Promoting and Incentivising Federated, Trusted, and Fair Sharing and Trading of Interoperable Data ASsets

EUDATA+ Use-Case Synergies

A document indicating the synergies between the use cases of the PISTIS, DATAMITE, EnrichMyData, FAME, Graph Massivizer, UPCAST and Extreme XP Projects to serve as a catalyst for collaboration.

Dr Shaun Topham, PISTIS Project.

Funded by the European Union under Grant Agreement 101093016. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the granting authority can be held responsible for them.

Executive Summary

This is an attempt to determine where exploitable synergies lie, between the forty-five use cases shared between the seven EUDATA+ projects.

A table sets out the first impressions from an overview of the seven projects as to where synergies may lie, with brief and rough descriptions of each of the use-cases. These will be revised to be more accurate in due course.

A matrix has been added to capture the synergies which we have already to be identified as potential areas for collaboration.

There follows a table listing all the partners involved in the EUDATA+ projects.

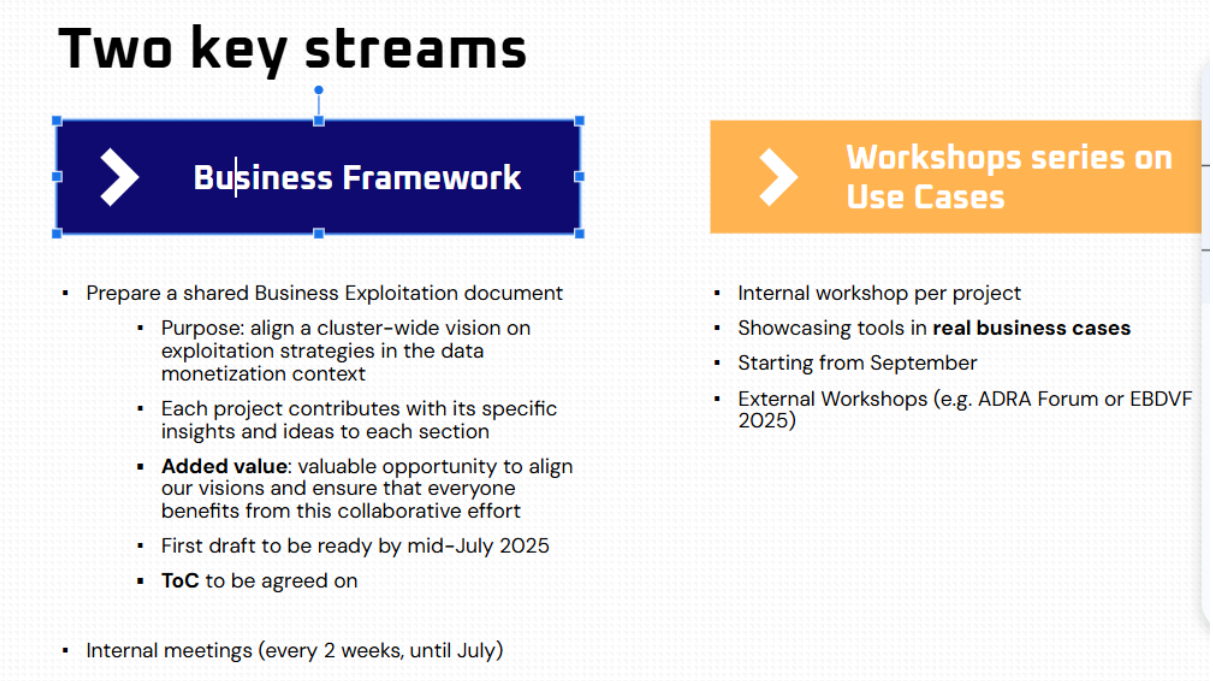
It concludes with short overviews of the seven projects based on their CORDIS overviews.

The paper may also make a contribution to other EUDATA+ documents which are in the pipeline, particularly relating to the Business Group strategy and plans.

INTRODUCTION

There are clear and obvious advantages for the EUDATA+ projects to cooperate in areas such as dissemination, knowledge sharing, market research etc. Additionally, there is potential for the seven projects to initially assist each other in improving the offering from the individual projects, but also, more ambitiously, to investigate if there is any potential "joint-offering" which could be made, complementing the basic business plans of the seven projects.

The EUDATA+ Business Group has the remit to pursue two streams, as indicated in the figure below. This work will contribute to the preparation of a shared Business Exploitation document.



.

DOCUMENT STRUCTURE

Following this introduction, a table sets out some of the initial synergies identified. The following content provides the key to the use-cases which are identified in the table and this at the moment is "rough and ready" as it is intended to be a stimulant to more concrete activities.

The seven main categories are: Mobility, Climate and Environment, Energy, Marketing, Industrial, Finance and City Commerce, all of which could be embraced by "Smart Cities".

A final table lists all the partners involved in the seven projects.

Synergies identified across the 45 Use-cases.	**1. Mobility**	**2. Climate & Environment**	**3. Energy**	**4. Marketing**	**5. Industrial**	**6. Finance**	**7. City Commerce**	**8. Airport**	**9. Corporate group.**	**10. Farming**	**12. Innovation**	**13. Cybersecurity**	**14. Exascale Comp**	**15. Health**	**16. Public Safety**
PISTIS (11)	PI 1.4 PI 3.1 PI 3.2	PI 4 "Weather Factory" supports use-cases.	PI 2.1 PI.2.2 PI 2.3 PI 2.4	Market-ing for EUDATA + as a whole?			PI 1.5	PI 1.1 PI 1.2- PI 1.3							
DATAMITE (6)		DA 6.	DA 3. DA 4.						DA 1 DA 2	DA 5					
EnrichMyData (6)				EN 1	EN 2 EN3 EN 6	EN4					EN 5				
FAME (8)	FA 2	FA6				FA 1 FA 2 FA 3 FA 4 FA 5 FA 7	FA 2.								
Graph-Massivizer (4)		GM 3			GM 2	GM 4							GM 1		
UPCAST (5)		UP 3	UP 6 Tool	UP 1 UP 5										UP 2 UP 4	
Extreme XP (5)	EX 4	EX.1			EX 5			EX 3? EX 5?				EX 2			EX 3

WHERE SYNERGIES HAVE BEEN IDENTIFIED

The use-cases below are listed as references from the table which identifies synergies between some of the use cases across the seven projects.

1.MOBILITY

PISTIS: PI 1.4: PUBLIC TRANSPORTATION PLANNING SUPPORT.

Data exchange between the airport and the municipality can help improve the overall planning of public transport, allowing the development of services that are able to predict the accuracy of the load within the day to improve route scheduling, deployment of vehicles and the efficient utilisation of its fleet, thus offering to commuters better services and lowering operational costs.

The outputs of the analysis will be valuable to parties such as the airport and the city, for the latter to be able to improve their own offer and services towards the commuters, or to relevant third parties (e.g., duty-free shops, local city businesses, etc.). This can be achieved by combining data for incoming passenger traffic to the airport from public transport (e.g., bus occupancy), data from the airport service handlers (e.g., expected queuing, check-in counters availability, security checkpoint staffing), and information such as airport routing, flight, and weather information, etc.

Ambition of use-case

The **aim** of this use case is to design a better service for passengers from and to the airport and the optimisation of the public transport service for the airport.

On top of that, the use case will try to set up a process to enable the live data sharing on road traffic conditions (kerbside), between AIA and OASA. This is to facilitate the OASA driver to make an informed decision on the route to follow so as not to get stuck in traffic and be delayed. The existing data coming from the access control system at the kerbside needs to be further processed so as to provide a meaningful piece of information that is worth sharing.

Target Audience Descriptions

Athens Urban Transport Organisation (OASA), Athens International Airport (AIA) and the City of Athens (DAEM) are the core actors involved in Use Case 1.4. In the public transportation planning support process, OASA acts both as a data seeker and a data owner. As data seeker, OASA will utilise the input data provided by the other actors with the aim of refining its services, whereas as data owner, the organisation will trade back data to the rest of the actors in order for the latter to improve their own services and offerings. The same applies to AIA who acts as a data owner, mainly feeding the PISTIS platform with the relevant data to be utilised by OASA and as a data seeker receiving data from the Public Transport Authority.

DAEM in Use Case 1.4 mainly acts as a data seeker, whereas UBIMET acts as a data owner with its weather data.

Summary of status after Alpha testing phase

Data plays an important role in successful public transport planning, directly impacting passenger satisfaction and operational efficiency. The PISTIS Data Factory facilitates seamless data exchange among stakeholders, enabling advanced analysis and predictive insights. This empowers public transport operators to optimise service planning, enhance user experience, and allocate resources more effectively to meet passenger demand.

As part of the alpha testing phase, OASA evaluated the core functionalities of the PISTIS Data Factory. The process involved ingesting and managing data samples through the following steps:

- Job Configurator: Datasets were uploaded by specifying the file name and description.
- Workflow Execution: The upload progress was monitored in real-time.
- Data Transformation: Datasets were processed using the Transformation Catalogue and Data Transformation tools.

Once published, datasets became available in "MyData", where their properties, such as pricing, availability and download limits could be configured. Additionally, the Marketplace was tested to explore available datasets and simulate purchasing functionalities.

In its full implementation, the PISTIS platform will provide comprehensive datasets for analysis via an interactive dashboard, ultimately refining public transport planning and decision-making processes.

PISTIS: PI 3.1: TRAFFIC QUALITY ASSESSMENT.

Driving patterns of vehicles will be analysed in their temporal, spatial and situational context using vehicle sensor data (connected car data via CARUSO). From this data, specific information is calculated, e.g. acceleration, speed, and fuel consumption patterns along a specific route at certain times and their correlation with weather and air quality. These insights are used as input parameters for urban analytics applications facilitating planning processes, such as emission models or the hot-spot detection of stop-and-go patterns. This generated information is also used for incentivising green driving styles and mobility decisions (such as reducing the share of motorised individual transport in the modal split of a commute) and optimising potentials for greener transport in the context of mobility management. Novel urban analytics components will be prototyped, including input parameters for emission models and traffic quality parameters, as well as mobility management services and mechanisms for companies to promote more sustainable driving styles.

Ambition of Use-case

The **goal** of this use-case is the development of two demonstrators: **urban analytics** and **corporate mobility management**. This will entail the exploration of new datasets, including previously unknown sources, and establish mechanisms for selecting only the right sources that are truly relevant and available on a stable basis.

a) Urban analytics

This will involve urban car traffic quality assessment with combination of geospatial time series data to generate an emission model. The combination of floating car data and detector data to estimate traffic volume, with acceleration patterns and locations of individual vehicles provided by CARUSO, allows the generation of an in-depth model of estimated air pollution levels for various road segments.

Subsequently, road segments with significant pollution levels are identified and clustered based on both temporal and spatial factors. This clustering not only enables a clear visual representation of pollution hotspots but also allows for detailed statistical analysis, offering deeper insights into the dynamics of traffic congestion and its environmental impact.

This results in an improved model of the environmental impact of road traffic on the surrounding environment, facilitating the implementation of targeted traffic planning measures in order to strategically minimise the burdens to the cities and communities.

b) Corporate mobility management

This will entail a dashboard integrating real-time data on weather conditions and emission levels across selectable locations. It actively encourages eco-friendly and healthy commuting by recommending sustainable transport options like bicycles or public transport. The recommendations consider user-specific factors such as the starting point and destination of journeys, but also generalisable factors such as current environmental conditions like temperature (provided by UBIMET), precipitation (provided by UBIMET), risk-zones (provided by VIF) and air quality (using the data generated in the urban analytics use case). Recommendations are based on textual hints, called nudges, that encourage the user to make an environmentally conscious choice of transport.

Target Audience Descriptions

TRAFFICON has access to vehicle data, floating car data and public transport data. These data sources are already in use and available. Additional road networks based on OpenStreetMap data are used as reference point to combine the data and use as a standardised reference. CARUSO can provide access to vehicle data used to characterise emissions based on vehicle type, speed engine RPM and other factors. UBIMET can provide access to current weather data. VIF can provide risk events for further use by Trafficon. Trafficon seeks data from UBIMET, VIF and CARUSO via the PISTIS Platform. In addition, external data described above is imported in other means.

Summary of status after Alpha testing phase

In Use Case 1, VIF developed service-based systems using APIs, including dashboards and a mobile application, that leverage a specific geographic location key (GSK-geospatial key) to query driving risk-related events in the vicinity. These systems analyse events within geographic areas or along the vehicle's driving direction and corridor, correlate them with weather conditions, and issue warnings based on the actual driving speed.

PISTIS: PI 3.2: DRIVING STYLE & RISK ASSESSMENT.

The assessment of driving style and driving risk by using multiple data sources is a major topic for safe driving research and applications. This is a comprehensive safety technology designed to analyse various factors related to driving behaviour, environmental conditions, and potential hazards on the road. The ultimate goal is to provide timely warnings to drivers, helping them proactively navigate potential risks and enhance overall road safety.

A driving risk can be computed by analysing data from connected vehicles (such as location, speed, or driving direction), data about past accident hotspots, data indicating individual driving behaviour (such as harsh braking, acceleration, speed or driver distraction), and data indicating weather conditions and changes, to name a few examples. Thereby, driving-risk relevant events extracted from various data sources form the input for a driving risk assessment pipeline. Drivers may undergo categorisation based on their distinctive driving behaviours, encompassing safety-related events like speeding, braking, use of assistance systems, or reckless driving, all derived from the vehicle's sensor data. To enhance the precision of the driving risk evaluation, the severity of driving events is carefully weighed by incorporating contextual data. Factors such as weather conditions, road topography, locations of harsh driving, or accident hotspot data are integrated to provide a more nuanced understanding.

In the subsequent phase, this rich dataset may become a cornerstone for personalised driver coaching. Leveraging gamification strategies, drivers are incentivised through point systems or rankings, fostering a sense of competition. This encourages the adoption of environmentally friendly and safe driving practices. By employing these engaging techniques, the driving experience transforms into a dynamic platform for continuous improvement, benefitting both individual drivers and the broader community.

