Strategy for Pricing Data Assets in the Era of AI

Jana Peliova — University of Economics in Bratislava
Samrat Gupta — University of Agder, Norway
IIM Ahmedabad, India

This Project has received funding from the European Union´s Research and Innovation programme under grant Agreement no 101092639

Data assets as factor of production

Traditional Perception of Factors of Production

Land

Natural resources used in production (e.g., minerals, water, land itself). Ricardo, D. (1817)

Labor

Human effort, both physical and intellectual, in production. Smith, A. (1776).

Capital

Man-made tools, machines, and infrastructure aiding production. Solow, R. M. (1956).

Data: The Emerging Factor of Production

Creates value through analysis, prediction, and innovation.

Fuels AI, digital economies, and decision-making. Brynjolfsson, E., & Hitt, L. M. (2000).

This project has received funding from the European Union's Horizon research and innovation programme under Grant Agreement no 101092639

Why is it hard to give pricetag to data assets?

1. Intangibility

Data assets lack physical form, making their perceived value highly subjective.

2. Non-Rivalry

Multiple users can consume the asset simultaneously without reduction, complicating marginal cost pricing.

3. Dynamic Nature

Rapid innovation and obsolescence impact the longevity of value.

4. Valuation Challenges

Determining the economic value of a digital asset depends on its utility, demand, and alternative uses.

This project has received funding from the European Union's Horizon research and innovation programme under Grant Agreement no 101092639

Examples of pricing strategies

Pricing Strategy
Competitive Benchmarking
Cost-Plus Pricing
Value-Based Pricing
Usage-Based Pricing
Dynamic Pricing
Auction-Based Pricing
Customer Segmentation

Value-Based Pricing

Pricing reflects value to the customer.

Used by: SAP, Nielsen.

Advertising Monetization

Free data services monetized through advertising.

Used by: Meta, Google, TikTok.

This project has received funding from the European Union's Horizon research and innovation programme under Grant Agreement no 101092639

Business Models Used by Data Producers

1. Freemium Model

Free access to basic features, advanced features are paid.

Used by: Google (Analytics), Meta (Facebook Ads).

2. Subscription Model

Monthly or yearly service fees.

Used by: Microsoft Azure, Amazon AWS, Bloomberg Terminal.

3. Usage-Based Pricing

Payment based on data volume or API calls.

Used by: AWS, Google Cloud, X (Twitter) API.

This project has received funding from the European Union's Horizon research and innovation programme under Grant Agreement no 101092639

When it gets wrong

Disney+ Hotstar in India (2023)

Problem: Underestimating the value of premium content

What Happened: Disney+ Hotstar lost streaming rights for IPL (Indian Premier League) cricket matches but did not adjust its pricing strategy. It underestimated how much the absence of premium content would affect subscriptions.

Result: A significant drop in subscribers and revenue in the Indian market.

Lesson: Pricing strategies must reflect the perceived value of the content or product being offered.

This project has received funding from the European Union's Horizon research and innovation programme under Grant Agreement no 101092639

When it gets wrong

X(formerly known as Twitter) API (2023)

Problem: Sudden and steep pricing model changes

What Happened: Twitter drastically increased the price of its API access, making it unaffordable for smaller developers and organizations. The pricing moved from a free or low-cost tier to plans starting at $100/month, with enterprise plans costing tens of thousands of dollars.

Result: Many users abandoned the platform, leading to reputational damage and reduced ecosystem engagement.

Lesson: Transparent and incremental changes in pricing are critical for maintaining trust and adoption.

This project has received funding from the European Union's Horizon research and innovation programme under Grant Agreement no 101092639

AI and LLMs Supporting Pricing Strategies

Benefits for buyers:

- Market Comparison: AI helps buyers understand if they are getting a fair deal by comparing the product with similar offerings in the market.
- Personalized Recommendations: AI can provide recommendations based on user preferences and past behavior, enhancing the buying experience.

Benefits for sellers:

- Market Analysis: AI can assist sellers in understanding market trends and the competitive landscape.
- Dynamic Pricing: AI can implement dynamic pricing strategies that adjust based on demand, user engagement, and market trends.

This project has received funding from the European Union's Horizon research and innovation programme under Grant Agreement no 101092639

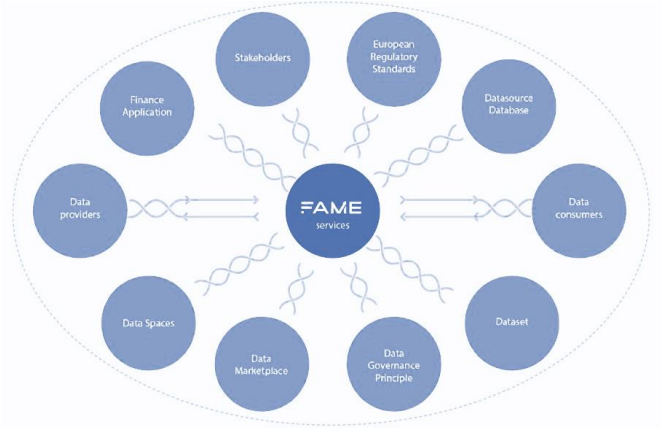
Innovation, Approach, Value

Business Model

- **Multi-Sided Powerhouse**: Connects data providers and consumers, fostering a win-win for both.
- **Financial Innovation Ecosystem**: Opens doors to scientific research and non-financial players, sparking fresh solutions.
- **Secure & Transparent by Design**: Blockchain ensures secure and traceable data transactions.

