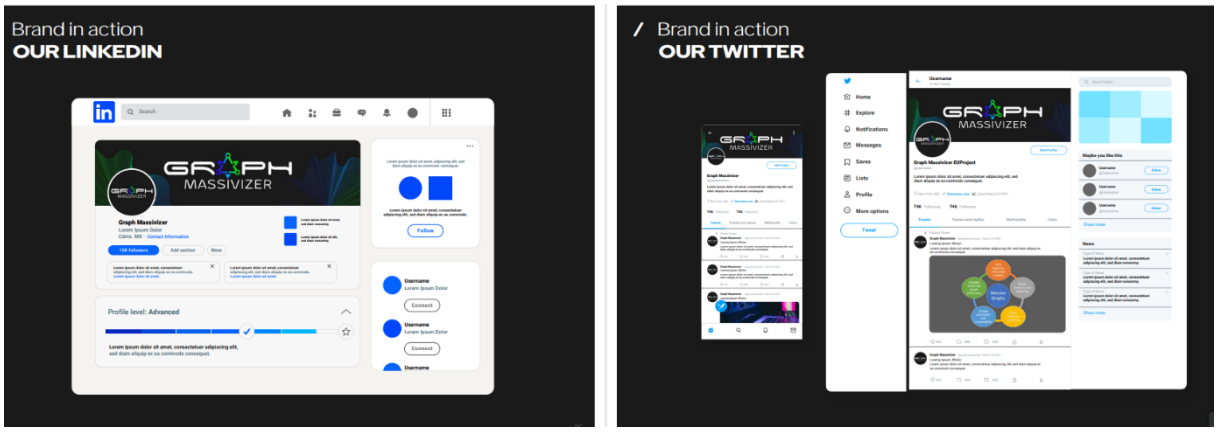


Figure 10: Graph-Massivizer: social media channels.

Website and social media channels will be an integral part of the communication campaigns that Graph-Massivizer will launch based on key milestones of the project (Figure 10). Internal work streams will be used to enhance the content as the project evolves and more resources become available.

4.2.4 EVENTS

Graph-Massivizer will be engaged in specialised project workshops to present results, interact with target stakeholder groups and attract scientific community interest. Partners will organise joint workshops with other EU projects at conferences where the target audience is present, aiming to align the project results with dissemination activities of similar initiatives. The project will be part of industrial fairs, exhibitions, and forums to demonstrate the Graph-Massivizer tools and the potential of its use cases, such as the Data Week, EBDVF, etc. While in the project's initial phase, we will pay greater attention to events in the R&I landscape due to the project's research focus, the second part of the project will shift to more industrially oriented events, including specific targets for the promotion of the use cases.

A list of typologies of events and gatherings for dissemination purposes is included here. The list will be periodically revised in the context of the dissemination group, and specific task forces will be set up for organizational aspects of those events that require more preparation and involvement.

- Target stakeholder groups gather numerous customers and collaborators of the project partners, including special focus groups for requirements elicitation brainstorming and beta-testers for validation campaigns (main focus of WP2)
- Specialised project workshops present results, interact with target stakeholder groups and attract scientific community interest.
- Industrial fairs, exhibitions, and forums present and demonstrate the Graph-Massivizer tools and the potential of their use cases, such as the European AI Alliance, AI & Big Data Expo, Big Data Expo, The Knowledge Graph Conference, Cloud Expo Europe and World Summit AI;
- Training, tutorials, and educational events organised as part of scientific conferences, summer schools, or during advanced MSc courses held by the academic partners (i.e., AAU, UTW, VU, UNIBO);
- Important scientific events and international conferences to present our work, exchange ideas, and discuss the latest trends in graph analytics:
  - **International Conference on Data Mining (ICDM**) is a premier research conference in data mining and includes topics related to graph analytics, such as social network analysis and graph-based machine learning;
  - **ACM SIGKDD Conference on Knowledge Discovery and Data Mining** is a leading conference in data mining and includes topics related to graph analytics, such as graph-based clustering and community detection;

- **IEEE International Conference on Big Data** is a leading conference in big data analytics and includes topics related to graph analytics, such as graph databases and graph-based machine learning;
- **International Conference on Web Search and Data Mining (WSDM**) is a conference that focuses on web search and data mining and includes topics related to graph analytics, such as link prediction and social network analysis;
- **IEEE International Conference on Data Engineering (ICDE)** is a conference that covers a wide range of topics related to data engineering and includes topics related to graph analytics, such as graph databases and graph-based data management;
- **International Conference on Advances in Social Networks Analysis and Mining (ASONAM)** is a conference that focuses on social network analysis and mining and includes topics related to graph analytics, such as community detection and link prediction;
- **International Supercomputing Conference and Supercomputing (ISC-HPC, SC)** focus on large-scale applications' high performance/efficiency;
- **International Symposium on Cluster, Cloud, and Internet Computing (CCGrid)** offers a forum for specialists in large-scale infrastructure and massive-scale data processing to discuss the latest technologies in the field.
- **International Conference on Cluster Computing (CLUSTER)** focuses on all the challenges around distributed cluster computing.
- **International Conference on Performance Engineering (ICPE)** is a forum for all performance engineering (including energy efficiency and sustainability) communities to discuss new developments in efficient and sustainable applications and systems.