In a nutshell this use case revolves around analysing shared data to identify safety-related events. Tailored to individual driving contexts, the system issues driver warnings, promoting safer driving practices. For transportation managers, a comprehensive overview of the entire data landscape is presented through an interactive dashboard, providing a holistic perspective on driver-risk relevant events.

Ambition of use-case

The **goals** are to create a data-driven driving risk prediction and warning system for (1) vehicle drivers that enhances driver awareness, reduces risky behaviours, and improves overall driving and road safety, integrating different data sources from different data providers. The knowledge of risky locations in the road network may also be of interest to (2) transportation managers who may want to establish risk mitigation mechanisms and are therefore provided with an interactive dashboard.

Therefore, the use case aims to

- Establish data trading partnerships or agreements with data providers (e.g., for vehicle and weather data) to ensure a continuous supply of current and historical data.

- Obtain data from relevant sources to develop algorithms and models for a more accurate and useful inference of driving risk and the development of respective driver warning and assistance systems.
- And to have access to large-scale car data alongside current and historic weather data, for a more scalable solution adaptable to a broader application context.

Target Audience Descriptions

VIF can provide access to data containing the location of past acceleration, braking and curve driving events including metadata such as location, time, driving speed and speed change, derived from data containing past road trips. CARUSO can provide access to vehicle data (personalised data) if the driver/owner of the vehicle has agreed to the use of this data for a specific service. CARUSO also has access to anonymous/synthetic vehicle data. UBIMET can provide access to current weather data as well as to historical weather data. VIF is seeking access to vehicle data (provided by CARUSO) as well as to current and historical weather data (provided by UBIMET).

Summary of status after Alpha testing phase

Data sharing is at the core of the automotive hub, which integrates weather, hazard, and vehicle data to provide driver risk warnings, emissions modelling, and visualization. By leveraging the PISTIS Data Factory on a cloud platform, testing was streamlined, enhancing both user experience and convenience by eliminating the need for local deployment by partners.

During the demonstration phase, partners rigorously tested the PISTIS Alpha version, ingesting diverse datasets—including vehicle data and traffic risk events—while transforming, enriching, and anonymising data, all via CSV files. For example, VIF shared risk-related event data, which TRAFFICON acquired and integrated into their dashboard. Beyond technical testing, hub partners established data-sharing agreements, facilitated multiple transactions, and monitored logs for cost and revenue tracking.

While minor UI issues were identified and can be easily resolved, the alpha version demonstrates strong potential for both data sharing and monetisation. We eagerly anticipate evaluating the next iteration.

In Use Case 2, TRAFFICON leveraged CARUSO Car Data to estimate motorised vehicle emissions, enhance or develop emission models, and visualise the results on a dashboard. The dashboard also integrated data from virtual weather stations and local driving hazards.

FAME: FA2 PARKING DATA

The scope of FAME Pilot 2 is to develop novel EmFi services in an urban context, where data from smart services are used to provide additional finance related services. To this end, Pilot 2 aims to create citizen's EmFi profile by leveraging existing parking data, onboarding other behavioural data of citizens and by using citizens profiles to generate and provide personalised citizen-centric offers and recommendations. On the other hand, Pilot 2 aims to extend citizen wallet to a broader range of services such as transportation services, payments of fines, etc.

In the following table, the details of the collected business requirements stemming from Pilot 2 are presented.

Analyse parking data in terms of locality, frequency, and time of parking FAME should optimize the data analysis part in order to be pinpoint parking habits

Request P2_BR2 UC1 data from other organisations that have loyalty programs in order to analyse the rewarding mechanisms FAME should offer an added value to an external entity in order to add their data and offer an authorized and usable onboarding/registration procedure

Acquire knowledge and training on the platform so as to leverage the city's IT personnel competencies Request and onboard data from other organizations so as to leverage the services of the city e.g., traffic data Functional User, technical, infrastructure, FAME should offer training feature (e.g., webinars, user guide, MOOCs, online training sessions etc) Public administration, management

FAME should offer an added value to an external entity in order to add their data and offer an authorized and usable onboarding/registration procedure Parking Management, Marketing Functional User, technical, infrastructure,

Reduce the costs that refer to the management of the parking system of the city FAME should optimize the data analysis in order to identify solutions that can make the system less costly for the city Parking Management, Marketing Functional technical, infrastructure,

Analyse parking and behavioural data in terms of locality in order to design new services e.g. if a citizen owns a shop, then short term FAME should optimize the data analysis part in order to be pinpoint parking and other habits and localities Parking Management, Marketing Functional User, technical, infrastructure, parking for customers could be foreseen

Ensure the sovereignty of the data being traded FAME should have mechanisms to ensure data usage control policies Governance Functional Infrastructure

Correlate data sources to offer new services FAME should offer a view of correlated data and usable visualizations for a city employee Public administration Functional User, technical

Exploit data on parking payments to design specific offers e.g., discounts FAME should offer an analysis of parking payments and usable visualizations e.g., time, duration, location etc Public administration, management Functional User, technical

Request and onboard data from other organizations so as to leverage the services of the city e.g. traffic data FAME should offer an added value to an external entity in order to add their data and offer a authorized and usable onboarding/registration procedure Parking Management, Marketing Functional User, technical

Analyse parking and behavioural data in terms of locality in order to foresee the addition of other municipal services FAME should optimize the data analysis part in order to be pinpoint parking and other habits and localities Parking Management, Marketing Functional User, technical, infrastructure,

Analyse parking and behavioural data in terms of high/low demand in order to develop a dynamic pricing scheme FAME should optimize the data analysis part in order to be able to foresee parking demand based on historical data Parking Management, Marketing Functional User, technical, infrastructure,

Offer citizen wallet to citizens Develop a citizen wallet for citizens to consume services in one app Marketing Functional

Develop a dynamic parking pricing scheme Develop a dynamic pricing mechanism based on parking demand Parking Management, Marketing Functional User, technical, infrastructure.

Ensure the sovereignty of the data being traded FAME should have mechanisms to ensure data usage control policies.

EXTREME XP: EX 4 FLEXIBLE TRANSPORTATION ANALYSIS AND VISUALISATION (MOBY)

Context description and objectives.

In the transportation platforms ecosystem, the main stakeholders/decision-makers are transportation authorities, transportation planners and modellers. They need to assess different transportation policies, using a variety of simulation software, advanced econometric modelling, or other data-driven tools. However, each stakeholder group (transportation planners, modellers, policy makers) has different requirements with respect to the granularity/level of analysis and corresponding outcome precision and quality.

There is no “one-size fits all” when it comes to data analytics and visualisations in the transportation domain. Therefore, ideally, the transportation analysis platform can adapt:

(i) the simulations and data analytics performed in the backend,
(ii) and the visualisation methods and techniques and learn the above for each user segment.

The transportation UC provides a rich space of data analytics and visualization alternatives that can be employed by different users according to their needs. For instance, users may have the need to evaluate a transportation policy before applying it or exploring potential future scenarios and corresponding policies/interventions.

This UC will identify several users' needs to improve decision making in scenarios such as:

(i) Introduction of zero-emission zones for freight transportation in urban areas;
(ii) Introduction of different level of autonomous vehicles in the transportation network;
(iii) Investigation of the relationship between different levels of remote work, impact on transportation and residential choice.

The present UC will explore the possible improvement for providing data-driven estimation, introduction of data pipelines for the calibration of models and adaptive visualisations that fit the purpose and needs of the various users and actors of the transport platforms. This

objective will cope with the different data availability and granularity by combining with smart and adaptive data usages and re-usage by delivering customised policy analysis and decision support. Additionally, each case will rely on receiving data from live sources.

The following ExtremeXP artefacts will be applied:

(i) Interactive and adaptive visualization of extreme data, for GIS visualizations of both reported (GPS traces) and predicted (simulation results) data of passenger or freight traffic flows.

(ii) User-driven Optimization of Complex Analytics, for estimation and application of econometric and other sub-modules on the datasets, generating the necessary results of the simulation platform along with latent classification of users, attitudes, and travel behaviour.

(iii) User Feedback via Serious Games, to perform stated preference experiment data collection using serious gaming techniques to ease the burden of reporting on the respondent.

(iv) Transparent and Interactive Decision Making, for scalable explanation of results based on customer/audience (from experienced modellers to non-technical policy makers), including the use of AR techniques.

Experiment definition

UC 4 aims at improving the decision-making process for transport planners and enabling better insights for forecasts into the future. The approach is twofold, firstly through the provision of an improved annotated dataset containing the habits and travel patterns of a given study area, and secondly through the utilization of a multilayer modelling approach. The outcome of the modelling process serves as the testbed for running various scenarios and evaluating the results through meaningful statistics and visualisations.

When it comes to data collection, this is done through Moby's mobile application, MobyApp which allows end-users to voluntarily share their daily trips and preferences, along with their precise locations throughout a period of time (typically 2-4 weeks). Their trips are computed by MobyApp's ML engine (i.e., Pythia) and then presented to the users for verification. Effectively, once users have verified their trips, what is made available is a thorough mapping of their travel patterns for the given study area and time period.

During data collection (i.e., MobyApp) the main actors are the user (or passenger), who is using the app, and secondarily the organisation/planner/researcher who is collecting the data. The main goals are, thus, mainly focusing on how to make the data collection process easier and more effective for the user, thus, resulting to a more accurate and detailed dataset in the end. To that end, the effort is on the minimization of manual data entry, by producing more accurate prediction of the user's activity, both for stationary and moving segments. Once identified by Pythia, all segments that are presented to the user need to be as close to reality as possible, to enable the user to have limited tampering with their trip diary, so as to be enticed to only fine-tune some missing details, if any.

Additionally, efficient data collection with reduced battery consumption, and the need for adaptive sensing, will allow for less dissatisfaction by the users, who may often see their mobile phone battery depleted due to the continuous stream of data from their phone. The current granularity is set to every 5 seconds, so there is a question of how much data is enough data to produce accurate predictions for their trip diaries. Real-time information of current/recent trips and activity is an essential milestone for allowing users to view their trips while on the go and will in turn enable the addition of a multitude of features. At the moment, location data for each user is batched into chunks and processed at the end of the day. Lastly, predictive accuracy that increases per user on each subsequent use of the app will enable the user to receive more tailored recommendation about their whereabouts and their selected modes of transport. Patterns across the entire userbase can be used for a specific individual, harnessing the power of machine learning, and calibrating per individual. For example, a user may have multiple verified trips by bike, therefore the produced predictions thereon may be deliberately skewed towards predicting more trips by bike than on other users.

Regarding variability points for the data collection by MobyApp, there is currently one algorithmic logic for converting GPS traces into meaningful trip diaries. This means that regardless of the application area of each survey, the governing logic is the same, whichever thresholds or assumptions used may not perform equally across regions. For example, bike may be overestimated in countries like Cyprus where biking is not preferred, while it may be underestimated in countries like the Netherlands, where cycling is much more common. These methods are applied to different geographic locations; thus, users exhibit significantly different travel behaviour depending on the location where the survey is taking place. Currently, any existing variability points are resolved during the design of the algorithm, where a set of possible parameters is tried out per case and decided upon per case. For instance, when using Density-based spatial clustering of applications with noise (DBSCAN) per each user per each day trip, the parameters are selected depending on the produced silhouette score, after clustering has taken place using a range of different parameters.

Another consideration would be to develop a close-to-real-time estimation of trips, instead of using batches of data. For example, the estimation by Pythia could only consider the past few hours of tracking or could be triggered only when a user stands idle. A real-time estimation of trips is also a possible approach, and it would require a drastic re-vamp of the ML module, to perhaps transition to a stop-detection based algorithm or other real-time methods. Once the data from MobyApp is transformed into meaningful trips by applying the above-described process, it is passed into HARMONY MS to be used for modelling purposes. Harmony-MS has been the outcome of the HARMONY H2020 project and constitutes a TRL 7 product which aims at satisfying different objectives corresponding to the different goals of actors of transport modelling sphere. It is a tool which aims at streamlining the modelling process for the different individuals included in the modelling circle.

There have been 3 main actors defined:

• Transport planners are the users who are interacting via the graphical user interface of the HMS and perform comparative analysis of transportation cases, interventions, and policies.

• Transport modelers are the users specifying the models of the available transportation cases, interventions, and policies that transport planners can compare. They are also interacting with the integrators to ensure the validity of the comparisons.

• Integrators (Components developers) ensure that the different platform components (transport models and simulators) at different levels (strategic, tactical, and operational) are correctly integrated to the HMS. The transportation cases, interventions, and policies that can be compared via the HMS, and hence the capabilities of the platform available to transport modelers and ultimately transport planners, depend on the components that are integrated to it. Integrators are the actors that are extending the capabilities of the HMS.

• Main Extension points for Harmony MS can be categorised into three major groups: Transferability of models and data means that whenever a city needs to test an intervention, but has no data, via Harmony MS they can use data other cities to extrapolate the impact of this intervention to the city in question.

• Adaptiveness stands for enabling the flexibility to select the needed tools, methods, and outputs. A city should be able to select the best model out of a pool of models that have the same purpose. In addition, they also need to select the most appropriate set of KPIs and related visualisations for the presentation of results.

• By streamlining, Harmony MS aims to provide a seamless environment for the users by adaptive error messaging. User-related actions and personalisation are also options to consider since different cases come with different requirements. Variability points for Harmony MS refer to the different datasets and different data needs across cases, as matching data needs to the available data is sometimes a significantly hard task. Different algorithms should be able to be selected, depending on the available models and the given scope. A set of possible visualizations must be available, depending on the output of the models and the needs of the user. There are various internal transaction surfaces, where endpoints can be created for facilitating third party access to data streams and internal endpoints.

The main steps of the process are the following:

1. Pre-processing

a. The MobyApp collects raw GPS data and feeds it to the backend in batches (not in real-time currently). b. Then an ML module (I.e., Pythia) utilises machine learning techniques to identify segments at which the user is stationary or moving.

2. Analytics a. For the moving segments: identification of the mode of transport and the type of public transport line is being used (via GTFS). b. For the stationary segments: identification of the type of activity (via POIs) or whether the user was at a Public Transport stop (GTFS) c. For the modelling tools on Harmony-MS, selection from pre-existing models, creation of models according to available data, pairing between inputs and outputs of modelling modules.

