Boosting the Impact of Extreme and Sustainable Graph Processing for Urgent Societal Challenges in Europe

Graph-Massivizer: A Horizon Europe Project

Nuria de Lama Sanchez
International Data Corporation
Madrid, Spain
ndelama@idc.com

Peter Haase
metaphacts GmbH
Walldorf, Germany
ph@metaphacts.com

Dumitru Roman
SINTEF AS
Oslo, Norway
dumitru.roman@sintef.no

Radu Prodan
University of Klagenfurt
Klagenfurt, Austria
radu.prodan@aau.at

ABSTRACT

We explore the potential of the Graph-Massivizer project funded by the Horizon Europe research and innovation program of the European Union to boost the impact of extreme and sustainable graph processing for mitigating existing urgent societal challenges. Current graph processing platforms do not support diverse workloads, models, languages, and algebraic frameworks. Existing specialized platforms are difficult to use by non-experts and suffer from limited portability and interoperability, leading to redundant efforts and inefficient resource and energy consumption due to vendor and even platform lock-in. While synthetic data emerged as an invaluable resource overshadowing actual data for developing robust artificial intelligence analytics, graph generation remains a challenge due to extreme dimensionality and complexity. On the European scale, this practice is unsustainable and, thus, threatens the possibility of creating a climate-neutral and sustainable economy based on graph data. Making graph processing sustainable is essential but needs credible evidence. The grand vision of the Graph-Massivizer project is a technological solution, coupled with field experiments and experience-sharing, for a high-performance and sustainable graph processing of extreme data with a proper response for any need and organizational size by 2030.

CCS CONCEPTS

• Social and professional topics • Professional topics • Computing industry • Sustainability

KEYWORDS

Extreme data, massive graphs, sustainability, serverless computing, computing continuum, green finance, global foresight, green automotive, sustainable exascale computing

ACM Reference format:

Nuria de Lama Sanchez, Peter Haase, Dumitru Roman and Radu Prodan. 2023. Boosting the Impact of Extreme and Sustainable Graph Processing for Urgent Societal Challenges in Europe: Graph-Massivizer: A Horizon Europe Project. In *Companion of the 2023 ACM/SPEC International Conference on Performance Engineering (ICPE '23 Companion). ACM, New York, NY, USA, 6 pages.* https://doi.org/XX.XXXX/XXXXXXXXX

1 Introduction

The use, interoperability, and analytical exploitation of graph data are essential for the European data strategy. Graphs are crucial to innovation, competition, and prosperity and establish a strategic investment in technical processing and ecosystem enablers. Graphs are universal abstractions that capture, combine, model, analyze, and process knowledge about real and digital worlds into actionable insights through item representation and interconnectedness. For societally relevant problems, graphs are extreme data that require technological innovations to meet the needs of the European data economy. For example, a study by IBM revealed that the world generates nearly 2.5 quintillion bytes of financial data daily, posing extreme analytics challenges. The complexity, diversity, and data multilingualism lead to highly complex graph operations when encoding media, news, and social-networking messages. Digital graphs help pursue the United Nations Sustainable Development Goals (UN SDG) by enabling better value chains, products, and services for more profitable or green investments in the financial sector and deriving trustworthy insight for creating sustainable communities. All science, engineering, industry, economy, and society-at-large domains can leverage graph data for unique analysis and insight, but only if graph processing becomes easy to use, fast, scalable, and sustainable.