- Events and conferences organised by associations and initiatives in the R&I landscape include flagship events of BDVA, GAIA-X, IDSA, and DSSC. Some of those include the **Data Week, European Big Data Value Forum**, **GAIA-X Summit**, **IDSA Summit**, and **Data Spaces Symposium**, most of them held annually.

4.2.4.1 GRAPH-MASSIVIZER EVENTSs

Graph-Massivizer will mainly work on organising specific workshops to satisfy the project's needs along the project lifecycle (requirements elicitation, validation of the toolkit and the different tools included, showcase of demonstrators, exploitation, etc.). Awareness and dissemination at a more general level will capitalize events, conferences and fairs organised by reference organizations in the fields of interest due to their ability to attract a wider or more specialised audience. It will also allow us to focus our efforts on the most relevant activities of the project and reduce the organizational aspects associated with setting up events, which is a challenging and time-consuming job.

Therefore, many of the events organised by the project will be categorised under third-party events hosted under a recognised or existing brand. It should not lead to neglecting the effort needed to propose, validate and organise such workshops.

The physical approach of these events will be combined with a series of webinars organised by the project. Topics and timings are still under discussion and will be published on our website, social media and other channels in advance.

4.2.4.2 THIRD-PARTY EVENTS

Most events organised by Graph-Massivizer will be integrated with well-recognised conferences and fairs, following the approach described before. The project has been very active from its inception phase before the contract entered into force, and a good amount of work can be reported already in M3, either as events already attended or as workshops under preparation.

Attended events (with the active participation of Graph-Massivizer) include:

- **2022 IEEE International Conference on Cluster Computing[10] (CLUSTER) (September 2022; Heidelberg, Germany)**. A scientific paper focusing on game theoretic provisioning of computing continuum resources to scientific applications was presented
- **IEEE Cloud Summit (October 2022; Mainz, Germany).** Graph-Massivizer was presented in the 2022 edition of the IEEE Cloud Summit[11] , including a peer-reviewed conference paper (check section 4.2.5). IEEE Cloud Summit is the flagship conference on Cloud with high-quality academic research presentations and industrial best practices.
- **European Big Data Value Forum[12] (October 2022; Prague, Czech Republic)** organised by the Big Data Value Association (BDVA), IT4Innovations (VSB‑Technical University of Ostrava), Plan4all and Charles University in Prague in collaboration with the European Commission (DG CNECT), the 2022 edition of EBDVF has been granted with the Czech Republic Presidency Auspice. From a thematic point of view, the event addressed big data and data spaces, illustrating how businesses can harness the power of Trustworthy AI and the role of high-performance computing as an enabler for digital transformation. Graph-Massivizer led the organization of a *"Fueling Industrial AI with Data Pipelines"* session to present the project ambitions and kick-start collaboration with other relevant EU projects.
- **High Performance, Edge and Cloud computing (HiPEAC)[13] (January 2023; Toulouse, France)** conference is the premier European forum for experts in computer architecture, programming models, compilers and operating systems for general-purpose, embedded and cyber-physical systems. Areas of focus and integration include safety-critical dependencies, cybersecurity, energy efficiency and machine learning. Graph-Massivizer performed a project presentation.
- ***"Get-to-know"* introductory workshop organised by the DSSC (February 2023; virtual).** The Data Spaces Support Center facilitated a workshop to introduce all the projects funded by the EC in areas of interest for data space deployment. The goal was for all the projects to know what their peers were doing and facilitate clustering and networking. Graph-Massivizer made a project presentation following the guidelines and will follow up on possible collaborations based on the strategy depicted in this document.
- **NESTiD Seminar 2022/23[14] (March 2023; Liverpool, UK).** Graph-Massivizer was presented at the be-weekly seminar talks co-organised with the Networks and Distributed Computing research group at the University of Liverpool as part of the Durham-Liverpool synergy.
- **Neuro-Symbolic AI and Graph Technologies seminar[15] (March 2023; Norway)**, part of the collaboration with the Norwegian Research Center for AI Innovation and OsloMet with a scientific audience of more than 50 attendees, fostered collaboration among DataCloud, enRichMyData, Graph-Massivizer, UPCAST, and NorwAI projects. Graph-Massivizer representatives from SINTEF provided an *"Introduction to Graph technologies in the era of AI"* and a presentation on *"Knowledge Graph storage systems: do they scale?"*. Some of these projects and other stakeholders are responsible for the upcoming Data Week edition initiative.

Upcoming events (with the active participation of Graph-Massivizer) include:

- **ICPE 2023 (15-19 April; Coimbra)** originated from merging the ACM workshop on software and performance prediction and a SPEC workshop focused on benchmarking and performance evaluation. This year, it will host a series of co-located workshops key to Graph-Massivizer.