Visualization workflows a. On the mobile app: visualizations on map and on timeline (this currently built-in on the mobile applications). b. On the client web app of MobyApp: KPIs, metrics, statistics, maps. c. On Harmony MS: a selection of more advanced KPIs.

Exploited datasets & data acquisition methodology Datasets that are collected by MobyApp require caution since user locations are considered sensitive data according to GDPR. The reason being that in some cases, a GPS trace can be uniquely pinpointed to a specific individual. For Harmony MS, the datasets are primarily secondary data used for modelling purposes. For the ExtremeXP project, the available data coming from the data collection pipeline are:

a. Collected raw GPS data that are fed to the backend in batches (currently not implemented in real-time) b. Annotation from users from their verified trips

On the Harmony MS front the datasets are: a. Sample models from past case studies. b. Pipelines and related association between datasets and models. c. Datasets used as inputs to the different models.. d. Datasets - expected outputs of the different models

2. CLIMATE AND ENVIRONMENT

PISTIS: PI 4. PISTIS WEATHER HUB.

PISTIS will be tested and validated in real-life settings in 3 large-scale demonstrators' hubs in Greece (Mobility and Urban Planning), Spain (Energy) and Austria & Germany (Automotive). Each "demonstrator hub" has at its centre a core domain and involves many different partners of the consortium (having the role of different actors in each scenario), not all of them belonging in the same sector or domain, to demonstrate the applicability of the solution in both vertical as well as cross-domain/sector scenarios. As such, the PISTIS demonstrator hubs include representative and complete value chains in cross-domain settings as well, to validate data sharing, exchanges and transactions among them. Aside the three demonstrator hubs, **PISTIS will also be deployed in the UBIMET partner to serve the demonstrators with weather data,** one in FHG to serve stakeholders with selected open data ingested form the EU Data Portal and one in SPH, to facilitate the activities of the Living Lab and those of the training.

Historical weather data (geo-spatial data, - 40 MB per year per weather station); Data on current weather situation (low frequency); Air quality data (15 MB per year per measurement station) and from The MISTRAL platform aims to process, preserve, and provide, through an Open Data Portal (ODP).

DATAMITE: DA 6. WEATHER

The MISTRAL platform aims to process, preserve, and provide, through an Open Data Portal (ODP), facilities for the re-use of meteorological datasets at a national level, thus facilitating value added service using HPC resources. The aim of the ODP is to collect in a centralized digital repository meteorological data from observation networks and those from the civil protection radar network. Currently, the platform enables a wide range of services.

However, there is still much room for enhancement in terms of increasing data monetisation by widening the range of target users, improving data ingestion performance, and implementing quality inspection modules.

Description of the pilot: Currently MISTRAL lacks features that include lack of data discoverability and quality checks modules, and services that allow the use of available public data to generate AI based prediction models. Several new modules will be developed within the DATAMITE framework. Data sharing policies will be established to expose the climatic data to more users/companies working in weather sector.

From <https://datamite-horizon.eu/portfolio/connecting-mistral-to-the-eu-ai-on-demand-platform/>

FAME: FA6 CLIMATE-RELATED INSURANCE.

The scope of the FAME Pilot 6 is to develop novel climate aware property insurance products. Hence, it will leverage the value-added datasets in FAME in order to project valuations of real-estate assets based on local-level predictions as well as to calculate climate aware VaR of entire portfolios of assets possessed by insurers, including bonds, stocks and real-estate assets. Finally, Pilot 6 will analyse different portfolios of assets in the light of climate change.

GRAPH MASSIVIZER: GM 3. GLOBAL FORESIGHT FOR ENVIRONMENT PROTECTION

Global foresight for environment protection focuses on geopolitical and business aspects of ESG, 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. A contextual graph built through data from the Common Crawl, Linked Open Data Cloud, and global media news provides unique insights into mass media's convergence of the three societal systems (economy, politics, science).

Objective

Create and mine a FOR-MG, built based on media news information and enriched with Common Crawl Web and Linked Open Data Cloud.

Result

Environment Protection Foresighter SaaS for subscription alerts.

The Global Foresight for Environment Protection Use Case in detail

Global foresight for environmental protection focuses on the geopolitical and business aspects of ESG, which includes climate action, responsible production and consumption patterns, clean water, and affordable energy.

This approach involves analysing news articles and extracting information about events happening in the world, particularly those concerning companies mentioned in the articles.

With the help of Graph-Massivizer, past events will be used to compute potential future events to detect trends and anomalies.

Use case overview

The main objective of this use case is to identify individual relations related to global companies in online media mentions. These relations are related to various topics concerning environmental protection and business operations. Based on these extracted relations, predictions about potential future events can be made. These predictions will be based on identified generalized sequences of events.

Using Graph Massivizer, we can identify shared sequences of events across multiple companies and use those patterns to predict new events based on a sequence of recent events about a company. Besides predicting future events, the aim is also to identify specific trends and anomalies in the data by observing data in individual industries, sectors, and geographies.

The outcome of this project is the ability to identify common sub-graphs related to individual companies in large graphs of data. These patterns can be used to predict, as accurately as possible, likely future events.

This Use Case relies on cutting-edge technology to help organizations predict future events based on past data patterns. Our workflow is based on the initial provision of data as JSON objects, which are then converted to a graph structure using an established ontology. The Graph Scrutinizer is our main component for computing patterns and trends, and it operates on data imported into Graph Massivizer. By leveraging its powerful capabilities, the Graph Scrutinizer can predict potential future events for a new subgraph based on common patterns. The Use Case relies primarily on Graph-Inceptor for ingesting new data and Graph-Scrutinizer for extracting common patterns and trends from the graph.

Expected Outcomes of the Use Case

The results of our use case are reported in the form of a list of potential events predicted for a given input sequence of past events. Each predicted event has an associated probability, giving a clear sense of the likelihood of it occurring based on the frequency of the pattern in past data.

Our prediction component is the perfect addition to your existing product, seamlessly integrating with it to offer you an additional service for an extra charge.

We are committed to offering this service to organizations in the geopolitical segment, hedge funds and investment companies, and general companies interested in monitoring their supply chain. As the primary stakeholders in this use case, analysts will benefit from our technology as it enables them to understand what is happening globally and the potential consequences of these events. By leveraging our cutting-edge tools and technology, you can be confident in your ability to predict future events and make informed decisions that will keep you ahead of the curve.

From <https://graph-massivizer.eu/project/global-foresight-for-environment-protection/>

UPCAST: UP 3. SHARED PUBLIC ADMINISTRATION FOR CLIMATE CHANGE AT THESSALONIKI METROPOLITAN AREA

Description:

The availability of environmental data has the potential to change the ways in which cities are governed for sustainability and climate change mitigation and adaptation. More than ever, better environmental data is required to address urban challenges to the climate crisis. The availability of such data is critical to improve the monitoring and management of urban systems, as well as enabling robust assessments of policy-making and planning interventions. This pilot concerns the 11 cities that comprise the metropolitan area of Thessaloniki and their need to realise data-driven environmental policy-making. Working under their umbrella organisation, the Major Development Agency Thessaloniki (MDA), and Open Knowledge Foundation Greece (OKF Greece) this pilot will use of the UPCAST plugins for integration and exchange of all data related to its environmental use case.

Major Development Agency Thessaloniki (MDAT S.A.) is one of the UPCAST partners, responsible for the Thessaloniki Use Case.

https://mdat.gr/2023/10/13/upcast-thessaloniki/

Contact person: Paraskevi Tarani

From <https://www.upcast-project.eu/shared-public-administration-for-climate-change-at-thessaloniki-metropolitan-area/>

EXTREME XP: EX.3: SITUATIONAL INTELLIGENCE AND DECISION MAKING FOR PPDR (PUBLIC PROTECTION AND DISASTER RELIEF) (ADS)

3.4.1 Context description and objectives In the third UC, the scenario focuses on the significant role of extreme analytics coupled with data personalisation and presentation for the detection, characterization, and operation of fire scenarios in an urban area, via an enhanced situational awareness and critical human operators' decision making. We will explore ExtremeXP capabilities, such as the experience-centred and assisted decision making, trustworthy and explainable AI, extremely precise prediction, in building strong situational awareness, as well as a precise prediction of operational evolutions that first responders will be able to share through their Mission Critical collaboration platform. This platform allows on-field users form different organisations to communicate with talk groups, or video groups, or messages with the control centre. All data collected from sensors are available on the control centre application on a map or in different menus (type of user, location, status, videos, pictures, etc.). The ExtremeXP framework will be applied to demonstrate the added value that combined complex analytics, personalization, assisted decision making, and mission critical

services could provide almost real-time in the scope of improving operational situational awareness and decision making, based on the mission, organisation, role, location, and context of the first responders.

• Detection phase: Currently operators survey multiple screens displaying videos from on field cameras. When an operator detects an issue, such as a fire or smoke, he/she would have to continuously monitor the situation on the camera, identify the location, choose users on the field and alert pertinent users. An analysis of all the videos with a form detection capability will send off alerts to the command control with all relevant data, such as pictures of the detected fire, location, type of object/vehicle in fire of detected, fire progress, etc. to allow the operator to validate the alert.

• Decision making & actuation triggering phase: In case of a validated alert, the operator would have firstly to detect all pertinent procedures (forest fire, vehicle/truck in fire, etc.) and secondly choose the corresponding categories of users on the command-and-control map. This choice entails a series of actions: search, and selection of the closest users in each selected category and secondly a status check of the availability of each user. If a user is available, he has to contact by phone or message each user and check if the user can reply to his request. In case of negative answer, the operator has to select other users of the same category closer to the alert location and contact this user in order to choose another unit. In case of positive answer, the operator sends required tasks to the user. Within the ExtremeXP's framework, the main objective is to decrease the duration of the decision making process, by providing the operator with a better view of the current situation, thus enhancing situational awareness and facilitate the actuation:

• Depending on the type of alert, the system will display a selection of actions to be launched on the control centre by the operator. Elected, on-field users will be displayed on the control centre map to facilitate the appointment of selected users as available and close to the alert location and allocation of corresponding tasks.

• Next the operator will have to validate the appointment, by clicking the validation or propose another choice on "another choice menu"

• Then the system will directly send a message to all selected users with a request for an answer. If the user responds to the request, all pertinent data will be sent to his/her smartphone.

• If the user rejects the order, a message is sent to the closest available user of the same category for a new selection. In this UC, we identify two classes of users: the operator, who is in charge of processing alerts received in the control centre and makes decisions, and on-field users in charge of responding to those decisions.

3.4.2 Experiment definition

The threat of fire is a significant concern for people all around the world. Traditional methods of detecting fire often rely on sensors that use fire parameters to identify the presence of

flames. Unfortunately, these sensors can take some time to detect the fire, which can be problematic as the flames can quickly grow out of control and cause significant property damage. To address this issue, a more efficient system for detecting indoor and outdoor fires should be developed, using video surveillance cameras. In recent years, computerized monitoring methods have become increasingly popular for identifying fires. However, accurately determining whether a region is actually on fire remains a challenge due to the changing colour of flames from red-yellow to nearly white.

The system will contain the following modules:

• Detector: Detecting and localizing fire.

• Recommender: Recognize the context and disseminate information until the appropriate first responder unit.

• Optimizer: Optimize first responder flow and response time.

The experiments will be split into three separate phases. The first phase will consist of effectively detecting the fire and pre-event situations as smoke with object recognition algorithm by considering spatial and temporal complexity. The second phase will consist of context recognition for situation as forest fire, car fire, or narrow street fire by representing content in visual relationship as Spatial-Temporal Graph Neural Network and suggest the appropriate first responder unit. The last phase will consist of managing and optimizing first responder, unit interventions by formulating as resource allocation problem with availability and localization constraint related to real world road constraint.

Early-stage fire detection and localization The main objective of fire detection is to immediately identify the presence of fire and signal an alarm to take swift action. As stated in |10], it is crucial to detect fire as quickly as possible to prevent any serious damage to property or lives. Various effective methods for detecting fire are available, but detection speed is of utmost importance. Therefore, it is essential to develop a reliable system using the available resources that can detect fire instantly and alert authorities promptly to prevent any potential hazards. The challenges of accurately detecting fire and sounding alarms urgently need to be addressed to allow for sufficient time for evacuation and extinguishing the fire.

This innovative system proposed by [11] uses three key features of fire, including colour, motion, and shape, which are enhanced with fuzzy logic for added accuracy. The main goal of this system is 38 to minimize false positives in fire detection and improve the effectiveness of the model in both indoor and outdoor settings.

3. ENERGY

PISTIS: PI 2.1 INCREASE THE HOSTING CAPACITY OF THE GRID.

The demo case focuses on the effective resolution of distribution grid congestions, through innovative flexibility-based grid management mechanisms. Smart metering data provided by local prosumers, together with distributed generation data (photo-voltaic, PV) and SCADA

(Supervisory control and data acquisition) information from the Distribution Service Operator (DSO) will be jointly analysed to extract accurate demand and generation forecasts (in the short and mid-term). Also to estimate anticipated events in the distribution grid thanks to the digital twins. The required flexibility to effectively address them will be procured by the Aggregator.

Ambition of use-case

The **aim** of this case study is to Increase the hosting capacity of the grid by accommodating DERs smoothly, whilst also negotiating new short-term flexibility products.

Thanks to the short-term flexibility market, one-off events due to grid overload can be avoided by making the flexible elements work, so that during the rest of the time it is possible to accommodate more distributed generation in the grid, thanks to controllability.

Target Audience Description

In this Use Case, the participating actors include all those from the Energy Hub, where Omie is the MO (Market Operator) and provides a local flexibility market platform, with Cuerva as the DSO, and Bamboo as the aggregator. In the case of Cartif, it will act as the developer providing the technology that in a real-world setting would be part of the DSO. Additionally, Cartif will also serve as an independent flexibility service provider, a role to which the aggregator turns to create the flexibility offering. UBIMET will provide meteorological data to Cartif for the generation of consumption, generation, and network event prediction data.