The Graph-Massivizer project [1] delivers an integrated toolkit with ambitious technological breakthroughs in line with the Horizon Europe Strategic Plan and data strategy of the European Union, providing a unique, fundamental building block in its target green and digital transformation. The project leverages the

world-leading roles of European researchers in graph processing and serverless computing. It uses leadership-class European infrastructure in the computing continuum, from pre-exascale high-performance computing (HPC) facilities to local computing clusters with state-of-the-art networking and in-place cybersecurity capabilities. Graph-Massivizer validates its innovation on four complementary use cases considering their extreme data properties and coverage of the three sustainability pillars (economy, society, and environment): sustainable green finance, global environment protection foresight, green artificial intelligence (AI) for the sustainable automotive industry, and data center digital twin for exascale computing. Graph-Massivizer targets 70% more efficient analytics than AliGraph [2], and 30% improved energy awareness for extract-transform-load (ETL) storage operations than Amazon Redshift. Furthermore, it aims to demonstrate a possible two-fold improvement in data center energy efficiency and over 25% lower greenhouse gas (GHG) emissions for basic graph operations.

2 Expected Products

Graph-Massivizer gathers a multidisciplinary team that researches, develops, integrates, deploys, and validates a novel integrated toolkit integrating five stand-alone tools for sustainable development and operation of extreme data processing, represented as *massive graphs (MG)*. The project validates its outcomes on four industrial and technological use cases. Table 1 summarizes the expected software products of the project.

2.1 Graph-Massivizer Toolkit

The project delivers the *Graph-Massivizer toolkit* of five open-source tools and findable, accessible, interoperable, and reusable (FAIR) graph datasets covering the sustainable lifecycle of processing extreme data as MG, as depicted in Figure 1. The toolkit creates an automated, sustainable graph operation, processing, and resource management loop with unprecedented capabilities and quality.

2.1.1 Data ingestion and graph creation. Graph-Inceptor tool translates extreme data from various static and event streams or follows heuristics to generate synthetic data, persist it, or publish it within a graph structure.

2.1.2 Graph analytics and reasoning. Graph-Scrutinizer tool implements three basic graph operations, graph enrichment, query, and analytics, which analyze and expand extreme datasets using probabilistic reasoning and AI algorithms for graph pattern discovery, low memory footprint graph generation, and low latency error-bounded query response. The output of this phase is a new graph, a query, or an enriched structured dataset.

2.1.3 Graph workload modeling. Graph-Optimizer tool analyses and expresses the given graph processing workload into a workflow of basic graph operations (BGO). It further combines parametric BGO performance and energy models with hardware models to generate accurate performance and energy consumption predictions for the workload running on a given multi-node, heterogeneous infrastructure of CPUs, GPUs, and FPGAs. The predictions indicate the most promising combinations of BGO optimizations and infrastructure, i.e., a codesigned solution for the given workload while guaranteeing its performance and energy consumption bounds.

Table 1: Graph-Massivizer software products.

Product	*Description*
Graph-Massivizer toolkit	Automated, scalable, and sustainable processing of extreme data based on MG representation
Graph-Massivizer workload archive	Open trace data resulting from validating the toolkit with external beta-testers
Graph-Inceptor tool	Extreme data ingestion, MG creation, and storage
Graph-Scrutinizer tool	MG analytics and reasoning
Graph-Optimizer tool	Graph processing workload modeling with performance and energy guarantees
Graph-Serverlizer tool	Scalable serverless graph analytics over a codesigned continuum infrastructure
Graph-Greenifier tool	Sustainable graph processing with low energy consumption and GHG emissions, evidence-based
Environment Protection Foresighter	Subscription foresight service focused on the four environmental SDGs of the UN
Green Financial Data Multiverse	Synthetic extreme financial data sets for improved green investment simulations
Green Manufacturing Line Diagnoser	Green AI for sustainable optimization of welding robots in automotive manufacturing
Green Data Centre Digital Twin	Graph data center operational model for sustainable exascale computing

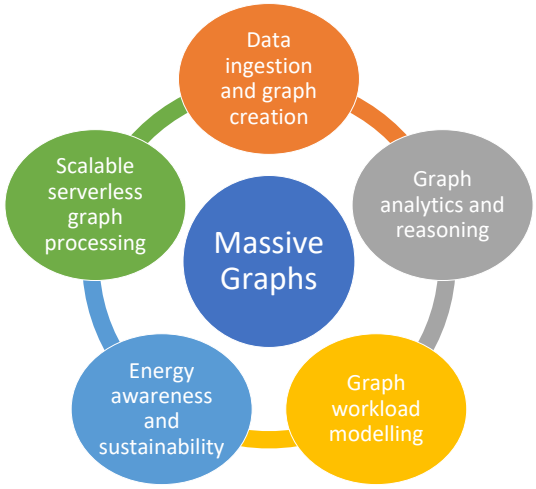

Figure 1: Sustainable operational lifecycle of extreme data represented as MG.