[10] https://clustercomp.org/2022

[11] https://www.ieeecloudsummit.org/2022-program

[12] https://european-big-data-value-forum.eu/2022-edition/

[13] https://www.hipeac.net/2023/toulouse/

[14] https://nestid.webspace.durham.ac.uk/nestid-seminars/

[15] https://www.eventbrite.com/e/ai-meets-complex-knowledge-structures-neuro-symbolic-ai-and-graph-tech-tickets-574353345317

- **The First Workshop on Serverless, Extreme-Scale, and Sustainable Graph Processing Systems (GraphSys '23)(15-19 April; Coimbra)** focuses on high-performance large-scale graph processing, covering state-of-the-art and emerging (future) graph processing systems from various viewpoints. The workshop brings together the community of graph processing systems with the key presence of talks and presentations by the Graph-Massivizer project. Key project researchers are prominent names of the organizing and program committees. This event will be a unique occasion to confront the project's technical progress with the community, with 11 papers accepted as part of the proceedings (see Section 4.2.5).
- **Knowledge Graph Conference[16] (8–12 May 2023; Cornell Tech, New York, USA)** will be the largest ever, with over 30 workshops and tutorials, over 80 presentations by leading experts, eight discussion panels, and more, showcasing only the best of knowledge graphs. Metaphacs, as a sponsor of the event, will bring Graph-Massivizer to the attention of this important audience through the masterclass *"Semantic Knowledge Modeling for Domain Experts and Business Users"*.
- **Symposium on AI, Data and Digitalization (SAIDD)[17] (9–10 May 2023; Sogndal, Norway)** will explore how data and AI set the stage for the digital revolution and contribute to solving societal challenges. The symposium brings together researchers, policy experts, and practitioners across AI, big data and a range of application domains, thus, representing the views and activities of the policy, industry, research and academia. Graph-Massivizer will bring an Introduction to the green finance use case and a paper presentation on the strategic foresight use case.
- **1st Workshop on Hybrid Artificial Intelligence and Enterprise Modelling[18] (12 June 2023; Zaragoza, Spain)** brings together researchers and practitioners from machine learning, knowledge representation and reasoning (including semantic technologies), and enterprise modelling to reflect on how combining the three fields contributes to engineering intelligent information systems. An introduction to the Graph-Massivizer project will be featured at the workshop.
- **Data Week 2023 (13-15 June; Luleå, Sweden)** is a community event organised by BDVA and the EUH4D project, following the former BDVA Summit run by the BDVE project. While the first edition of Data Week in 2021 suffered from the virtual format imposed by the pandemic, the 2022 edition enabled the data community to come together and present main novelties in the data field. In 2023 an Open Call for sessions was launched following the edition's motto, *"Data meets infrastructure at the edge"*, attracting many players. Graph-Massivizer will organise a session titled *"Are current infrastructures suitable for extreme data processing? Insights, examples and discussion about infrastructure requirements and limitations of the Computing Continuum for extreme analytics, including sustainability and energy-efficiency aspects"*. The session is a joint effort between the projects Graph-Massivizer (leading) and DataCloud, enrichMyData, with a possible link with ETP4HPC/EuroHPC, data centres in the region and computing continuum initiatives. The main objectives pursued by Graph-Massivizer with this session are:

  - Present current needs of extreme data processing for infrastructures, elaborating on major challenges and opportunities for Europe;
  - Include energy efficiency and sustainability aspects as part of the problem equation;

[16] https://www.knowledgegraph.tech/

[17] https://www.bigdata.vestforsk.no/saidd2023

[18] https://hybridaims.com/

- Introduce ongoing work in the domain and propose solutions to overcome main limitations;
- Showcase examples that range from the use of HPC to computing clusters;
  - Identify stakeholders willing to engage in validation campaigns in the context of ongoing projects so that industrial needs are well aligned with the technical work proposed by the scientific community;
  - Converge towards an agenda of common themes for several projects and initiatives and can benefit from joint work to foster collaboration.

A full description can be found in Annex II.

Figure 11 Data Week'23: Data meets Infrastructure at the Edge.

- **European Big Data Value Forum (25-27 October; Valencia, Spain)** is the flagship of the ecosystem around data running for over ten years, connecting stakeholders engaged in BDVA. It normally gathers close to 1,000 participants networking with similar projects and discussing initiatives in the field. It is typically organised in the country that holds the European presidency, usually part of the official agenda. In the second part of 2023, it will be held in Valencia through the support of local organiser ITI, as Spain will take over the EU presidency. Graph-Massivizer will apply to the different engagement possibilities as soon as they are communicated.

4.2.5 SCIENTIFIC PUBLICATIONS

Graph-Massivizer partners achieved the following scientific publications in the first three months of the project:

- Towards Extreme and Sustainable Graph Processing for Urgent Societal Challenges in Europe (IEEE Cloud Summit)
- SimLess: simulate serverless workflows and their twins and siblings in federated FaaS (SoCC '22: Proceedings of the 13th Symposium on Cloud Computing)
- Matching-based Scheduling of Asynchronous Data Processing Workflows on the Computing Continuum (2022 IEEE International Conference on Cluster Computing (CLUSTER))
- Extreme and Sustainable Graph Processing for Green Finance Investment and Trading (SAIDD 2023 Symposium on AI, Data and Digitalization)
- AI-based Strategic Foresight for Environment Protection. Finance Investment and Trading (SAIDD 2023 Symposium on AI, Data and Digitalization)
- Graph neural networks for anomaly anticipation in HPC systems (Graphsys'23)
- Graph-Serverlizer: A Sustainable and Serverless Solution for Massive Graph Processing on Computing Continuum (Graphsys'23)