Summary of status after Alpha testing phase

Following the tests carried out with the PISTIS platform, using datasets already available and currently being used in the parallel development of the use cases, it has been demonstrated that the platform's integration can streamline certain processes, such as data anonymisation and the handling of specific issues, while adding significant value through data enrichment.
In addition, the ability to "offer" these datasets for usage or exchange simplifies their control and traceability within the development process, where maintaining a structured development backlog is essential.
Once Use Case 1 was fully developed alongside the platform, it was also demonstrated that it can be fully integrated into the short-term flexibility market workflow, verifying data communications and maintaining a comprehensive log. This serves as an example not only for Use Case 1, but also for Use Case 2, which focuses on the long-term flexibility market.

After conducting tests with various datasets related to Use Cases 1, 2, and 3, a TAM (Technology Acceptance Model) questionnaire was completed to gather feedback on key aspects of the platform, including its usability, interface, and areas for improvement.
In general, the platform was perceived as user-friendly and intuitive, with the expectation that it will not only add value to the data itself but also to the processes derived from the market trial, supporting more efficient and structured workflows.

PISTIS: PI 2.2: INVESTMENT DEFERRAL.

This Use Case focuses on investment reduction thanks to the available data and the flexibility of the system. The most benefited actor is the DSO that can plan the grid operation, including the use of flexibility agents, and create a long-term local flexibility market. In this case, the

market is planned to create commitments with the DSO for the following months or years, thus effectively providing a new management tool for the DSO. These data are interchanged over time thanks to the platform, being thus registered, and if it is necessary, apply tradability or penalties in the validation process.

Ambition of use-case

The **goal** for this use-case is to focus on studying the impact of utilising data and flexibility in grid development, as an alternative to traditional approaches that address issues like overvoltage and congestions. There is the need to facilitate the exchange of valuable information and resources, leading to more efficient grid management and cost savings in the energy sector.

Target Audience Descriptions

The interaction with the platform and all the processes in the flexibility market are the same as stated above in the use-case #2.1 Hosting Capacity of the Grid. Similarly, the data to be shared are the same and the only difference is the period of the events, and the flexibility offers.

Summary of status after Alpha testing phase

In the case of Use Case 2, both the data inputs and the overall structure of the use case mirror those of Use Case 1. The key difference lies in the time horizon: while Use Case 1 focuses on the operation of a short-term flexibility market, Use Case 2 explores the dynamics of a long-term flexibility market, which introduces different requirements and perspectives, particularly in terms of forecasting, planning, and investment decision-making.
As both use cases are still under development, it is expected that the specific differences will become clearer once each is fully implemented. For now, the testing of the PISTIS platform for Use Case 2 has been performed using the same datasets as in Use Case 1, as they are technically valid for both applications.
This shared structure has allowed the team to validate core functionalities of the platform — such as data uploading, processing, anonymisation, and interaction with services — while also identifying potential adjustments that may be needed once the long-term market-specific data becomes available.
The ability to reuse and adapt data pipelines between use cases highlights the flexibility and scalability of the platform, reinforcing its role as a central enabler in future market-oriented data ecosystems.

PISTIS: PI 2.3 P2P TRADING BETWEEN USERS OR ENERGY COMMUNITIES.

Currently, in view of the promotion of self-consumption, shared self- consumption and the growing number of new energy communities, users are increasingly wishing to be aware of the origin of the energy purchased from the grid and the destination of the energy fed into the grid. At present this information is unknown, beyond the guarantees of origin offered by some traders. Peer-to-Peer (P2P) trading is proposed to allow the user to choose which grid user to sell to or buy from, thus enabling the creation of these peer-to-peer exchanges and, therefore, empowering the users to know exactly what they are consuming and where it comes from.

This could be enabled by new platforms operated by the MO, where different resources trade energy in local level. In this case, the DSO is not seeking to solve a problem with the supply quality, but it cannot allow events to occur due to exchanges between users that could lead to unexpected situations. Therefore, the distributor's role is to act as a controller of these exchanges for prior validation before technical dispatch.

Ambition of use-case

The **aim** is to create peer-to-peer exchanges with a new platform operated by the MO, to facilitate free negotiation among energy resources and assets (that is -without any DSO call).

It is also interesting for the DSO to study how P2P affects not only the grid, but also short and long-term flexibility markets, and the need to intervene in P2P to ensure the quality of supply.

Target Audience Descriptions

In this use case, the primary roles are covered by the Market Operator and the Aggregator, using the short-term market platform as a free market platform. In this scenario, end-users could place energy demand and supply on the platform outside of the time periods when the DSO requires flexibility for specific moments. However, it is essential to consider how these P2P exchanges affect the grid and whether they generate new flexibility requirements.

Summary of status after Alpha testing phase

In the case of Use Case 3, the same user consumption and network data previously used in Use Cases 1 and 2 have been employed for testing purposes. While the objective of Use Case 3 differs—focusing on a different dimension of data usage within the energy system—the underlying datasets remain relevant and applicable due to their foundational role in representing real operational conditions.

This approach has allowed for the reuse of validated data pipelines, streamlining the integration with the PISTIS platform and enabling consistent testing of core functionalities such as data upload, processing, and enrichment. The use of familiar datasets also facilitated the assessment of the platform’s performance in a comparable context, making it easier to identify usability improvements, validate system responses, and ensure interoperability across different use cases.

Although Use Case 3 introduces its own specific analysis and goals, leveraging the same datasets has proven effective for the initial validation phase, and sets a solid baseline for future expansion with more targeted or domain-specific data inputs. This further demonstrates the flexibility and adaptability of the PISTIS platform to support a variety of energy-related data use cases.

PISTIS: PI 2.4: MONETISATION OF ENERGY EXCHANGE DATA

This Use Case will explore the possibility to monetise the data, which is owned by the different actors participating in the energy exchange over the grid, to third parties for different uses, such as energy as a service (EaaS). The potential consumers are aggregators, installers, energy service companies, retailers, consultancy or advisory companies, private research groups and universities, EV charging companies and software companies. They would be able to optimise their existing services and develop new ones thanks to the additional information obtained through use of the data.

Ambition of use-case

The **Goal** for this case study is to define the process of data valorisation to target potential consumers of energy assets, who may use them for Energy as a Service (EaaS) including aggregators, installers, energy service companies, retailers, consultancy firms, research groups, universities, EV charging companies, and software companies.

Target Audience Descriptions

In this scenario, every member of the Energy Hub acts as both a data owner and a data seeker, collaboratively exploring potential external entities that might be interested in acquiring the data produced.

Summary of status after Alpha testing phase

In the case of Use Case 4, the platform has not yet been tested with datasets specifically tailored to this use case, as its main focus lies in the valorisation of data, rather than in operational processes or technical simulations like the previous use cases. However, the initial testing phase using the shared datasets from Use Cases 1–3 has proven highly valuable in assessing how the platform handles data ingestion, processing, enrichment, and traceability.
Although no direct monetisation scenarios have been tested yet, the experience so far has provided a clear overview of the technical and economic value that PISTIS can add to raw datasets. Features such as data cataloguing, metadata generation, quality control, and usage tracking demonstrate how the platform can support the creation of standardised, reusable, and traceable data products — all of which are essential building blocks for the future development of data-driven business models.
This early evaluation has not only improved our understanding of the platform's potential in the context of data monetisation, but has also helped to establish the conceptual and technical foundations for the implementation of Use Case 4. Going forward, these insights will support the definition of relevant datasets, the identification of value creation mechanisms, and the design of data-sharing frameworks aligned with market and regulatory needs.

DATAMITE: DA 3 AND DA 4 COVERING THE ENERGY SECTOR

System planning is a critical task for Distribution System Operators (DSOs) like HEDNO.

At least 3 modules are required to meet this goal:

a) optimal sizing/siting methodology;

b) AI-based load and DER generation forecasts module;

and c) AI-based state estimation algorithms.

To perform these tasks, an Energy services provider, like CERTH, may require requesting & consuming data from a DSO through Energy Data Spaces. Sharing data with service providers is far from being an easy or automated process. The lack of automation not only complicates sharing data with service providers, but also exploring monetisation or collaboration options like the exploitation of data markets or data spaces.

Description of the pilot: DATAMITE facilitates the automation of data sharing through its different modules:

- With Governance tools, facilitating Findability and Accessibility to data.
- The Quality Module ensures that data meets sufficient quality criteria to be shared with third parties
- With Security tools, defining fine-grained access and authorisation policies, and facilitating anonymisation, privacy preservation or encryption processes.
- With the data sovereignty component and IDSA components integrated in the sharing module, that allows the definition and enforcement of precise terms of use.

Data sharing to DataSpaces will be validated by publishing data in Energy DataSpaces like, potentially, those provided by DATA CELLAR.

https://datamite-horizon.eu/portfolio/offering-data-to-service-providers-with-dataspaces/

UPCAST: UP 6 (TOOL) ENVIRONMENTAL IMPACT OPTIMIZER

Deliverable D3.4 presents the finalized environmental impact optimizer plugin, developed in task T3.3 of UPCAST, to enhance sustainability in systems that process data. It introduces two innovations: the Environmental Impact Estimator (REST API) for predictive energy consumption modelling and the Monitoring Tool with a real-time dashboard for tracking workflow energy usage.

This deliverable marks the completion of the environmental impact optimizer plugin (D3.4), developed as part of task T3.3 in work package 3 of the UPCAST project. It provides a public record of the plugin's development, supported by practical demonstrations that showcase its functionality. Building upon D3.3, which introduced the initial plugin architecture and API, this deliverable advances the work with two significant contributions: the environmental impact estimator REST API and the environmental impact monitoring tool.
The environmental impact estimator REST API builds on re designed and extended experiments analysing processing energy consumption, while the monitoring tool includes a visual dashboard for real-time tracking of energy usage in workflows. These advancements aim to align with UPCAST's mission of fostering sustainability in data-driven systems, ensuring compliance with European environmental regulations, and promoting energy efficient practices. This document is accompanied by a video demonstration showcasing the use of the final version of the plugin, the monitoring tool, and the estimator tool, as well as the experiments conducted to gather data for the estimator.
The video is presented here: https://www.youtube.com/watch?v=V1He0v5BeHk.

Environmental Impact Monitoring Tool This chapter details the features and development of the environmental impact monitoring tool. The main objective of this tool is to provide continuous observation of key environmental metrics during the execution of a data processing workflows (DPWs). It begins with presenting the purpose and scope of the tool and outlines its system architecture and core components. The chapter also provides guidelines for configuring and using the integrated data collection and visualization tools that support the monitoring process. 3.1 Purpose and scope of the environmental

monitoring tool The environmental monitoring tool offers a clear and user-friendly solution for tracking energy usage in real-time within a data processing workflow (DPW). It allows users, such as data consumers, to monitor and manage power consumption remotely, providing continuous insights into energy performance. These insights are displayed through Grafana dashboards, which visually present key energy usage metrics and power trends. This empowers users to pinpoint inefficiencies and implement targeted optimisation effectively.

4. MARKETING/MONETISATION

Both FAME and PISTIS have produced monetisation tools.

UPCAST: UP 1. DIGITAL MARKETING DATA AND RESOURCES

Description:

Today, data on the performance of digital marketing campaigns is only used to determine the most adequate optimization actions to increase the engagement and ROI of the active campaigns. However, there is high value in the data that is not exploited at all. By analysing this type of data it is possible to determine the real user interests in a wide variety of business verticals (classified based on the Google categories taxonomy) and locations. For that reason, this pilot is developing a new data-as-service business model where data consumers can decide which data is needed and how the data and insights should be delivered.

The main components of the pilot are:

(i) the service request represents the interface where the data consumer can check some examples of the data sets, sign and log in, and define the service request based on a pre-defined set of filters and features;

(ii) the service signature is the component responsible for calculating the price of the service requested, define and sign the contract and

(iii) the service delivery, where the query is automatically generated based on the data consumer needs, and the different services are implemented (data sharing, and reporting).

In summary, this marketing related data as a service offers the possibility of getting access to new information and insights depending on the data consumer needs, who will be able to know more about real topics and interests in their targeted markets.

https://www.jot-im.com/innovation/upcast/

Contact person:

fernando.perales@jot-im.com

UPCAST: UP 5 CACTUS

Description:

Cactus creates customized strategies and experiences capable of making any business distinguish in its industry, turning it into a leading force.

Cactus digital marketing data will be used by the Upcast Project in order to better understand how these data will be monetised. Apart from this, **Cactus** wants to better understand how the data economy will affect the business environment in general.

Website: www.cactusweb.gr/en *Contact person:* Alexandros Lemperos

From <https://www.upcast-project.eu/cactus-2/>

ENRICHMYDATA: EN 1 MARKETING CAMPAIGN MANAGEMENT.

Digital marketing campaign management involves both configuration and daily optimization of many variables related to keywords and ads. The marketing campaign configuration requires to be aligned with the social behaviour and interests strongly impact the final performance (visits) and profit (revenue). They need to be updated with new keywords very frequently as they reflect the user trends. For that purpose, JOT must relate to external data sources to automate the keyword dataset update. Therefore, more impactful and profitable marketing campaign management requires access to new data and the implementation of analytical solutions to enrich the information linked to the keywords and support all the structural configuration and optimization actions for the campaign

From <https://enrichmydata.eu/business-cases/>

5. INDUSTRIAL APPLICATIONS

ENRICHMYDATA: EN 2

Bosch proposes a business case from the automotive sector in the context of Catena-X1, an automotive alliance for secure and standardized data exchange that aims at enabling continuous data reuse and sharing for all contributors along the automotive value chain. The business case is centred around automated robotic welding where BO provides components and software solutions for welding machines that are used by factory owners, including BO and its renown car producers, for building body-in-white (BIW) of cars and other car related components. The goal of the business case is to create a data enrichment pipeline for the automation of data preparation for AI-based analytics related to automatic welding robots.

https://enrichmydata.eu/business-cases/

ENRICHMYDATA: EN 3 SERVICE DATA ENRICHMENT FOR SMART MAINTENANCE

Philips designs and produces capital intensive medical imaging systems (e.g., Magnetic Resonance Imaging, Computed Tomography) which are used for diagnosis and image guided therapy in hospitals. Over the years the functionality of this equipment has grown dramatically, which makes maintenance of system, design and verification methods harder with every system release. The main objective of the business case is to enrich the un/semi-structured datasets using external data sources. The most critical information related to system failure can be extracted from the semantically enriched data and used in downstream tasks to help engineers perform the system troubleshooting remotely or on-site.