2.1.4 Energy awareness and sustainability. Graph-Greenifier tool collects, studies, and archives performance and sustainability data from large-scale operational data centers and national energy suppliers. This phase simulates multi-objective infrastructure sustainability profiles for operating graph analytics workloads, trading off performance and energy (e.g., consumption, CO_2, methane, GHG emissions) metrics. Its ultimate purpose is to model the impact of specific graph analytics workloads on the environment for evidence-based decision-making.

2.1.5 Scalable serverless graph processing. Graph-Serverlizer tool uses performance and sustainability models and data from the previous phases to deploy serverless graph analytics on the computing continuum. It relies on novel scheduling heuristics, infrastructure partitioning, and environment-aware processing for scalable orchestration of serverless graph analytics with an accountable performance and energy consumption tradeoff.

2.1.6 Graph-Massivizer workload archive. Graph-Massivizer aims to release an open FAIR dataset generated from the toolkit beta-testing and validation campaigns. The dataset of approximately 10 TB in size seeks to support future simulation research targeting sustainable graph processing problems.

2.2 Use Cases

Graph-Massivizer selects four real-world use cases with complementary economic, social, and environmental sustainability profiles, depicted in Figure 2. They must tackle complementary extreme data processing and MG analytics challenges, going in order of magnitude beyond big data in at least three "V" characteristics. Graph-Massivizer proposes a novel characteristic called *"Viridescence"*, representing the sustainability of processing extreme data from an environmental perspective. Table 1 summarizes seven characteristics enhanced through Graph-Massivizer from "big-to-extreme" dimensions and their balanced distribution across the four use cases.

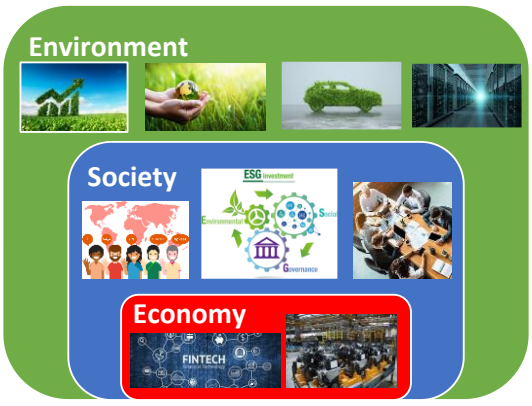

Figure 2: Sustainable Graph-Massivizer use cases, constraining economy and society by environmental limits.

Table 1: Big-to-extreme data enhancements in Graph-Massivizer use cases.

Use Case	*Volume*	*Velocity*	*Value*	*Veracity*	*Variety*	*Viscosity*	*Viridescence*
1	✓		✓	✓			✓
2	✓				✓	✓	✓
3		✓		✓	✓		✓
4	✓	✓			✓		✓

2.2.1 Synthetic Financial Data Multiverse. Graph-Massivizer aims to remove the limitations of financial market data providers (limited volume, reduced accessibility, very high costs) by enabling fast, semi-automated creation of realistic and affordable synthetic extreme financial data sets, unlimited in size and accessibility. Graph-Massivizer employs the financial multiverse for improved AI-based green investment and trading simulations, free of critical biases such as prior knowledge, overfitting, and indirect contaminations due to present data scarcity.