- AI, What Does the Future Hold for Us? Automating Strategic Foresight (Graphsys'23)
- Metaphactory for Massive Graphs (Graphsys'23)
- Semi-Supervised Prediction on Events with Neural Networks (Graphsys'23)
- Graph Greenifier: Towards Sustainable and Energy-Aware Massive Graph Processing in the Computing Continuum (Graphsys'23)
- Graph-Scrutinizer: Towards Massive Graph Analytics, Querying, and Reasoning (Graphsys'23)
- Large-scale graph processing and simulation with serverless workflows in federated FaaS (Graphsys'23)
- Graph-Inceptor: Towards Extreme Data Ingestion, Massive Graph Creation and Storage (Graphsys'23)
- Boosting the Impact of Extreme and Sustainable Graph Processing for Urgent Societal Challenges in Europe (Graphsys'23)
- Scaling the Data Value Quest: A Holistic Semantic Approach at Bosch (Graphsys'23)
- Incremental Multilayer Resource Partitioning for Application Placement in Dynamic Fog (IEEE Transactions on Parallel and Distributed Systems)
- Proactive SLA-aware Application Placement in the Computing Continuum (37th IEEE International Parallel and Distributed Processing Symposium (IPDPS))

4.2.6 OTHER ACTIVITIES

As part of the operational plan of Graph-Massivizer to generate awareness about the project, we will create agreements and collaborations with organizations that could act as potential multipliers, trying to maximise the usage of their marketing channels and resources. These activities will not be restricted to some of the organizations mentioned in this document, such as BDVA, ADRA or GAIA-X, which manage social media channels with relevant outreach, newsletters, and other resources. However, they will be extended to other types of actors such as influencers in our scientific area of work, policymakers and representatives of national channels (Ideal-IST network, ministries in charge of the Digital Agenda or similar strategies and notably, communication services of the European Commission) and specialised press, when affordable by the project. A specific communication-press kit will be created to ease this process with online resources, allowing other organizations (not initially identified by the project) to follow specific guidelines to promote Graph-Massivizer.

4.2.7 SUPPORTING MATERIALS

The project aims to minimise its environmental impact and maximise its outreach possibilities using digital materials. Printed promotional materials will only be distributed at events, enhancing different website sections and disseminating the project outcomes through various channels.

The project plans to produce promotional videos for advertising purposes. These videos will provide engaging content and attract the attention of social media users. One video will focus on presenting the project's benefits and value proposition, while short videos will showcase the results of the specific use cases. The possibility of generating other types of videos or video clips will be explored depending on specific needs. All videos will be published on the project's YouTube channel and website with the informed consent of all individuals featured.

Visual materials, such as infographics, booklets, flyers, and leaflets, will be developed during this phase to increase awareness and highlight project achievements by looking at the Graph-Massivizer's presence at physical events. The final format will be selected based on the purpose, information, and channel. Specifically, the team will create a leaflet, brochure/flyer with important project topics for uploading on the website and for partner engagement actions.

Humans are emotional, and attracting people to booths at exhibition areas is increasingly difficult. The project will reward visitors and active contributors with project-branded goodies, which still work quite well for the first attraction. Then, the value of the project will need to be prominently presented for

follow-up actions. In addition, using or wearing specific project articles by all the team members at events will make us recognizable to external people and contribute to our brand's promotion. Figure 12 gives examples of goodies currently under analysis based on production timing and cost).

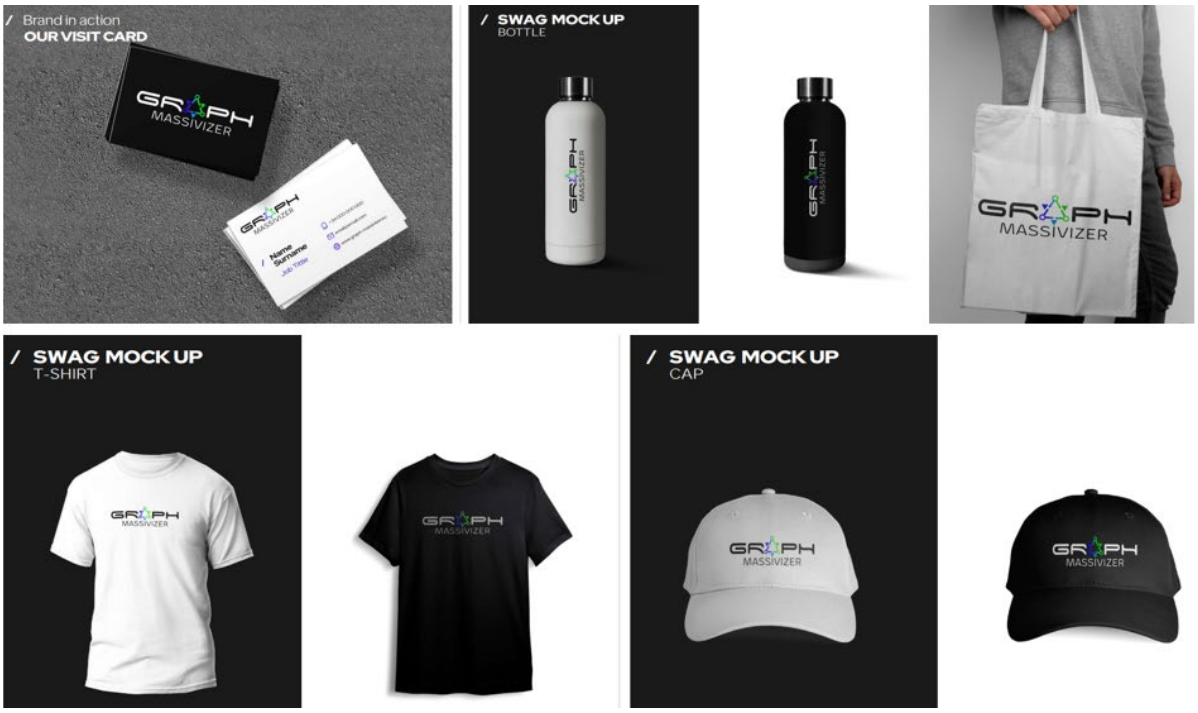


Figure 12: Graph-Massivizer: communication materials and goodies with branding.