Governance Model

- **Empowering Participants**: FAME adopts a decentralized approach, giving data providers and consumers control (data sovereignty) and fostering trust through EU-compliant practices.
- **Secure Collaboration**: A shared framework of agreed-upon principles ensures a secure environment for everyone involved.
- **Operations**: The governance module streamlines member onboarding and secure data transactions.



Data Assets

- **Classical Data and Datasets**: this category includes raw and processed data applicable to various financial situations.
- Value-Added Assets: algorithms and pre-made technology components like code snippets and software modules ready to be used in financial applications.
- **Running Services**: operational services that can be directly integrated into user environments. These include, but are not limited to, Energy Efficient (EE) Analytics, Artificial Intelligence (AI), Machine Learning (ML) models, Situation Aware Explainability (SAX) and Explainable AI (XAI) models.
- **Digital Content**: educational and training materials such as courseware and tutorials to help users learn and develop their skills on the platform.

Market Focus

- **Beyond generic data marketplaces**: Designed specifically for the growing field of Data-Driven Finance Applications.
- **Focus on a niche market**: Caters to the specific needs of the financial sector, from banks, insurance, to startups.
- **Unified platform for diverse data**: Enables easy discovery and use of various financial data assets.
- **Empowering innovation**: Provides the tools for financial institutions to develop new services efficiently.

Pricing Advisory Tool

- **Data's Unique Value**: data as a valuable and reusable resource.
- **Pricing Made Easy**: pricing advisory tool helps users determine fair prices for their data assets.
- **Similar Data, Similar Price**: The tool analyzes properties like accuracy, completeness, and uniqueness to recommend a price based on similar data assets already sold.

This project has received funding from the European Union's Horizon research and innovation programme under Grant Agreement no 101092639

Innovation, Approach, Value

Pricing Advisory Tool

- **Data's Unique Value**: data as a valuable and reusable resource.
- **Pricing Made Easy**: pricing advisory tool helps users determine fair prices for their data assets.
- **Similar Data, Similar Price**: The tool analyzes properties like accuracy, completeness, and uniqueness to recommend a price based on similar data assets already sold.

This project has received funding from the European Union's Horizon research and innovation programme under Grant Agreement no 101092639

Key principle for pricing data in FAME

Data possesses a concept of pairwise distance, which is not typically found in other goods (Agarwal et al., 2019).

Agarwal, A., Dahleh, M., & Sarkar, T. (2019, June). A marketplace for data: An algorithmic solution. In *Proceedings of the 2019 ACM Conference on Economics and Computation* (pp. 701-726).

This project has received funding from the European Union's Horizon research and innovation programme under Grant Agreement no 101092639

Extrinsic and non-extrinsic Characteristics of Data Assets

Extrinsic

- **Logical** Characteristics of Data Assets
- **Ordinal** Characteristics of Data Assets
- **Numeric** Characteristics of Data Assets
- **Categorical** Characteristics of Data Assets

Non-Extrinsic

- **Data Distribution** within Data Assets
- **Textual** Data Associated with Data Assets
- **Intrinsic Characteristics** of Data Assets

This project has received funding from the European Union's Horizon research and innovation programme under Grant Agreement no 101092639

Parameters related to extrinsic characteristics of data assets

- Quality of data (Accuracy, validity, completeness)
- Costs (data collection, data storage)
- Estimated cost for customer
- Uniqueness
- Environmental Sustainability

This project has received funding from the European Union's Horizon research and innovation programme under Grant Agreement no 101092639

Questions for extrinsic characteristics of data assets

- Quality of data (Accuracy, validity, completeness)
  - How credible is the source of the asset? (Likert)
  - Have these data been validated against independent sources or standards? (Y/N)
  - How many information points does the asset contain? (Numeric)
- Costs (data collection, data storage)
- Estimated cost for customer
- Uniqueness
- Environmental Sustainability

This project has received funding from the European Union's Horizon research and innovation programme under Grant Agreement no 101092639

Questions for extrinsic characteristics of data assets

- Quality of data (Accuracy, validity, completeness)
- Costs (data collection, data storage)
  - Is it necessary to use additional software to manage/read/update data? (Y/N)
  - How much resources were required to assemble and prepare the asset (in MD)? (Numeric)
- Estimated cost for customer
- Uniqueness
- Environmental Sustainability

This project has received funding from the European Union's Horizon research and innovation programme under Grant Agreement no 101092639