2.2.2 Environment Protection Foresighter. Graph-Massivizer removes the limitations of classic strategic geopolitical and business foresight (e.g., expert surveys) using MG encoding crawling Web data and multilingual live media news. The AI-based analysis enables frequent horizon scanning, megatrend analysis, and higher scalability for processing and combining different data sources and events while avoiding human-prone biases. This use case provides an innovative subscription-based foresight service targeting the four environmental SDGs of the UN: climate action, responsible production and consumption, clean water and sanitation, and clean and affordable energy.

2.2.3 Green Manufacturing Line Diagnoser. Graph-Massivizer enables new graph-based encoding that captures several value-chain stages to predict their outcome better and detect anomalies. Better and quicker analysis prevents defect propagation and unnecessary waste, contributing to a sustainable, circular, and climate-neutral automotive industry. By combining graph-based AI methods with digital twins, Graph-Massivizer provides new insights and boosts the efficiency and scalability of the diagnosis beyond that of more expensive alternatives (e.g., excessive sensor deployment for continuous monitoring). The insights gained will help optimize manufacturing operations and improve the operational quality of the resulting products.

2.2.4 Green Data Center Digital Twin. Graph-Massivizer targets "sustainable science throughput" through scalable energy-aware, exascale operation, and traceable "total cost of ownership" (TCO) understanding, including sustainability indicators and their environmental effects (e.g., GHG emissions). The Graph-Massivizer tools will enable the creation of a novel, graph-based digital twin of a data center; this digital twin will further support the construction of sustainable exascale computing operational models to support scientific discovery in the next decade.

2.3 Commercialization Plan

In addition to developing open-source tools, Graph-Massivizer identifies exploitable commercial-use products and devises business models tailored to their stakeholder needs. The use cases are a particular target for developing innovative products in preparation for commercialization after the project completion.

Creating a ready-to-use integrated platform is central to achieving maximum adoption in enterprise environments. To this end, and to complement the usage scenarios of the individual tools, Graph-Massivizer develops an enterprise-class commercial version based on the metaphactory knowledge graph platform that tightly integrates the tools in an easy-to-use-and-deploy integrated offering. metaphactory [2] is a comprehensive graph management platform that further extends Graph-Massivizer with its existing capabilities for knowledge graph modeling, end-user interfaces for interacting with graphs, and a low-code environment for graph application development. It also provides maximum interoperability in graph data management using open standards around the graph data models of the semantic Web standards stack (i.e., RDF, SPARQL, OWL, SHACL). In particular, the support for RDF-Star will provide the basis to bridge between the two major graph data models of RDF-based semantic graphs and labeled property graphs.

No partner has complete knowledge and technical capacity to commercialize the Graph-Massivizer toolkit on a specific market. Therefore, various project partners will form independent joint ventures as virtual collaboration enterprises targeting commercial market opportunities. The interplay of all strengths has a leverage effect on the common offering and market outreach capability. Finally, each partner can access local EU markets that are much harder to reach abroad, which is essential for SMEs.

In summary, the commercial version aims to provide a ready-to-use, integrated offering with enterprise-level support, training and consulting services to lower the entry barrier for Graph-Massivizer to reach a broader market share, including the use cases of the project and beyond.

3 Market Analysis

Graph databases support any type and depth of relationship analysis and data mining, well-positioned for various applications such as fraud detection, social media influence analysis, system operation, and scientific breakthrough. The graph's core abstraction uniquely enables predictive analysis, network analysis, and problem determination. However, most companies are neither aware nor skilled enough to benefit from graph analysis. Many research institutions with expertise in the field still publish and disseminate research too abstract and not easily accessible to industry and SMEs. According to the International Data Corporation, the market saturation is under 5%, with banking, financial, and energy services taking the lead, but with more aggressive investment predicted in a 3–5-year timeframe. Existing adapters are typically large enterprises building and hosting their tools due to the lack of graph analytics tools. Performance inefficiency can lead to unsustainable operations at the European scale, so a lack of graph-processing solutions and evidence of their sustainability is vital to ensure commercial success.