From <https://enrichmydata.eu/business-cases/>

ENRICHMYDATA: EN 6 INDUSTRIAL DATA ENRICHMENT FOR MINERAL PROCESSING OPTIMIZATION

Following the integration of industrial internet, Big Data, AI, and other new generation information technology with the traditional process technology in the mining industry, the digital and intelligent upgrade of mineral processing has become the development trend. Combining the digital twin technology and characteristics of the mineral processing industry, a digital simulation platform will enhance the application effect of equipment improvement, process optimisation and intelligent control in the whole production cycle of a concentration plant. Industrial data enrichment for mineral processing optimisation will help BGRIMM increase its competitiveness in optimal service for mineral processing plants worldwide.

From <https://enrichmydata.eu/business-cases/>

GRAPH-MASSIVIZER: GM 2 GREEN AI FOR SUSTAINABLE AUTOMOTIVE INDUSTRY

The automotive value chain involves extreme data flows of heterogeneous, distributed, fast-growing, disconnected, or hardly compatible information. ML methods face new challenges and opportunities to holistically analyse the massive and unprecedented data integrated across these chains, to support decisions that fundamentally change automotive manufacturing processes towards a sustainable, circular, and climate-neutral automotive industry. **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 ML 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).

Objective

Predict “best” production configurations for a given BiW type and welding machines over simulated data with predictable manufacturing KPIs (BiW quality).

Result

Manufacturing Graph Generator with extremely controlled scaling of Man-MG in multiple dimensions (number of sensors, production lines, processes).

The Green AI for Sustainable Automotive Industry Use Case in Detail

The automotive industry involves a complex network of data flows that are often incompatible and diverse. To address this issue, we plan to adopt the Graph-Massivizer toolkit, which uses Machine Learning (ML) methods to analyse this data holistically. This will pave the way for a sustainable, circular, and climate-neutral automotive industry.

Graph-Massivizer uses a unique graph-based encoding that captures multiple value-chain stages, improves prediction accuracy, and anomaly detection. This enhanced analysis will help to reduce defects and waste, promoting industry sustainability. By integrating Graph-Massivizer’s graph-based ML methods with digital twins, we aim to improve diagnostic efficiency and scalability, surpassing more expensive alternatives such as continuous sensor monitoring.

The primary objective of this Use Case is to develop a solution that can predict the best production configurations for a given material type and welding machines. We will use simulated data to achieve the desired quality levels.

The Use Case aims to improve the accuracy of predicting the produced spot diameter based on an input welding program during simulation. This will help in determining the quality of the welding result and predicting the parameters required to produce a specific welding quality.

The logic or workflow is as follows. Firstly, the sensor data, welding machines, welding programs, and quality specifications are onboarded. Then, these are mapped to the welding quality ontology. Next, an algorithm is chosen (such as a neural algorithm) and a graph embedding model is trained. Once the algorithm is selected, the necessary hardware resources for the algorithm and data are acquired. Finally, the algorithm is executed, and its performance is monitored.

We plan to use the GM platform to effectively manage the streams of data produced by the numerous sensors that monitor the welding process and the machines used. This data will be integrated with other sources in the graph, and symbolic and sub-symbolic reasoning will be applied on the graph. To facilitate the integration of the sensor data in the KG, the Graph Inceptor is crucial. The Scrutinizer will be utilized to mine the produced graph. We also intend to use the Optimizer and the Greenified to pick the best resources for the selected algorithm. This will help us avoid overusing resources and most importantly, reduce failures and repeated computations caused by insufficient resources.

Expected Outcomes of the Use Case

The expected results will be obtained by generating KG embedding of the constructed welding graph and using the embedding model to predict the quality of the welding spot. This will also help in determining the most suitable program for a given material and spot.

The results will be presented in the form of an interactive view of the developed ontology and a graph generated from synthesized data. Additionally, there will be a query interface available for stakeholders to test the system and obtain the desired results.

The plan for exploitation involves the extension of the project output to create an internal platform for industrial data that is similar to the welding scenario. The aim is to adapt and expand this output to more Bosch plants, thereby addressing the needs of the industrial floors. The stakeholders involved are multiple Bosch plants that are involved in welding procedures. The sensor data is expected to provide inferred information that can assist technicians, engineers, and managers in optimizing the production process.

EXTREMEXP: EX5 MANUFACTURING

How to maximize machinery uptime in an industrial plant ?

ExtremeXP leverages Industry 4.0 techniques, such as predictive maintenance, remaining lifetime assessment, and critical machine element diagnosis, to ensure maximum machinery uptime. These methods help maintain high system availability and enhance competitiveness by providing precise insights into machine performance, enabling operators to act before failures occur.

How can ExtremeXP help prevent erroneous warnings and maintain production capacity ?

To avoid false alarms that can reduce production efficiency, ExtremeXP offers a solution that focuses on delivering accurate information to operators. The goal is to provide a dashboard specifically designed for the technical service team, helping them make informed decisions about incipient or imminent failures of critical machine components like heads, ball screws, or axes.

ExtremeXP's approach includes a comprehensive dashboard that gives the technical team the ability to monitor and assess the health of key machine elements. With access to real-time data, teams can make proactive decisions regarding potential failures, ensuring minimal downtime and improved system reliability.

6. FINANCE

GRAPH-MASSIVIZER: GM 4 GREEN AND SUSTAINABLE FINANCE

One goal of the Graph-Massivizer project is to eliminate the constraints posed by financial market data providers, such as limited volume, restricted accessibility, high costs, and limited historical relevance of historic data. The project achieves this by facilitating the rapid semi-automated generation of realistic and cost-effective synthetic financial datasets in massive volumes (petabytes), with limitless scalability and accessibility.

What is synthetic data? Synthetic data is artificially generated data that mimics real-time and historical data in terms of essential characteristics, capturing and reflecting the statistical properties of real-world data. In the context of investment and trading, synthetic data is useful for simulating various market conditions and investment scenarios, providing a rich and diverse dataset for analysis.

It can be used for different main topics at the heart of investment and trading, such as:

Stress Testing Financial Algorithms

Stress testing financial algorithms means that algorithms are exposed to a multitude of extreme market conditions, sudden economic shifts, and unforeseen geopolitical events. Rigorous and in-depth stress testing helps identify potential weaknesses and vulnerabilities, enabling traders and investors to refine and fortify their algorithms pre-emptively.

https://graph-massivizer.eu/project/green-and-sustainable-finance/

Through FAME's seven pilots, project partners seek to test and validate FAME technologies in real-world environments, offering solutions for financial institutions and other end-users. Each pilot engages with different aspects of the financial services ecosystem, including citizen wallets, EU fund applications, ESG scorecards, family financial planning, and more. Check out all of FAME's pilots below.

FAME: FA1 FAMLY – A POWERFUL FINANCIAL RECOMMENDATION ENGINE FOR FAMILIES

Pilot #1: FaMLy – A powerful financial recommendation engine for families

FAME: FA 3. PERSONALISED COLLABORATIVE INTELLIGENCE FOR ENHANCING EMFI SERVICES

Pilot #3: Personalized Collaborative Intelligence for Enhancing EmFi Services

FAME: FA 4. EU FUNDS APPLICATION PROCESS MADE EASY

Pilot #4: The EU Funds Application Process Made Easy

FAME: FA 5. ESG Scorecard Ranking & Sustainable Portfolio

Pilot #5: ESG Scorecard Ranking & Sustainable Portfolio Optimisation

FAME: FA 6. Climate Aware Property Insurance

The scope of Pilot 6 is to develop novel climate aware property insurance products. Hence, it will leverage the value-added datasets in FAME in order to project valuations of real-estate assets based on local-level predictions as well as to calculate climate aware VaR of entire portfolios of assets possessed by insurers, including bonds, stocks and real-estate assets. Finally, Pilot 6 will analyse different portfolios of assets in the light of climate change. In the following table, the details of the collected business requirements stemming from Pilot 6 are presented.

Pilot #6: Embedding Climatic Predictions in Property Insurance Products

FAME: FA 7. Assessing the Quality and Monetary Value of Data Assets

Pilot #7: Assessing the Quality and Monetary Value of Data Assets

PISTIS: Monetisation Task

Pistis has a task covering monetisation and placing value on datasets.

EnrichMyData: EN 4. European Register of Entities from Known Actions EUREKA

Spend Network has built and maintains the largest Open Contracting Data Standard Data (OCDS) database in the world, with over 180 million lines of data. OCDS is a data format that allows procurement data to be standardised, providing more robust linked data for the $13 trillion USD global public procurement market. The objective of the European Register of Entities from Known Actions (EUREKA) business case is to build the EUREKA tool as an open data list of organisations using Spend Network's proven experience of releasing procurement data from a free to use, open tool.

7. CITY COMMERCE

FAME: FA 2. Embedding Finance Services in a Personalized Citizen Wallet

The scope of Pilot 2 is to develop novel EmFi services in an urban context, where data from smart services are used to provide additional finance related services. To this end, Pilot 2 aims to create citizen's EmFi profile by leveraging existing parking data, onboarding other

behavioural data of citizens and by using citizens profiles to generate and provide personalised citizen-centric offers and recommendations. On the other hand, Pilot 2 aims to extend citizen wallet to a broader range of services such as transportation services, payments of fines, etc.

Analyse parking data in terms of locality, frequency, and time of parking FAME should optimize the data analysis part in order to be pinpoint parking habits

Request data from other organisations that have loyalty programs in order to analyse the rewarding mechanisms FAME should offer an added value to an external entity in order to add their data and offer an authorized and usable onboarding/registration procedure

Acquire knowledge and training on the platform so as to leverage the city's IT personnel competencies Request and onboard data from other organizations so as to leverage the services of the city e.g., traffic data Functional User, technical, infrastructure, FAME should offer training feature (e.g., webinars, user guide, MOOCs, online training sessions etc) Public administration, management

FAME should offer an added value to an external entity in order to add their data and offer an authorized and usable onboarding/registration procedure Parking Management, Marketing Functional User, technical, infrastructure,

Reduce the costs that refer to the management of the parking system of the city FAME should optimize the data analysis in order to identify solutions that can make the system less costly for the city Parking Management, Marketing Functional technical, infrastructure,

Analyse parking and behavioural data in terms of locality in order to design new services e.g. if a citizen owns a shop, then short term FAME should optimize the data analysis part in order to be pinpoint parking and other habits and localities Parking Management, Marketing Functional User, technical, infrastructure, parking for customers could be foreseen

Ensure the sovereignty of the data being traded FAME should have mechanisms to ensure data usage control policies Governance Functional Infrastructure

Correlate data sources to offer new services FAME should offer a view of correlated data and usable visualizations for a city employee Public administration Functional User, technical

Exploit data on parking payments to design specific offers e.g., discounts FAME should offer an analysis of parking payments and usable visualizations e.g., time, duration, location etc Public administration, management Functional User, technical

Request and onboard data from other organizations so as to leverage the services of the city e.g. traffic data FAME should offer an added value to an external entity in order to add their data and offer a authorized and usable onboarding/registration procedure Parking Management, Marketing Functional User, technical

Analyse parking and behavioural data in terms of locality in order to foresee the addition of other municipal services FAME should optimize the data analysis part in order to be pinpoint parking and other habits and localities Parking Management, Marketing Functional User, technical, infrastructure,

Analyse parking and behavioural data in terms of high/low demand in order to develop a dynamic pricing scheme FAME should optimize the data analysis part in order to be able to foresee parking demand based on historical data Parking Management, Marketing Functional User, technical, infrastructure,

Offer citizen wallet to citizens Develop a citizen wallet for citizens to consume services in one app Marketing Functional

Develop a dynamic parking pricing scheme Develop a dynamic pricing mechanism based on parking demand Parking Management, Marketing Functional User, technical, infrastructure,

Ensure the sovereignty of the data being traded FAME should have mechanisms to ensure data usage control policies

Pilot #2: Embedding Finance Services in a Personalized Citizen Wallet

PISTIS: 1.5: INSIGHTS FOR CITY COMMERCIAL BUSINESSES.

Added value services will be offered across specific areas of interest in the city of Athens based on the analysis and prediction of the load expected within the city, informing local businesses on how people are expected to move within the commercial zones of the city. The intention is to improve local entrepreneurship and boost businesses turnover, and to deliver services that can improve the mobility experience of the citizens and visitors of the city, which will lead to an optimisation of mobility services in specific areas (e.g. dynamic PT timetables, less queuing at touristic sites) and could help the Public Transport Operator to adapt, in the longer term, its routes and timetables with respect to touristic/heavy load destinations.

Ambition of use-case

The intention is to enable data exchange between the airport, the public transport agencies and the city to offer better services to citizens, to improve local entrepreneurship in Athens according to expected loads in specific areas and to leverage the mobility as well as the touristic experience of the city according to foreseen mobility flows.

Target Audience Descriptions

The actors in UC1.5 are the Greek cluster partners of PISTIS, namely AIA, OAG and DAEM that collaborate for the preparation and implementation of the Hub 1 Demonstrator. The actors can be identified as data owners and seekers, and they are divided as internal (the PISTIS partners) and external from other stakeholders, bodies and organisations.

Summary of status after Alpha testing phase

The Alpha version testing for the Hub was finalised at M28. The relevant guidance was shared with the Hub 1 partners in terms of links, accounts, credentials, steps to take, etc. The evaluation of the testing included a first demo testing of AIA and followed by the rest of the partners proceeding with their processes. DAEM tested the relevant factory (daem.pistis-market.eu). An account and the json file was shared prior to the testing.

A sample dataset was used from the registry of businesses in Athens converted to a compatible format of UTF-8 encoding.

Initially the sample was uploaded and the json file was copied. The response was sent back from the PISTIS platform. Then the testing involved the whole workflow of the Alpha testing, namely the Job Configuration, Workflow execution and Publish of the datasets in the marketplace.