4.3 COMMUNICATION AND DISSEMINATION KPIS

If indicators are not defined in advance and continuously monitored, assessing the effectiveness of a dissemination and communication strategy is impossible. Table 4 and Table 5 present the most crucial KPIs associated with communication and dissemination, regularly monitored during the project. The strategic plans, dissemination, and communication strategy review will be based on these KPIs yearly. The list of KPIs is subject to modification based on the insights gained from the project, in line with the overall dissemination and communication plan. The following tables present the yearly and total KPIs for dissemination and communication.

The consortium has established a tracking system to monitor these indicators relying on two factors: the development of templates for project partners to report on their activities and the responsibility of each partner to monitor their progress in terms of events, articles, etc. Our monitoring approach aims to gather timely and regular information on the various activities. All project partners must report on their progress on at least a monthly basis and more frequently for specific events or occasions. The information requested from project partners includes but is not limited to the type of activity (e.g. event, article, webinar, press partnership), the target audience, the date of the activity (if applicable), the number of stakeholders reached, and a detailed description of the activity. This information will be used to develop best practices and lessons learned from successful activities. Collecting and monitoring this information regularly allows the team to review the current strategy yearly with sufficient data and insights.

Table 4: Graph-Massivizer dissemination KPIs.

Activity	Metric	Year 1	Year 2	Year 3	KPI
Scientific and technical publication	Peer-reviewed conference papers	10	10	20	40
	Peer-reviewed journal articles	1	2	2	5
	Peer-reviewed journal special issues	0	0	1	1
	Popular website articles	1	1	1	3
Presentation	National and international presentations	30	30	40	100
Event participation	Workshops organisation (≈ 30 participants)	1	1	1	4
	Toolkit software demonstrations	1	1	1	3
	Use case demonstrations	0	4	4	8
	Poster presentations	12	12	12	36
	EC concertation meeting presentations	3	3	3	9
	EU project collaborations	3	3	3	9
Training	Graph-Massivizer toolkit tutorials	0	2	2	4
	Use case tutorials	0	2	2	4
	University lectures	2	4	6	12
Collaboration	EU project collaborations	2	3	5	10
	Use case association memberships				2
Open science	Graph-Massivizer workload archive			10 TB	10 TB
	Open-source software tools	0	0	5	5
	Reproducible science artefacts	0	4	4	8
Standardisation	Contributions to standardisation bodies				2

Table 5: Graph-Massivizer communication KPIs.

Activity	Metric	Year 1	Year 2	Year 3	KPI
General media	Press releases	2	3	5	**10**
	Newspaper or magazine articles	1	2	3	**6**
	Newsletters	3	4	4	**11**
Website	Unique visitors	2,000	3,000	5,000	**10,000**
	Number of blogs	12	12	12	**36**
Social media	Number of LinkedIn and Twitter followers	100	250	500	**500**
	Number of tweets and posts	300	300	400	**1,000**
	Number of YouTube videos	4	6	8	**18**
Public events	Promotional keynote presentations	1	3	6	**10**
Open science	*"Science citizens"*	0	0	50	**50**
Policymaking	EC concertation meetings	3	3	3	**9**
	Government, NGO			1	**1**

Even though a formal account for such KPIs will be developed in M12, we can already anticipate some figures that show the proactivity and performance of the project:

- **16 peer-reviewed papers**, going beyond the target set for the first year;
- **2 workshops with over 30 attendees** also beyond the first year KPI;
- **Over eight national and international presentations**;
- **One poster** presentation;
- **Several preliminary use case** demonstrations and tutorials.

5 CONCLUSIONS

D7.1 provides the basis of the work that Graph-Massivizer will develop to maximise its impact through communication, dissemination and exploitation measures. The document starts by compiling all the relevant KPIs that characterise the impact expected by the project, including some KPIs of technical nature associated with the performance of the toolkit, its tools, supporting technologies and use cases, as well as KPIs of a business nature associated mainly to the four use cases implemented in the project. They will drive some of the directions of our activities to ensure the fulfilment of such indicators. Still, they will also be a reference for elaborating convincing messages towards the target audience.

The plan for exploitation activities combines a series of activities aiming at producing a joint exploitation plan for the toolkit and individual exploitation plans based on partners' interests and capabilities. The plan is anticipated by describing the expected key exploitable results as a comprehensive view of the Graph-Massivizer Toolkit, its tools and added value for the market needs. A list of potential users for the proposed solution helps to understand the benefits of Graph-Massivizer in practical terms. The document also introduces organizational structures that could enable the exploitation of the project outcomes as well as potential business models, all of them as possible choices that the consortium will have to consider as one or several sustainability paths so that the project can run preparatory activities for a successful implementation of the desired strategy.

With a clearer idea of the KPIs' impact, project outcomes and possible exploitation roots, the document brings the strategy for dissemination and communication to the project interests. The strategy includes objectives for communication and dissemination, phases, and the definition of target audiences and channels. Understanding the project's positioning is crucial in defining messages and actions. This document analyses the centre of gravity of Graph-Massivizer, potential added value and relationship with other projects and initiatives in the R&I landscape, and its competitive position businesswise, looking at its value proposition for the current market offering in graph processing. The collaboration, clustering and networking strategy is based on the positioning in the R&I landscape, while the second analysis complements and supports the exploitation and go-to-market activities.