Graph-Massivizer selects use cases from four large markets to boost its impact beyond the large-scale graph analytics tools:

1. The *financial analytics market* will reach $25.38 billion by 2028, growing at a compound annual growth rate (CAGR) of 11.53% from 2021 to 2028. The US market size for financial data was roughly $15.8 billion in 2022, growing at a CAGR of $0.5 billion.
2. *The environmental, social, governance (ESG), and sustainability consulting market* amounted to approximately $1 billion in 2022 and will reach $16 billion by 2027, growing at a CAGR of 17%.
3. The *automotive market* will reach $90.5 billion worldwide by 2027, growing since 2019 at a CAGR of 2.0%. The welding equipment market will reach $18.79 billion by 2026, with a CAGR of 6.3%. The worldwide development of welding hardware speeds up at a CAGR of almost 7%, between 2021 and 2025, with the Asia Pacific contributing by 51%.
4. The *HPC market* is central to European industrial development, with an average of €867 increased revenues and €69 profit for each invested Euro. The European Commission has recognized the importance of HPC in the European data economy and expanded more than €8 billion in the forthcoming Multiannual Financial Framework (2021–2027). The sustainable use of this investment becomes, therefore, a priority.

4 Target Impact

We present the target impact of the Graph-Massivizer project, covering its contribution to the Horizon Europe strategic plan for European leadership in the data economy and urgent societal challenges for improved social and economic benefits.

4.1 Horizon Europe Strategic Plan Contribution

4.1.1 Next-generation computing. Graph-Massivizer supports seamless serverless graph processing operations on the computing continuum, aggregating the complete range of codesigned HPC, Fog, and Edge resources. The toolkit promises 70% more efficient analytics than AliGraph and 30% improved energy awareness for ETL storage operations than Amazon Redshift. Furthermore, it aims to demonstrate the possibility of a two-fold improvement in data center energy efficiency and over 25% lower GHG emissions for BGO processing, considering Tier 1 and national suppliers' and public energy data. The Green Data Centre Digital Twin modeling relationships between cooling equipment, computing nodes, and HPC workloads promise to pave the road towards the next generation of European sustainable exascale computing with 10% improved power efficiency and 20% improved utilization.

4.1.2 Data technologies and infrastructures. Graph-Massivizer aims for scientific leadership in sustainable MG processing of extreme data on a codesign computing continuum infrastructure, addressed in Section 4.2.1. These are groundbreaking technological results agreed upon by research and industry.

4.1.3 European single market for data. Graph-Massivizer drives innovation within a single European market in two extreme data-centric financial and industrial domains, addressed in Sections 4.2.2 and 4.2.3.

4.1.4 Trustworthy AI ecosystem. Graph-Massivizer drives innovation within a single European market in two extreme data-centric financial and industrial domains, addressed in Sections 4.2.2 and 4.2.3.

4.2 European Leadership in Data Economy

4.2.1 Scientific leadership in sustainable MG processing extreme data on the computing continuum. Graph-Massivizer open-source software toolkit brings a new dimension to European ICT security products and services scale and complexity of extreme data analytics. Graph-Massivizer aims to lower the extreme -data streaming ingestion latency to 500 milliseconds and targets an analytics throughput of over 3,000,000 triples per second for MG with 10 billion nodes and 100 billion edges. Graph-Massivizer will achieve this scalability through 80% accurate BGO performance and energy consumption models of codesigned commodity and specialized hardware accelerators while reducing energy use two-fold and GHG emissions by 25% for BGO. A serverless operational engine validates the promise of targeting 40% faster deployment and 70% faster graph analytics than business enterprise solutions like AliGraph [2]. Graph-Massivizer promises to give the European Union a winning position in this area with few vendors and no uncertain paths.