Sample files were both uploaded and downloaded satisfactorily with an ease-of-use. The final steps of the workflow of Data enrichment and quality assessment are not implemented in the Alpha version.

In summary, the testing resulted in a simple interface, process and usable marketplace. Finally, a TAM questionnaire was completed.

8. PISTIS: PI 1.1, PI 1.2 AND PI 1.3 RUNNING AN AIRPORT

PISTIS: PI 1.1. BAGGAGE HANDLING MANAGEMENT.

Ground handlers receive baggage-related timestamps from the airport that enable proper baggage management and make it possible for the early identification of irregularities in the baggage delivery process to the passengers. Post-operation analysis of such data is very important to investigate further incidents or non-optimal performance of the baggage delivery process.

Ambition **of use-case**

To **improve operational performance** of baggage handling, reduce irregularities in the process and predict the probability of flights delays due to baggage handling. Efficient baggage handling reduces the risk of delays and coordination between the airport and ground handling would help mitigate operational anomalies.

Target Audience Descriptions

The actors in UC1.1 are the Greek cluster partners of PISTIS, namely AIA and Goldair that collaborate for the preparation and implementation of the Hub 1 demonstrator. The actors can be identified as data owners (AIA) and seekers (Goldair), and they are divided as internal (the PISTIS partners) and external - from other stakeholders, bodies and organisations such as other Ground Handling Companies operating in Athens Airport or state authorities.

Summary of status after Alpha testing phase

During the Alpha testing phase, AIA and Goldair successfully collaborated to test the core functionalities of the PISTIS platform for baggage handling management data exchange. The testing focused on establishing the data pipeline between AIA as the data owner and Goldair as the data seeker. Sample baggage-related timestamp datasets were prepared in CSV format and uploaded to the PISTIS Data Factory, where they were processed and made available through the marketplace.

The testing validated the platform's ability to handle baggage handling data securely while maintaining data sovereignty. Both organisations verified that the platform could ingest, process, and deliver the relevant baggage management data needed to identify potential

irregularities in the baggage delivery process. The data workflow execution was completed successfully, demonstrating the platform's core functionality for data exchange in this specific use case.

While the basic data marketplace scenario was validated, advanced features like data enrichment, quality assessment, and monetisation were not implemented in the Alpha version, as was expected. The testing provided valuable insights for future iterations, particularly regarding the integration of real-time baggage data feeds that would be required for operational use. The foundation has been laid for more complex testing in the Beta phase, where these advanced functionalities will be incorporated.

PISTIS: PI 1.2. TRANSFER PASSENGER MANAGEMENT.

The impact of delayed transfer passengers who transit at the Athens International Airport (AIA) is being investigated. Information on the number of passengers' transfers at AIA and on the delay of arriving flights is useful for departing flights which are waiting for transfer passengers, to calculate the impact of such delay on their schedules.

Sharing information on scheduled and actual transfer passengers between airlines, ground handlers and the airport is also able to help the optimal allocation of aircraft stands to specific flights to minimise passengers' connection times between arriving and departing aircrafts.

Ambition of use-case

The goal for this use case is to optimise allocation of aircraft stands for specific flights and minimise passengers' connection times.

Target Audience Descriptions

The actors in UC1.2 are the Greek cluster partners of PISTIS, namely AIA and Goldair that collaborate for the preparation and implementation of the Hub 1 demonstrators. The actors can be identified as data owners (AIA) and data seekers (Goldair), and they are divided as internal (the PISTIS partners) and external from other stakeholders, bodies and organisations such as other Ground Handling Companies operating in Athens Airport or the state authorities

Summary of status after Alpha testing phase

For the Transfer Passenger Management use case, the Alpha testing focused on establishing the data exchange pathway between AIA (providing data on passenger transfers and flight delays) and Goldair (receiving and utilising this data). Sample datasets with anonymised passenger transfer information and flight delay data were successfully uploaded to the PISTIS platform, processed, and retrieved by the data seeker.

The testing validated the platform's ability to handle passenger-related data with appropriate privacy measures while ensuring that the core information needed for stand allocation optimisation remained intact. Both stakeholders were able to complete the end-to-end data sharing workflow through the Alpha version of the marketplace, demonstrating the feasibility of this data exchange for future operational use.

The testing also highlighted areas for improvement in subsequent versions, particularly regarding data format standardisation and the need for near-real-time data processing capabilities that would be essential for operational decision-making about aircraft stand

allocation. These insights will inform the development of more advanced functionalities in the Beta version, where additional features like data quality assessment and usage analytics will enhance the value proposition for this use case.

PISTIS: PI 1.3. AIRCRAFT TURNAROUND PROCESS.

Information about the Target Off Block Time (TOBT) and the turnaround times for the aircraft's servicing and the underlying processes, such as catering, fuelling, cleaning etc. is very important for it to be exchanged between the involved stakeholders. Additionally, information on the aircraft turnaround process enables us to estimate whether the daily schedule will be performed as planned. It is very important for the airport to know as soon as possible any issues regarding the turnaround of the aircraft and to exchange real-time information with the handlers and airlines regarding issues or incidents in a secure manner. The sooner the airport or handler become aware of any irregularities, the shorter the reaction time for mitigation measures will be, thus minimising impact.

Ambition of use-case

The **ambition** for this use-case is to Optimise processes to achieve an increasingly efficient turnaround process and avoid delays compared to target off-block time.

Target Audience Descriptions

The actors in UC1.3 are the Greek cluster partners of PISTIS, namely AIA and Goldair that collaborate for the preparation and implementation of the Hub #1 demonstrator. The actors can be identified as data owners (AIA) and seekers (Goldair), and they are divided as internal (the PISTIS partners) and external- those from other stakeholders, bodies and organisations such as other Ground Handling Companies operating in Athens Airport or State authorities.

Summary of status after Alpha testing phase

The Alpha testing for the Aircraft Turnaround process use case successfully demonstrated the basic capabilities of the PISTIS platform for exchanging turnaround-related data between AIA and Goldair. Test datasets containing Target Off Block Time information and timestamps of various aircraft servicing processes (catering, fuelling, cleaning, etc.) were prepared and processed through the platform's data exchange workflow.

Both organisations were able to verify the platform's ability to ingest, process, and deliver this operational data securely, while maintaining its usefulness for turnaround process optimisation. The testing confirmed that the core data marketplace functionality works as designed, enabling the sharing of critical turnaround process information that would help identify potential delays early in the process.

The Alpha testing identified the need for enhanced notification capabilities and more sophisticated data processing features in future versions, which would be crucial for real-time operational response to turnaround irregularities. These requirements have been documented for implementation in the Beta version. Additionally, the testing established a baseline for understanding how this data sharing could be expanded to include other stakeholders in the aircraft turnaround process, setting the stage for more comprehensive testing in the next phase of the project.

Possibly a synergy with Extreme XP use case EX 3 covering "public safety"?

9. SHARING CORPORATE DATA

DATAMITE DA 1. **CORPORATE GROUP.**

BACKGROUND: Grupo Gimeno (GIDI) is a corporation composed of many individual companies, which cover diverse areas within citizen services sector. Although all companies belong to the same corporation, they operate totally isolated, as modular units with their own vision and objective.

Description of the pilot: The proposal consists of the development of a system/platform that allows transactions and the exchange of data between companies within a large corporation. This system offers exchanging capabilities in a framework of security and flexibility. The strategy will be a catalyst to these data centralization strategies, especially allowing them to be turned into tangible projects.

DATAMITE DA 2. SECOND CORPORATE USE CASE

BACKGROUND: Currently OTE subsidiaries have their own data repositories that mainly ingest, store, process and analyse their own produce datasets. When a subsidiary needs information that is not part of the own business, a data exchange process is initiated to request data based on business needs. Overall data exchange process is lengthy, complex and resources demanding, resulting in data and process duplication, increased delivery times, and development and operational efforts.

Description of the pilot: The scope of the pilot is to utilize the DATAMITE framework to overcome the challenges that OTE faces concerning intra-corporate data exchange between different subsidiaries of OTE Group of companies.

The DATAMITE framework will produce an efficient data exchange process, by creating a metadata repository reducing complexity of governing data, allowing users to have uniform, fine grain and rule based security controls, as well as, providing a global point to define quality rules.

https://datamite-horizon.eu/portfolio/corporate-multi-site-data-exchange/

The Remaining use-cases where synergies are not yet established.

10. FARMING

Possibly linked to the “environment cluster” or to the “sharing of corporate data group”?

DATAMITE:DA 5. FARMING USE CASE

BACKGROUND: eDWIN is a national advisory platform for farmers, operated by WODR and PSNC. Besides the data provided by the farmers, a network of meteorological ground stations provides agro-meteo data and statistical information, a Farm Management System, and ground observations of the occurrence of diseases and pests. However, integrating only the data and services from public institutions is limiting the overall offer for farmers as many of the services are provided by the commercial sector. The current roadmap for eDWIN emphasizes the need to enable secure data sharing to potentiate data monetisation, ensuring data interoperability and quality that allows for comparing and validating data, its fusion and that it is trustable so it can be used in decision support services.

Description of the pilot: The DATAMITE framework provides key mechanisms necessary for the implementation of an AgriDataMarketPlace pilot integrated with eDWIN so it can be exploited by all its customers, and by companies building services on top of it. DATAMITE governance and quality modules help to improve data management, and quality standards and validation of the acquired data. The sharing module should also provide mechanisms to smoothly publish data into the Marketplace in a secure way as well as keeping control over the data through its data sovereignty component. It should also provide accounting mechanisms to facilitate data trading and billing.

https://datamite-horizon.eu/portfolio/connecting-edwin-to-data-markets/

11. INNOVATION LIFECYCLE

ENRICHMYDATA: EN 5. INNOVATION KNOWLEDGE GRAPH FOR UNDERSTANDING INNOVATION LIFECYCLE

The business case aims to build and maintain an “Innovation Knowledge Graph” (IKG) by joining several independent streaming data sources constituting the main phases of the innovation lifecycle. The goal is to interconnect, model and understand the relevant aspects of the global innovation ecosystem from the inception of the ideas to their potential realization in business, up to the decision-making at the research policy level. In particular, the resulting knowledge graph aims to analytically support the narrative of stages which could be understood as a “journey of an innovation” throughout the innovation lifecycle.

https://enrichmydata.eu/business-cases/

12. CYBERSECURITY

EXTREME XP: EX 2: INCREASED CYBERSECURITY SITUATION AWARENESS FOR EFFICIENT THREAT MITIGATION

Context description and objectives

Through cybersecurity situation awareness, this UC aims to provide solutions to the fundamental challenges faced by the cybersecurity research community through the simulation of real-world attack scenarios.

(i2CAT & ITU)

Could they test some of the other 6 projects?

13. EXASCALE COMPUTING

GRAPH MASSIVIZER: GM 1. DATA CENTRE DIGITAL TWIN FOR SUSTAINABLE EXASCALE COMPUTING

High Performance Computing (HPC) plays a crucial role in scientific progress, but as systems approach exascale, maintenance becomes increasingly difficult. This Use Case introduces a graph-based digital twin of HPC centers that addresses anomaly handling, energy efficiency enhancement, and carbon emissions reduction.

The main objective of this Use Case is to develop and implement a sustainable framework to tackle various challenges encountered in HPC systems, including handling anomalies, enhancing energy efficiency, reducing carbon emissions, and ultimately optimizing system performance.

The logic of the Use Case involves the integration of various components, such as Graph Inceptor to output the graph representation of the telemetry data of the HPC for anomaly prediction models, as well as the Graph Scrutinizer to execute the BGOs, as well as the Optimizer and Choreographer to provide the computational resources for performing inference/graph queries.

Expected Outcomes of the Use Case

The targeted outcomes are the implementation of a sustainable model for data centers based on graphs, the development of an ontology for HPC systems, and the creation of a data center data model. The outcomes will be achieved by creating a graph-based digital twin of HPC

centers to optimize system performance and make data centers more scalable and sustainable.

This solution will enable the implementation of complex queries that make the work easier for facility managers and engineers that are not directly possible for current monitoring systems.

The solution plans to export open-source ontologies of the Marconi100 public dataset, as well as to provide technologies to be tested in data centers in production, and it will target data centers and HPC systems for exploitation. The stakeholders involved in this use case are data center owners, operational data analytics framework developers, and data center operators.

14. DATA-ENRICHMENT PIPELINE

ENRICHMYDATA: EN 2 ARTIFICIAL INTELLIGENCE -BASED WELDING ANALYTICS

Bosch proposes a business case from the automotive sector in the context of Catena-X1, an automotive alliance for secure and standardized data exchange that aims at enabling continuous data reuse and sharing for all contributors along the automotive value chain. The business case is centred around automated robotic welding where BO provides components and software solutions for welding machines that are used by factory owners, including BO and its renown car producers, for building body-in-white (BIW) of cars and other car related components. The goal of the business case is to create a data enrichment pipeline for the automation of data preparation for AI-based analytics related to automatic welding robots.

15. HEALTH

UPCAST: UP 2. BIOMEDICAL AND GENOMIC DATA SHARING

Description:

Several barriers still retain a great amount of **Biomedical and Genomic** data underexploited. Privacy-related concerns, ownership and intellectual property issues, utilisation of different analytical pipelines and different vocabularies for data description, as well as storage of data in diverse file formats, are among the challenges that have to be faced in order to facilitate integration and interpretation of data from different sources and facilitate data sharing.

Contribution to UPCAST:

NHRF is going to pilot privacy-preserving biomedical and genomic data sharing as well as integration of heterogeneous genomic data from different sources. In the frame of UPCAST, we are working on the development of semi-automated agreements to facilitate the establishment of multi-disciplinary collaborations between clinicians, biologists and data scientists, ensuring compliance with ethical and legal issues. In parallel, we are adapting our

analytical tools and workflows to agree with the FAIR principles and international standards to support harmonisation and integration of heterogeneous data.

http://www.eie.gr/nhrf/institutes/icb/research_groups/PapadodimaOlga_group_en.html

Contact person: Olga Papadodima & Konstantinos Voutetakis

UPCAST: UP 4. SMART4FIT

LESS PAIN MORE MONEY: How you will lose calories and earn money in Gym of the future

Description:

Personal fitness/health data, collected during various physical activities using wearable devices, has an extreme value not only for the data producer (trainee), but also for many service providers (fitness, healthcare, wellbeing) and product vendors (e.g. vendors of the fitness equipment, different supplements).