A more operational D&C plan brings the strategy into specific activities and channels in action. In particular, we expose the plan and activities already implemented by the project concerning project branding, templates, dissemination materials, events (both third-party events and those organised by us) and publications, among others. This chapter is complemented by specific KPIs defined by the project for the D&C outreach and a summary of the status of Graph-Massivizer at M3. Details on the reporting and evaluation methodologies are also available.

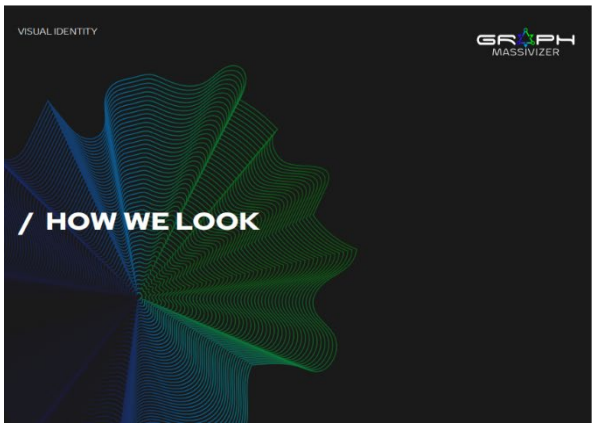
Besides analysing the context and the master lines of the strategy behind the Graph-Massivizer communication, dissemination and exploitation activities, this document illustrates that the project has been very active in setting up the procedures and basic elements to create awareness, with the following early results and monthly achievements:

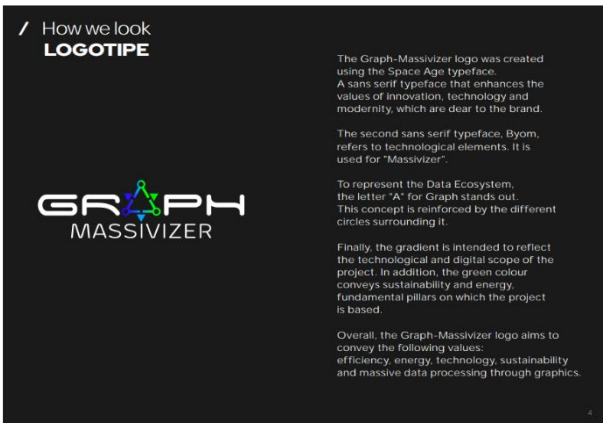
- **M1:** logo and branding guidelines;
- **M2:** templates, initial project website and social media channels (LinkedIn, Twitter, YouTube);
- **M3:** 16 peer-reviewed papers, organised two workshops with over 30 attendees and gave over eight national and international presentations, including one poster presentation.

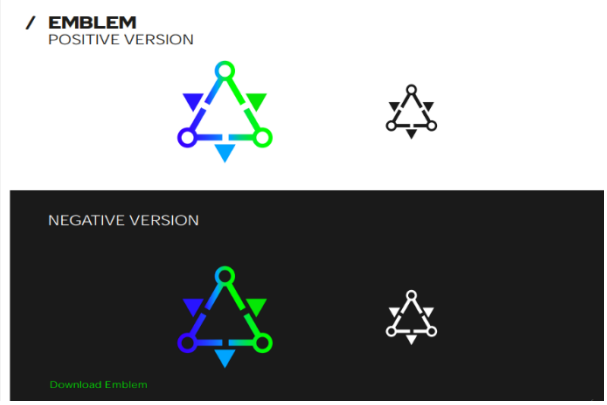
ACRONYMS

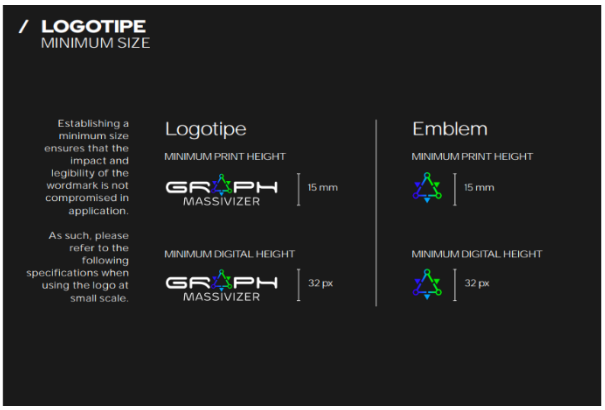
Acronym	Description
ADRA	AI, Data and Robotics Association
AI	Artificial Intelligence
BDVA	Big Data Value Association
D&C	Dissemination and Communication
DIH	Digital Innovation Hub
DSSC	Data Space Support Center
ETP4HPC	European Technology Platform for High-Performance Computing
EU	European Union
EUH4D	EU Hubs for Data
HPC	High-Performance Computing
IDSA	International Data Spaces Association
IEEE	Institute of Electrical and Electronics Engineers
IoT	Internet of Things
IPR	Intellectual Property Rights
IT	Information Technologies
KPI	Key Performance Indicator
M*[number]*	Month*[number]*
ROI	Return on Investment
SME	Small and Medium Enterprise
UC	Use Case
WP	Workpackage