4.2.2 Technological leadership in the green and sustainable financial market. Graph-Massivizer innovates an extreme data product called Green Financial Data Multiverse, targeting ten prospective customers by the end of the project and covering at least six relevant securities, including stocks, futures, commodities, bonds, exchangeable trade funds, and currencies. Graph-Massivizer strives for a 30% increased return on investment (ROI) by commercializing the synthetic data two years after the project and continuously releasing data samples to support open science.

4.2.3 Industry 4.0 leadership in automotive welding solutions. Graph-Massivizer aims to strengthen Europe's leading position in the automotive industry as a component, technology, and smart manufacturing provider, particularly in resistant spots, ultrasonic, and other robotized welding. The vision is to ensure that each product is either produced with the help of AI or contains AI. Graph-Massivizer looks to share its AI-based automation of welding solutions with hundreds of stakeholders from the Catena-X automotive network for continuous data exchange, standardization, and sustainable automation along the European automotive value chain. Such massive scaling and impact require AI to scale over MG. The partners strive for at least 40% ROI using Graph-Massivizer technology within the first three years of exploitation of the project results.

4.2.4 Leading European supercomputing and big data hub. Graph-Massivizer will integrate its technology with the Bologna Technopole, aggregating the next-generation supercomputers of Cineca, the National Institute of Nuclear Physics, and the European Centre for Medium-Range Weather Forecasts powered at 60MW and hosting over 4,000 scientific researchers. The Green Data Centre Digital Twin demonstrated on the pre-exascale Leonardo system operating at approximately 6 MW (10%) provides essential information for HPC facilities' future engineering and sustainable operation. Critical to evidence-based TCO and green practices, capturing graph data of complicated topographical and temporal linkages between cooling equipment, compute nodes, and MG workloads. The partners intend to apply their product in the European Pilot for Exascale codesign.

4.3 Social and Economic Benefits

4.3.1 Greener financial algorithms and better investments. Graph-Massivizer aims for better green ESG investments by improved financial algorithmic simulations for in-depth analysis, increased returns, and de-risking. The Green Financial Data Multiverse will offer 20%–40% cheaper yet complete, high-quality synthetic streaming data with five times higher volumes. This high-quality synthetic streaming data of potentially "unlimited" volume allow investors to optimize their portfolios by in-depth simulations and increase their investment performance by 2%–4% and their excess return by 1%–2% with a quick ratio higher than 1.5, reflecting healthy investments with lower risk and higher returns.

4.3.2 Daily strategic societal foresight in extreme global media news. Graph-Massivizer will enable policymakers, governments, non-governmental organizations, and business strategists to extract timely trends and scenarios by applying MG analytics to global media news, where the three societal systems (economy, politics, science) converge. The partners expect to speed up foresight creation between seven and 30 times with a latency below 24 hours (daily).

4.3.3 Green AI-based manufacturing line automation. The AI-based diagnosis of welding robot operations, welding spots, and welding caps in bodies-in-white manufacturing aims to improve the quality, time, and manufacturing line downtime by 5% –15%, deployed at dozens of factories. Graph-Massivizer recognizes the role of the European Green Deal in Industry 4.0 development and looks to improve its AI portfolio with green solutions for significant reductions in waste and energy consumption.

4.3.4 Sustainable green computing for improved societally relevant science throughput. Graph-Massivizer aims to maximize the societal and economic impact through high computation and scientific throughput per investment. Despite operating over 1,700 million core yearly processor hours, Cineca strives to increase the "science throughput" towards socially relevant fields. Last year, it partnered with national and international researchers for urgent societal problems such as COVID-19 drug discovery and energy production from fusion reactors. Graph-Massivizer pursues this goal by modeling the new Leonardo supercomputer as a digital twin and gaining unique insights into its sustainable operation and

TCO through 100% capturing spatial interaction, 10% improved power usage effectiveness, and 20% improved utilization.