Smart4Fit will validate the methods for sharing personal fitness data and monetarize its value properly.

Smart4Fit (TRL9) is the system for real-time monitoring in fitness, based on wearables, which are part of a bigger IoT environment (fitness club with plenty of connected fitness device) and are used in collaborative scenarios, like group training and collaborative gamification.

Main challenge is how to determine the right value of data in a particular context, maximizing the monetary value. This leads to the situations that a trainee which is exercising more intensively, is losing more calories, but also creating the data of a higher monetary value.

Moreover, this will enable a data market where the price will be defined through the supply and demand.

The vison is to enable predicting the value of particular data types and define training plans accordingly. smart4.fit *Contact person:* nenad.stojanovic@nissatech.com

16. PUBLIC SAFETY

EXTREMEXP: EX 3. HOW TO DETECT AND VALIDATE AN ALERT IN REAL TIME ?

ExtremeXP offers an innovative solution for real-time alert detection through video stream analysis, using fire detection algorithms such as Yolo. When a fire is detected, an alert is sent to the operator with the exact location and a short video of the incident. The operator can then review the alert and interact with the system. If the alert is false, the operator fills out a brief form to explain the non-validation. If the alert is confirmed, the operator sends a request to the closest field user to go to the location and verify the situation.

How does ExtremeXP optimize field user management during an alert ?

Based on the received alerts, ExtremeXP provides the operator with a list of available users closest to the incident location, using a routing algorithm that takes into account multiple parameters (availability, location, type of vehicle, etc.). The operator can validate, modify, or reject this list. Once validated, the system contacts the users and, if they confirm their availability, sends them all the necessary information to handle the situation quickly and efficiently.

How does ExtremeXP manage variability in incident detection and response ?

ExtremeXP incorporates several variability points to maximize the effectiveness of both detection and response. These include video characteristics (data rate, frames per second, definition), the fire detection algorithm (frames per second analyzed), and the routing algorithm for field users (status, location, type of vehicle). This ensures that responses are adapted to specific conditions and that incidents are managed optimally.

Partners in the EUDATA+ projects

Partner	PISTIS	DATAMIT F	EnrichMy Data	FAME	Graph-Massivizer	UPCAST	Extreme XP
FRAUNHOFER GESELLSCHAFT ZUR FOERDERUNG DER ANGEWANDTEN FORSCHUNG E.V.	**PI**						
ATOS IT SOLUTIONS AND SERVICES IBERIA SL	PI						
ATOS SPAIN SA	PI						
IDC ITALIA SRL	PI						
IMPERIAL COLLEGE OF SCIENCE TECHNOLOGY AND MEDICINE	PI						
ATHINA-EREVNITIKO KENTRO KAINOTOMIAS STIS TECHNOLOGIES TIS PLIROFORIAS, TON EPIKOINONION KAI TIS GNOSIS	PI						
ASSENTIAN EUROPE LIMITED	PI						
SUITE5 DATA INTELLIGENCE SOLUTIONS LIMITED	PI						
FUNDACIO EURECAT	PI						
DEEP BLUE SRL	PI						
EUROPEAN DIGITAL SME ALLIANCE	PI						
UBITECH LIMITED	PI						
ARTHUR'S LEGAL BV	PI						
SPACE HELLAS (MALTA) LTD	PI						
POLITECNICO DI MILANO	PI						
INSTITUTE OF COMMUNICATION AND COMPUTER SYSTEMS	PI						
UBIMET GMBH	PI						
UBIMET GERMANY GMBH	PI						
OAG AVIATION WORLDWIDE LIMITED	PI						
ATHENS INTERNATIONAL AIRPORT S.A.	PI						
GOLDAIR EXYPIRETISEIS EDAFOUS ANONIMI ETAIREIA	PI						
DIMOS ATHINAION EPICHEIRISI MICHANOGRAFISIS	PI						

Partner	PISTIS	DATAMITF	EnrichMyData	FAME	Graph-Massivizer	UPCAST	Extreme XP
ORGANISMOS ASTIKON SYGKOINONION ATHINON AE	PI						
MONTAJES ELECTRICOS CUERVA S.L.	PI						
BAMBOO ENERGY S.L.	PI						
OMI-POLO ESPANOL SA	PI						
FUNDACION CARTIF	PI						
VIRTUAL VEHICLE RESEARCH GMBH	PI						
CARUSO GMBH	PI						
TRAFFICON - TRAFFIC CONSULTANTS GMBH	PI						
UNIVERSITÄT KLAGENFURT,					**GM**		
IDC4EU					GM		
PERACTON LTD.					GM		
SINTEF AS					GM		
UNIVERSITY OF TWENTE					GM		
METAPHACTS GMBH					GM		
VRIJE UNIVERSITEIT AMSTERDAM					GM		
CINECA CONSORTIO INTERUNIVERSITARIO					GM		
EVENT REGISTRY					GM		
ALMA MATER STUDIORUM – UNIVERSITÀ DI BOLOGNA					GM		
ROBERT BOSCH GMBH					GM		
JOZEF STEFAN INSTITUTE					GM		
INSTITUTO TECNOLOGICO DE INFORMATICA		**DA**					
UNIVERSITAT POLITECNICA DE VALENCIA		DA					
FRAUNHOFER GESELLSCHAFT ZUR FORDERUNG DER ANGEWANDTEN FORSCHUNG EV		DA					
UNIVERSITAT KOBLENZ		DA					
FIR E.V. AN DER RWTH AACHEN		DA					

Partner	PISTIS	DATAMIT F	EnrichMy Data	FAME	Graph-Massivizer	UPCAST	Extreme XP
UNIVERSITY COLLEGE CORK - NATIONAL UNIVERSITY OF IRELAND, CORK		DA					
FUNDACION TECNALIA RESEARCH & INNOVATION		DA					
ETHNIKO KENTRO EREVNAS KAI TECHNOLOGIKIS ANAPTYXIS		DA					
EREVNITIKO PANEPISTIMIAKO INSTITOUTO SYSTIMATON EPIKOINONION KAI YPOLOGISTON		DA					
INSTYTUT CHEMII BIOORGANICZNEJ POLSKIEJ AKADEMII NAUK		DA					
ORGANISMOS TILEPIKOINONION TIS ELLADOS OTE AE		DA					
STICHTING EGI		DA					
MTU AUSTRALO ALPHA LAB		DA					
INTERNATIONAL DATA SPACES EV		DA					
PRAVO I INTERNET FOUNDATION		DA					
DYAD LABS LTD		DA					
DIACHEIRISTIS ELLINIKOU DIKTYOU DIANOMIS ELEKTRIKIS ENERGEIAS AE		DA					
IDFS SPOLKA ZOO		DA					
GIMENO DIGITAL TECHNOLOGIES SOCIEDAD LIMITADA		DA					
ECLIPSE FOUNDATION EUROPE GMBH		DA					
ZENTRUM FUR SOZIALE INNOVATION GMBH		DA					
GLOBAZ, S.A.		DA					
WIELKOPOLSKI OSRODEK DORADZTWA ROLNICZEGO W POZNANIU		DA					
DIN DEUTSCHES INSTITUT FUER NORMUNG EV		DA					
CINECA CONSORZIO INTERUNIVERSITARIO		DA					
1001 LAKES OY		DA					
E-REDES - DISTRIBUICAO DE ELETRICIDADE SA		DA					
SINTEF AS			EN c				
PHILIPS			EN				

Partner	PISTIS	DATAMITE	EnrichMyData	FAME	Graph-Massivizer	UPCAST	Extreme XP
BOSCH			EN				
SPEND NETWORK			EN				
JOT INTERNET MEDIA ESPANA SL			EN				
CS GROUP-ROMANIA S.A.			EN				
EXPERT.AI S.P.A.			EN				
ONTOTEXT			EN				
UNIVERSITY OFMILANO BICOCCA			EN				
JOZEF STEFAN INSTITUTE			EN				
UNIVERSITY OF COPENHAGEN			EN				
GATE INSTITUTE			EN				
BGRIMM			EN				
GFT ITALIA SRL				**FA**			
FUJITSU TECHNOLOGY SOLUTIONS				FA			
ATOS IT SOLUTIONS AND SERVICES IBERIA SL				FA			
IBM ISRAEL - SCIENCE AND TECHNOLOGY LTD				FA			
ENGINEERING - INGEGNERIA INFORMATICA SPA				FA			
MC SHARED SERVICES SA				FA			
UNIVERSO IME SA				FA			
DIMOS ATHINAION EPICHEIRISI MICHANOGRAFISIS				FA			
JRC CAPITAL MANAGEMENT CONSULTANCY & RESEARCH GMBH				FA			
MOTOR OIL (HELLAS) DIILISTIRIA KORINTHOU A.E.				FA			
LEANXCALE SL				FA			
UBITECH LIMITED				FA			
INQBIT INNOVATIONS SRL				FA			
KM CUBE ANONYMI ETAIREIA PAROCHIS EPENDYTIKON YPIRESION				FA			

Partner	PISTIS	DATAMITE	EnrichMyData	FAME	Graph-Massivizer	UPCAST	Extreme XP
UNPARALLEL INNOVATION LDA				FA			
JOT INTERNET MEDIA ESPANA SL				FA			
INNOV-ACTS LIMITED				FA			
TRAILBLU YAZILIM ANONIM SIRKETI				FA			
NOVOMATIX IDIOTIKI KEFALAIOUCHIKI ETAIREIA				FA			
UNIVERSITY OF GALWAY				FA			
NORSK REGNESENTRAL				FA			
UNINOVA-INSTITUTO DE DESENVOLVIMENTO DE NOVAS TECNOLOGIAS-ASSOCIACAO				FA			
UNIVERSITY OF PIRAEUS RESEARCH CENTER				FA			
INSTITUT JOZEF STEFAN				FA			
EKONOMICKA UNIVERZITA V BRATISLAVE				FA			
UNIVERSITETET I AGDER				FA			
BANKING & PAYMENTS FEDERATION IRELAND COMPANY LIMITED BY GUARANTEE				FA			
ECO VERBAND DER INTERNETWIRTSCHAFTEV				FA			
INTERNATIONAL DATA SPACES EV				FA			
INNEUROPE INITIATIVE S.L.				FA			
ARTHUR'S LEGAL BV				FA			
CREDIT AGRICOLE GROUP SOLUTIONS SOCIETA' CONSORTILE PER AZIONI				FA			
ELIS INNOVATION HUB SRL				FA			
ETONEC GMBH				FA			
IDC ITALIA SRL						**UP**	
UNIVERSITY COLLEGE DUBLIN, NATIONAL UNIVERSITY OF IRELAND, DUBLIN						UP	
ICT ABOVO TECHNOLOGIES PLIROFORIKISKAI EPIKINONION IDIOTIKI KEFALEOUCHIKI ETERIA						UP	
MAGGIOLI SPA						UP	

Partner	PISTIS	DATAMITE	EnrichMyData	FAME	Graph-Massivizer	UPCAST	Extreme XP
ANAPTYXIAKI MEIZONOS ASTIKIS THESSALONIKIS AE - ANAPTYXIAKOS ORGANISMOS TOPIKIS AUTODIOIKISIS						UP	
OPEN KNOWLEDGE FOUNDATION GREECE						UP	
KATHOLIEKE UNIVERSITEIT LEUVEN						UP	
SINTEF AS						UP	
JOT INTERNET MEDIA ESPANA SL						UP	
NOKIA NETWORKS FRANCE						UP	
LSTECH ESPANA SL						UP	
PRIVREDNO DRUSTVO ZA PRUZANJE USLUGA ISTRAZIVANJE I RAZVOJ NISSATECH INNOVATION CENTRE DOO						UP	
ETHNIKO IDRYMA EREVNON						UP	
STAMATIS ANESTIS						UP	
NOKIA SOLUTIONS AND NETWORKS OY						UP	
DAWEX SYSTEMS						UP	
UNIVERSITY OF SOUTHAMPTON						UP	
ATHINA-EREVNITIKO KENTRO KAINOTOMIAS STIS TECHNOLOGIES TIS PLIROFORIAS, TON EPIKOINONION KAI TIS GNOSIS							EX
ACTIVEEON							EX
AIRBUS DS SLC							EX
BITSPARKLES							EX
CS GROUP - France							EX
UNIVERZITA KARLOVA							EX
DEUTSCHES FORSCHUNGSZENTRUM FUR KUNSTLICHE INTELLIGENZ GMBH							EX
FUNDACIO PRIVADA I2CAT, INTERNET I INNOVACIO DIGITAL A CATALUNYA							EX

Partner	PISTIS	DATAMITE	EnrichMyData	FAME	Graph-Massivizer	UPCAST	Extreme XP
EREVNITIKO PANEPISTIMIAKO INSTITOUTO SYSTIMATON EPIKOINONION KAI YPOLOGISTON							EX
IDEKO S COOP							EX
INTERACTIVE 4D							EX
INTRACOM SINGLE MEMBER SA TELECOM SOLUTIONS							EX
MOBY X SOFTWARE LIMITED							EX
SINTEF AS							EX
TECHNISCHE UNIVERSITEIT DELFT							EX
UNIVERZA V LJUBLJANI							EX
UNIVERSITAT POLITECNICA DE CATALUNYA							EX
ITHINKUPC SL							EX
STICHTING VU							EX
BOURNEMOUTH UNIVERSITY							EX

Description from CORDIS of the Seven EUDATA+ Projects

1 EXTREME XP

A new framework for experimentation-driven analytics

Extreme data characteristics represent a challenge for advanced data-driven analytics and decision-making in critical domains such as crisis management, predictive maintenance, mobility, public safety and cyber-security. Data-driven insights must be timely, accurate, precise, fit-for-purpose and reliable, considering and learning from user intents and preferences. The EU-funded ExtremeXP project will create a next-generation decision support framework that integrates novel research from big data management, machine learning, visual analytics, explainable AI, decentralised trust, and knowledge engineering. The framework will aim at optimising the properties of complex analytics processes (e.g. accuracy, time-to-answer, specificity, recall, precision, resource consumption) by associating different user profiles with computation variants, promoting a human-centered, experimentation-based approach to AI and complex analytics. The project will perform five pilot demonstrations.