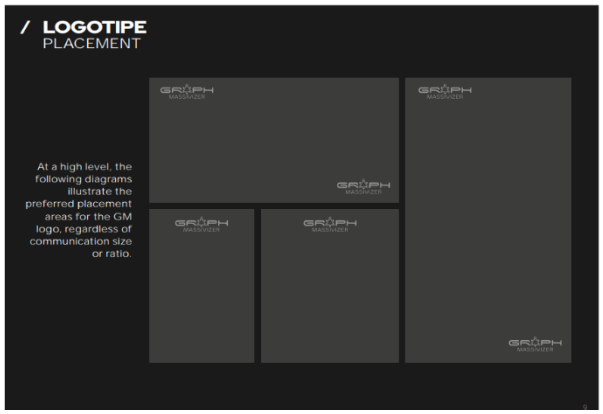
ANNEX 1: BRANDING GUIDELINES

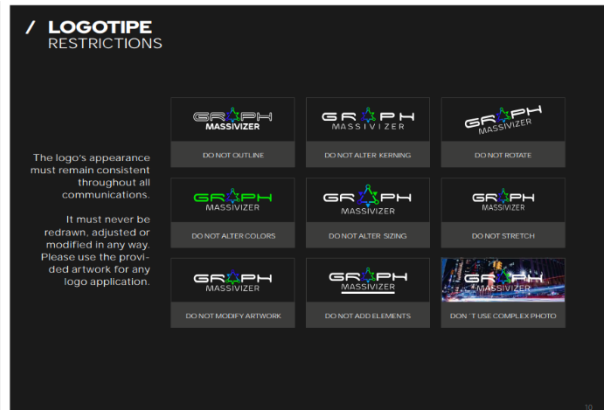












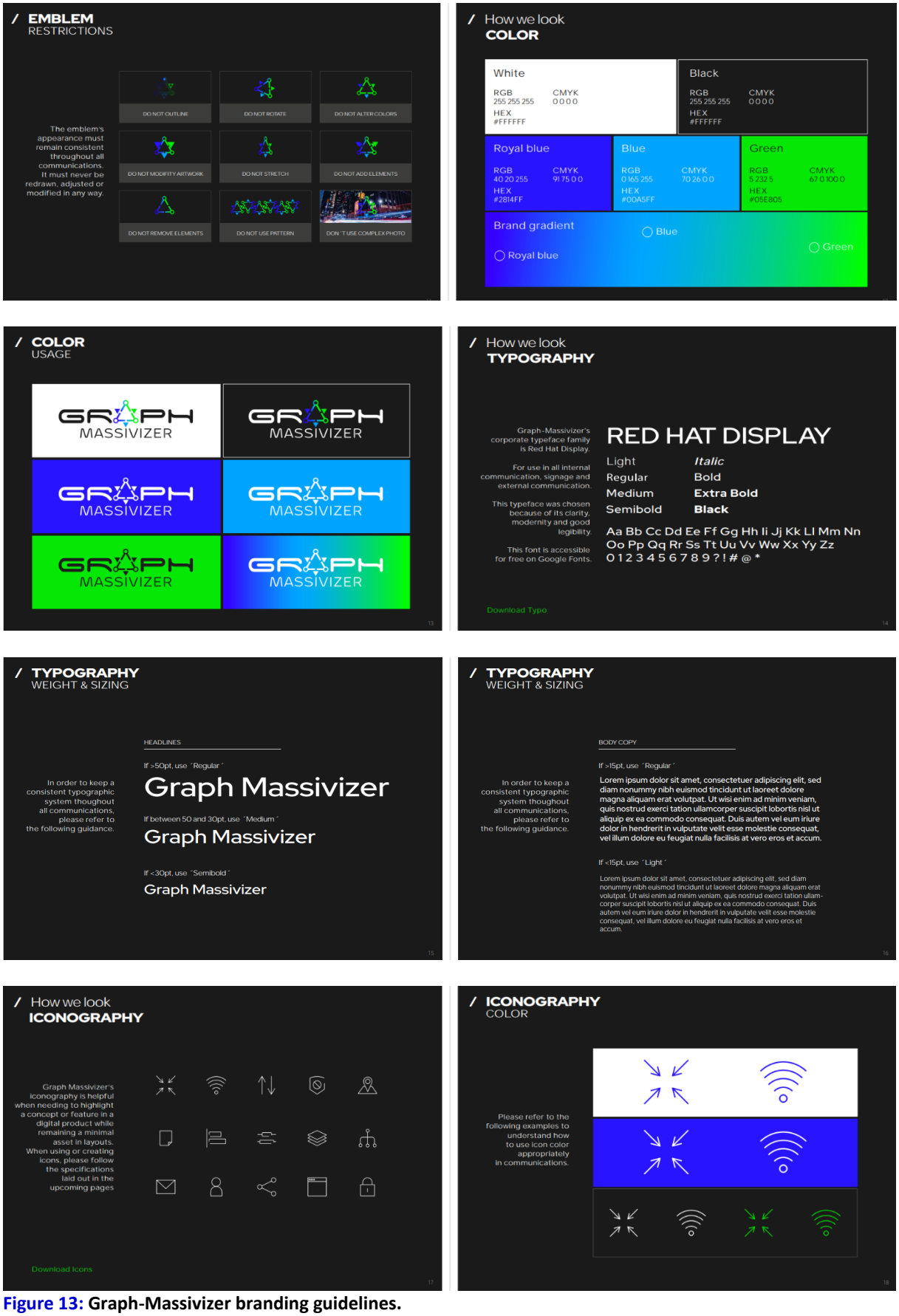


Figure 13: Graph-Massivizer branding guidelines.