4.4 Urgent Societal Challenge Management

4.4.1 Geopolitical and business foresight for environment protection. Environment Protection Foresighter aims to be an innovative foresight subscription service for automatic AI-based horizon scanning and megatrend analysis across extreme global news data sources, targeting the four environmental SDGs of the UN: climate action, responsible production and consumption patterns, clean water and sanitation, and clean and affordable energy. Graph-Massivizer targets at least one public (policymaker, governmental organization) and one private sector stakeholder (business strategist), implementing at least 25% of relevant environment protection insights.

4.4.2 Improved data center sustainability. The Graph-Greenifier tool based on the open-source OpenDC simulator [4] proposes an accurate sustainability-performance analysis of extreme graph processing workloads running on large-scale infrastructure, including a sustainability benchmark and green labeling of BGO. The tool aims to demonstrate the possibility of a two-fold improvement in data center energy efficiency and over 25% lower GHG emissions, considering Tier 1 and national suppliers' and public energy data. The tool avoids 'greenwashing' and strengthens Europe's ability to pursue real sustainability by providing quantitative evidence.

5 Conclusions

Graph-Massivizer allows European green financial investments, automotive, and media industries to accelerate at supercomputing speed and get a competitive advantage on graph-based powerful analytics, with evidence of improved performance and sustainability. While other forms of analysis rely on present assumptions about "what happened" or "what happens", correctly building and employing graphs can further reveal the predictive patterns suggesting what "might happen" with clear evidence for each connection or inference step. Graph processing facilitates solving problems in many use cases driven by metrics related to costs, fraud, equipment failure, and inefficiencies and enables extra revenues from better intelligence. The large-scale graph analytics tools market still traverses a developing phase, hampered by the lack of technology research and adoption use cases. Graph-Massivizer provides for Europe these missing links.

The future work follows a development and validation cycle of the Graph-Massivizer toolkit consisting of three iterations involving two stakeholder groups:

- *Five focus groups* of specialized experts participate in the first requirements elicitation through brainstorming sessions with the consortium members: a technological focus group engages with tool requirements and the other four use cases. After each prototype release, the focus group members test the toolkit and the use cases in joint sessions and provide feedback for analysis and incorporation in subsequent design iterations.
- After each prototype release, *five external beta-testing groups* validate the toolkit functionality and the four use cases. Graph-Massivizer consortium plans to connect with relevant, interested beta-testing communities such as the European Network of Living Labs and engage them in the validation, feedback collection and various crowdsourcing scenarios. The third validation campaign verifies the project technology and its use case KPIs.

ACKNOWLEDGMENTS

Graph-Massivizer receives funding from the Horizon Europe research and innovation program of the European Union. Its grant management number is 101093202.

REFERENCES

[1] Radu Prodan, Dragi Kimovski, Andrea Bartolini, Michael Cochez, Alexandru Iosup, Evgeny Kharmalov, Jože Rožanec, Laurențiu Vasiliu, Ana Lucia Vărbănescu. Towards Extreme and Sustainable Graph Processing for Urgent Societal Challenges in Europe. *2022 IEEE Cloud Summit,* 2021, pp. 23-30.

[2] Rong Zhu, Kun Zhao, Hongxia Yang, Wei Lin, Chang Zhou, Baole Ai, Yong Li, Jingren Zhou. AliGraph: a Comprehensive Graph Neural Network Platform. Proceedings of the VLDB Endowment, 12(12):2094–2105, 2019.

[3] Peter Haase, Daniel Herzig, Artem Kozlov, Andriy Nikolov, Johannes Trame, metaphactory: A platform for knowledge graph management. Semantic Web, 10(6):1109–1125, 2019.

[4] Fabian Mastenbroek, Georgios Andreadis, Soufiane Jounaid, Wenchen Lai, Jacob Burley, Jaro Bosch, Erwin van Eyk, Laurens Versluis, Vincent van Beek, Alexandru Iosup. OpenDC 2.0: Convenient Modeling and Simulation of Emerging Technologies in Cloud Datacenters. *2021 IEEE/ACM 21st International Symposium on Cluster, Cloud and Internet Computing (CCGrid),* pp. 226-236.