Objective

Extreme data characteristics (volume, speed, heterogeneity, distribution, diverse quality, etc.) challenge the state-of-the-art data-driven analytics and decision-making approaches in many critical domains such as crisis management, predictive maintenance, mobility, public safety, and cyber-security. At the same time, data-driven insights need to be extremely timely, accurate, precise, fit-for-purpose, and trustworthy, so that they can be useful. ExtremeXP will handle the complexity of matching extreme needs with complex analytics processes (i.e. processes that involve and combine ML, data analysis, simulation and visualization components) by placing the end user at the centre of complex analytics processes and relying on user intents and running experiments (i.e. trial and error) to prune the vast solution space of possible analytics workflows and configurations i.e. “variants”. Its main goal is to create a next generation decision support system that integrates novel research results from the domains of data integration, machine learning, visual analytics, explainable AI, decentralised trust, knowledge engineering, and model-driven engineering into a common framework. The overarching idea of the framework is to optimise the properties of a complex analytics process that the end user cares about (e.g. accuracy, time-to-answer, specificity, recall, precision, resource consumption) by associating user profiles to computation variants. The framework is envisioned as modular and extensible, orchestrating different services around an Experimentation Engine: Analysis-aware Data Integration, Extreme Data & Knowledge Management, User-driven AutoML, Transparent & Interactive Decision Making, and User-driven Optimization of Complex Analytics. The framework will be validated in five pilot demonstrators.

2 UPCAST

Components for safe, fair, interoperable data

Data exchange, monetisation and trading should be safe, fair and interoperable. This is essential in the digital data market. The EU-funded UPCAST will provide a set of universal, trustworthy, transparent and user-friendly data market plugins. These will enable actors in the European data spaces to design and deploy data exchange and trading operations. To ensure the automation of data sharing and processing agreements between businesses, public administrations and citizens, the project will pave the way for the automatic negotiation of agreement terms and improve data-asset discovery. UPCAST will also cover commercial and administrative confidentiality requirements, and provide low environmental footprint. Four real-world pilots across Europe will operationalise a set of working platform plugins for data sharing, monetisation and trading.

Objective

UPCAST provides a set of universal, trustworthy, transparent and user-friendly data market plugins for the automation of data sharing and processing agreements between businesses, public administrations and citizens. Our plugins will enable actors in the common European data spaces to design and deploy data exchange and trading operations guaranteeing (i) automatic negotiation of agreement terms, (ii) dynamic fair pricing, (ii) improved data-asset discovery, (iii) privacy, commercial and administrative confidentiality requirements, (iv) low environmental footprint, as well as ensuring compliance with (v) relevant legislation and (vi) ethical and responsibility guidelines. UPCAST will support the deployment of Common European data spaces by consolidating mature research in the areas of data management, privacy, monetisation, exchange and automated negotiation, considering efficiency for the environment as well as compliance with EU and national initiatives, AI regulations and ethical procedures. Four real-world pilots across Europe will operationalise a set of working platform plugins for data sharing, monetisation and trading, deployable across a variety of different data marketplaces and platforms, ensuring digital autonomy of data providers, brokers, users and data subjects, and enabling interoperability within European data spaces. UPCAST aims at engaging SMEs, administrations and citizens by providing a transferability framework, best practices and training to endow users in order to deploy the new technologies and maximise impact of the project.

3 FAME

New tech tools for the data marketplace

Data is all around us, in every sector and industry. The management of data, data technologies, the data economy and digital finance is more important than ever. The EU-funded FAME project will develop a secure federated data marketplace for embedded finance (EmFi). The aim is to demonstrate the full potential of the data economy. FAME will develop a federated cloud environment with multiple providers of EmFi data assets, including datasets,

AI/ML models, and more. By connecting with more than a dozen data marketplaces, the project will roll out seven pilots. To unlock the potential of the data economy, FAME will also establish a learning centre for tech and non-tech users, as this is a key prerequisite for unlocking the potential of the data economy.

Objective

FAME is a joint effort of world-class experts in data management, data technologies, the data economy, and digital finance to develop, deploy and launch to the global market a unique, trustworthy, energy-efficient, and secure federated data marketplace for Embedded Finance (EmFi). The FAME marketplace will alleviate the proclaimed limitations of centralized cloud marketplaces towards demonstrating the full potential of the data economy. In this direction, the project will enhance a state of the art data marketplace infrastructure (i.e. H2020 i3-Market marketplace) with novel functionalities in three complementary directions namely:
• Secure, interoperable, and regulatory compliant data exchange across multiple federated cloud-based data providers in-line with emerging European initiatives like GAIA-X.
• Decentralized, programmable, data assets trading and pricing leveraging blockchain tokenization techniques (including support for accruing data assets value in NFTs).
• Integration of trusted and Energy Efficient (EE) analytics based on novel technologies such as Quantitative Explainable AI, Situation Aware Explainability (SAX), incremental EE analytics, and edge analytics.
FAME will become operational in a federated cloud environment with multiple providers of EmFi data assets, including datasets, AI/ML models, and more. It will become interconnected with more than 12 data marketplaces that are operated by the project partners, as well as with other data infrastructures that will support the implementation of 7 pilots. Through this process, the catalog of the FAME marketplace will be populated with a critical mass of 1000+ data assets.
Furthermore, FAME will establish a Learning Center (LC) for tech and non-tech users, as this is a key prerequisite for unlocking the potential of the data economy. FAME will build a vibrant community of EmFi stakeholders around the FAME platform, which will serve as a catalyst for the sustainability of the project's results.

4 PISTIS

A sustainable solution for data trading and monetising

With so much data available, it is not easy to manage it effectively. There is a massive need for solutions to facilitate the collection, sharing, storage, processing, trading and reuse of data. The EU-funded PISTIS project will bring forward a reference federated data sharing/trading and monetisation platform for secure, trusted and controlled exchange and usage of proprietary data assets and data-driven intelligence. It will advance techniques and technologies, such as federated data discovery and sharing, distributed ledger technologies, data non-fungible tokens, AI-driven data quality assessment and monetisation, to build trust among stakeholders and assuage their safety concerns. PISTIS will consider the different data

supply and demand perspectives and train stakeholders to evaluate the sustainability of their data sharing strategies.

Objective

PISTIS brings forward a reference federated data sharing/trading and monetisation platform for secure, trusted and controlled exchange and usage of proprietary data assets and data-driven intelligence. PISTIS will advance the available techniques and technologies, such as federated data discovery and sharing, DLTs, data non-fungible tokens (NFTs), AI-driven data quality assessment and monetisation, to build trust among stakeholders and assuage their concerns. Such stakeholders will formulate a distributed network of existing and new data spaces with built-in governance brought by PISTIS to eliminate silos while accruing the actual data value and multiplying it through derivative assets in a fair and transparent manner. Taking into consideration the data supply and demand perspectives, PISTIS will establish the methodological and technical foundations across different axes:
-PISTIS Federated Data Management, Interoperability & Governance that aims at collecting, curating, securing and fully controlling the data made available through each organisations data space.
-PISTIS Federated, Secure Data Sharing that concerns the effective management and on-chain storage of (multi-party) data contracts in an inherently human-understandable manner, as well as the secure peer-to-peer data transfer and usage monitoring mechanisms for appropriately retrieving, provisioning, self-serving on-demand and tracking the appropriate data slices according to the relevant contract provisions.
-PISTIS Data Valuation and Monetisation to systematically articulate and recommend an appropriate target value, indicatively taking into consideration the cost approach, the income approach, and the market approach.
-PISTIS Data Sharing Skills Cultivation that provides the training material to educate stakeholders into how they can assess their data sharing maturity, proceed with data, deploy/operate the PISTIS technologies and implement an effective and sustainable data sharing strategy.

5 GRAPH-MASSIVIZER

Innovative graph processing addressing societal challenges

European researchers are making important strides in graph processing and serverless computing. The EU-funded Graph-Massivizer project will develop a high-performance, scalable and sustainable platform for information processing and reasoning relying on the massive graph representation of extreme data. It will deliver five open-source software tools and FAIR graph data sets. The tools will focus on holistic usability (extreme data ingestion and massive graph creation), automated intelligence (analytics and reasoning), performance modelling and environmental sustainability trade-offs. The project will validate the innovation on four complementary use cases covering the economy, society and the environment. It expects to show a two-fold improvement in data centre energy efficiency and over 25 % lower greenhouse gas emissions for basic graph operations.

Objective

Graph-Massivizer researches and develops a high-performance, scalable, and sustainable platform for information processing and reasoning based on the massive graph representation of extreme data. It delivers a toolkit of five open-source software tools and FAIR graph datasets covering the sustainable lifecycle of processing extreme data as massive graphs. The tools focus on holistic usability (from extreme data ingestion and massive graph creation), automated intelligence (through analytics and reasoning), performance modelling, and environmental sustainability tradeoffs, supported by credible data-driven evidence across the computing continuum. The automated operation based on the emerging serverless computing paradigm supports experienced and novice stakeholders from a broad group of large and small organisations to capitalise on extreme data through massive graph programming and processing.

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 AI for the sustainable automotive industry, and data centre digital twin for exascale computing. Graph Massivizer promises 70% more efficient analytics than AliGraph, and 30% improved energy awareness for ETL storage operations than Amazon Redshift. Furthermore, it aims to demonstrate a possible two-fold improvement in data centre energy efficiency and over 25% lower GHG emissions for basic graph operations.

Graph-Massivizer gathers an interdisciplinary group of twelve partners from eight countries, covering four academic universities, two applied research centres, one HPC centre, two SMEs and two large enterprises. It leverages the world-leading roles of European researchers in graph processing and serverless computing and uses leadership-class European infrastructure in the computing continuum.

6 ENRICHMYDATA

The importance of building rich, high-quality and valuable data sets

Big Data is essential to powering the evolution of AI. In this context, the EU-funded enRichMyData project is developing a new paradigm for building rich, high-quality and valuable data sets to feed data analytics and Al applications. The paradigm supports the scalable execution of data enrichment pipelines, including key operations such as discovery, understanding, selection, cleaning, transformation and integration of Big Data from a variety of sources. The paradigm will be easily accessible to a wide range of organisations that encounter difficulties in delivering suitable data to feed their data analytics solutions. The project will validate its solutions through a strong selection of complementary business cases in various domains, including digital marketing, public procurement, innovation ecosystems, Industry 4.0 predictive maintenance of medical devices, and mineral processing.

Objective

enRichMyData provides a novel paradigm for building rich, high-quality and valuable datasets to feed Big Data Aalytics and AI applications. The paradigm consists in facilitating the specification and scalable execution of data enrichment pipelines, with a focus on supporting various data enrichment operations such as discovery, understanding, selection, cleaning, transformation, integration of Big Data from a variety of sources. enRichMyData makes this paradigm easily accessible to a wide set of large and small organizations that encounter difficulties in the process of delivering suitable data to feed their data analytics solutions due to the lack of suitable tools and expertise to support cost- effective and energy-efficient management of data enrichment pipelines. enRichMyData delivers a software toolbox consisting of tools and infrastructure services for setting-up, deploying, executing and managing data enrichment pipelines. The main exploitation targets the operation and monetization of the toolbox primarily in European markets. enRichMyData validates its plan through a strong selection of complementary business cases offered by commercially- focused organizations targeting development of novel products in a wide range of domains, including digital marketing, Industry 4.0 medical devices maintenance, public procurement, innovation ecosystems, and mineral processing. The balanced consortium consists of 13 partners from 11 countries. With a mixture of R&D/technology providers and business case providers, enRichMyData gathers five large companies, three SMEs, two research institutes, and three universities.

7 DATAMITE

Open-source framework for reliable data

European companies can benefit from technologies and solutions for data monetisation, interoperability, trading and exchange. The EU-funded DATAMITE project will develop a modular, open-source and multi-domain framework to improve such technologies and solutions in the form of software modules, training and business materials. The framework will provide users with the tools they need to improve the quality management of their data, adhering to FAIR principles and allowing upskilling on technical and business aspects that make data more reliable. At the external level, it will provide new sources of revenue and opportunities for interaction with other stakeholders. The project envisages an architecture that enables digital innovation hubs' sandboxing, becoming a potential instructor in their onboarding of SMEs into the data economy.

Objective

DATAMITE empowers European companies by delivering a modular, open-source and multi-domain Framework to improve DATA Monetizing, Interoperability, Trading and Exchange, in the form of software modules, training, and business materials. DATAMITE unleashes the monetization potential at two levels. At internal level, users will have tools to improve quality management of their data, the adherence to FAIR principles, and will be able to upskill on technical and business aspects thanks to the multiple open-source

training materials the project will generate. Therefore, data will become trustable and more reliable also in other paradigms like AI.
At external level, keeping users in control of their data will provide new sources of revenue and interaction with other stakeholders. The architecture envisioned for DATAMITE enables DIHs sandboxing, becoming a potential instructor on their onboarding of SMEs and low-tech SMEs into the data economy. Together, DATAMITE's solutions will function as a catalyst to boost data monetization in the European productive fabric.
DATAMITE will validate the results in 3 different use cases with a total of 6 pilots, demonstrating that the Framework is interoperable and usable in different domains and user needs, as: 1) Intra-corporate, multi-domain data exchange; 2) Data trading among Data Spaces; 3) Integration with other initiatives as Data Markets, EU AI-on-demand platform, or DIHs. Sectors covered by the pilots are: agriculture, energy, industrial and manufacturing, and climate.
To achieve this, the project relies on a consortium of 27 partners from 13 countries, bringing together key actors of the Data Value Chain: Data Spaces technical and business stakeholders, multiple key communities (IDSA, Gaia-X, EUHUBS4DATA, AI4EUROPE, EOSC), key experts in Legal and SSH aspects to guarantee legal and societal compliance, and facilitators on open-source community building and standardisation activities to accelerate the transfer to the market.