ANNEX 2: PROPOSAL FOR A SESSION AT DATA WEEK 2023

DATAWEEK 23

JOIN.LEARN.SHARE.GET VALUE

Are current infrastructures suitable for extreme data processing?

Insights, examples and discussion about infrastructure requirements and limitations of the Computing Continuum for extreme analytics, including sustainability and energy-efficiency aspects

Topics to be linked (from the list provided by organisers):

Main:

- Convergence HPC + DM + ML (extreme data analytics)
- HPC and connection to industry transformation Green and Digital

Others (to less extent): Edge-Cloud continuum, Energy + Green deal

Objectives

- Present current needs of extreme data processing with respect to infrastructures, elaborating on major challenges and opportunities for Europe;
- Include energy efficiency and sustainability aspects as part of the problem equation;
- Introduce ongoing work in the domain and propose solutions to overcome main limitations;
- Showcase examples that range from the use of HPC to computing clusters;
- Identify stakeholders willing to engage in validation campaigns in the context of ongoing projects so that industrial needs are well aligned with the technical work proposed by the scientific community;
- Converge towards an agenda of common themes for several projects and initiatives and can benefit from joint work to foster collaboration.

Abstract

The Graph-Massivizer project is a first-in-class project that has understood the potential of Graph processing to solve multiple economic, societal or environmental challenges. However, current technologies are not suited to large-scale graph analytics. Graph-Massivizer will develop tools to overcome these limitations, where **underlying computing infrastructure plays a major role**. The project uses **European infrastructure in the computing continuum, from pre-exascale HPC facilities to local computing clusters** with state-of-the-art networking and in-place cybersecurity capabilities. The session will revolve around analysing the suitability of current infrastructures for extreme analytics, including large-scale graph processing, providing insights on the ongoing work in the field and including examples. Concrete tools and technologies that will be shared and discussed with the audience include graph workload modelling, energy awareness and sustainability and scalable serverless graph processing.

The session will feature a diversity of examples so that technologies can be put in an operational and concrete problem-solving context, and in particular, we will showcase a pilot on Green Data Center Digital Twins, targeting sustainable science throughput through scalable energy-aware, exascale operation and total traceable cost of ownership understanding, including sustainability indicators and their environmental effects (e.g., greenhouse gas emissions). The audience will see how the proposed

solutions will enable the creation of a novel, graph-based digital twin of a data centre; this digital twin will further support the construction of sustainable exascale computing operational models to support scientific discovery in the next decade.

Projects like DataCloud (addressing the complete life cycle management of Big Data pipelines through discovery, design, simulation, provisioning, deployment and adaptation across the computing continuum) and enrichMyData (highly scalable and replicable data enrichment pipelines) are confronted with similar challenges. They will join the discussion following the analysis-technological solutions-example approach.

The session will foster the creation of a collaborative agenda for joint work between the projects promoting the session (**Graph-Massivizer, DataCloud, enrichMyData**) and invite other projects and initiatives on stage to share their experiences. We will also identify **major industrial and scientific players based in Luleå, well-known for establishing Data Centres, to participate in the session and** encourage the different stakeholders to engage in future validation campaigns.

Preliminary Agenda:
Estimated duration: 90 minutes

Setting the scene (5 min): a short introduction to the content and objectives of the session, ensuring full understanding for all the participants.

Tentative speaker: Nuria de Lama (IDC)

Panel I: How suitable are current Infrastructures to support extreme data processing (40 minutes)
The panel will include short presentations of major projects and initiatives working in a range of infrastructures of the Computing Continuum, including the promoters of the session, which are flagship projects in the data domain (Graph-Massivizer, DataCloud, enrichMyData), as well as those with a more focused view on the evolution of such infrastructures, such as the CSAs OpenContinuum and Unlock.cei or major European endeavours like EuroHPC/ETP4HPC representatives. Presentations will be followed by a panel discussion for a more interactive approach.

Tentative speakers: Radu Prodan (Klagenfurt University), Till Christopher Lech/ Titi Roman (SINTEF), Irena Pavlova (GATE), a representative from EuroHPC/ETP4HPC, potential representative of the EUCLoudEdgeIoT.eu initiative.

Panel II: A practical view on the topic: get to know frontrunner examples (30 minutes)

This part of the session will showcase examples/pilots already working on those challenges. Among others: Green Data Centre Digital Twin (CINECA), 1–2 examples expected to be added from companies settled in Luleå focusing on adding sustainability and energy efficiency as key features of their computing infrastructures.

Tentative speakers: Green Data Centre Digital Twin pilot (CINECA representative), a potential representative from another pilot of the related projects, and 1-2 representatives of Data Centers in Lule.

Closing session and next steps (15 minutes)

The last part will wrap up major aspects debated during the session with closing statements from all the participants, interaction with the audience and an invitation to define joint paths towards technical challenges addressed by a set of projects and initiatives, strengthening collaboration among them. Of special relevance will be attracting stakeholders to the validation activities to be developed in the upcoming months so that a better alignment between supply and demand, industrial needs and scientific work can be attained.

Tentative speakers: speakers from Panel I and II (or a subset of them), audience moderated by Nuria de Lama.