Questions for extrinsic characteristics of data assets

- Quality of data (Accuracy, validity, completeness)
- Costs (data collection, data storage)
- Estimated cost for customer
  - How suitable are data in the asset for a wide range of analyses? (Likert)
  - Will the information in the asset regularly updated? (Y/N)
- Uniqueness
- Environmental Sustainability

This project has received funding from the European Union's Horizon research and innovation programme under Grant Agreement no 101092639

Questions for extrinsic characteristics of data assets

- Quality of data (Accuracy, validity, completeness)
- Costs (data collection, data storage)
- Estimated cost for customer
- Uniqueness
  - Are there copyrights related to the data asset? (Y/N)
  - How unique Is the dataset ? (Likert)
- Environmental Sustainability

This project has received funding from the European Union's Horizon research and innovation programme under Grant Agreement no 101092639

Questions for extrinsic characteristics of data assets

- Quality of data (Accuracy, validity, completeness)
- Costs (data collection, data storage)
- Estimated cost for customer
- Uniqueness
- Environmental Sustainability
  - To what extent was renewable energy used in the process of asset creation? (Likert)

This project has received funding from the European Union's Horizon research and innovation programme under Grant Agreement no 101092639

Similarity measurement between data assets

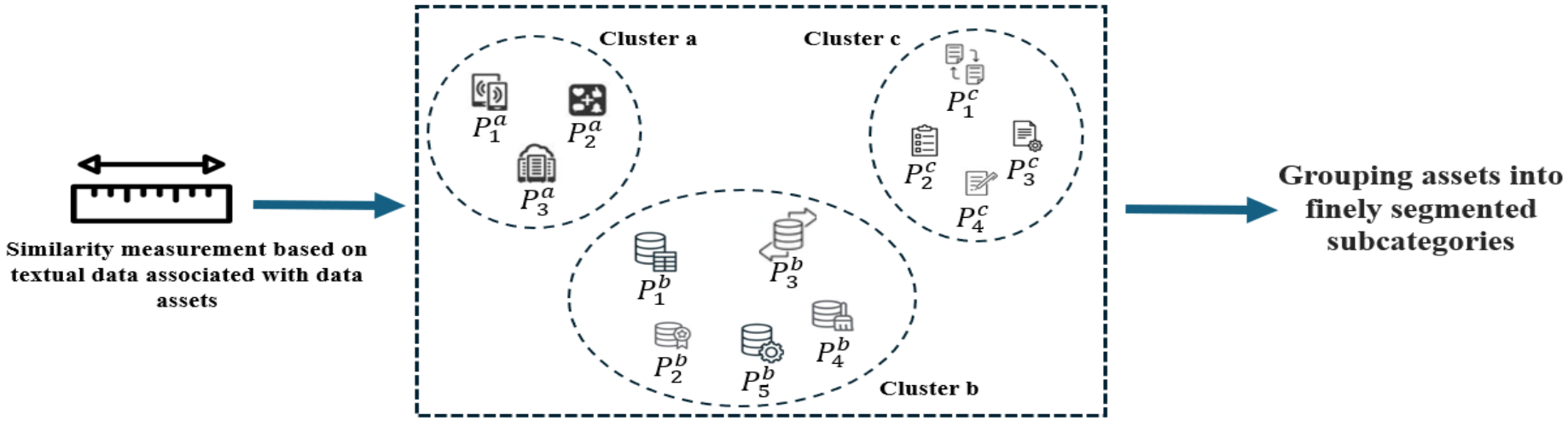
A similarity metric is a function, → [0, 1], that satisfies four properties which are as follows:

- Limited Range: $0 \leq SM \leq 1$;
- Reflexive: SM(X,Y) = 1 *if and only if* X=Y;
- Symmetry: SM(X,Y) = SM(Y,X);
- Triangle Inequality: $dSM(X,Y) + dSM(Y,Z) \geq dSM(X,Z)$ *where* $dSM(X,Y) = 1\text{-}SM(X,Y)$

This project has received funding from the European Union's Horizon research and innovation programme under Grant Agreement no 101092639

Two step approach for similarity assessment in FAME

The first phase uses Natural Language Processing (NLP) and SBERT techniques to extract textual data from asset titles and descriptions provided by users within the Asset Offering interface and subsequently preproces it for clustering.



This project has received funding from the European Union's Horizon research and innovation programme under Grant Agreement no 101092639

Two step approach for similarity assessment in FAME

The second phase initiates a similarity analysis leveraging advanced algorithmic comparisons within each identified subgroup.

Pseudo Code: Asset Matching in a Data Marketplace

Input:

$\{DA_1, DA_2, \ldots, DA_n\} \in DA$: A set of data assets hosted by the data marketplace

DA_f: A focal data asset for which price is to be recommended

Output:

$\{DA_{f1}, DA_{f2}, \ldots, DA_{f10}\}$: A set of 10 most similar data assets to the focal data asset DA_f

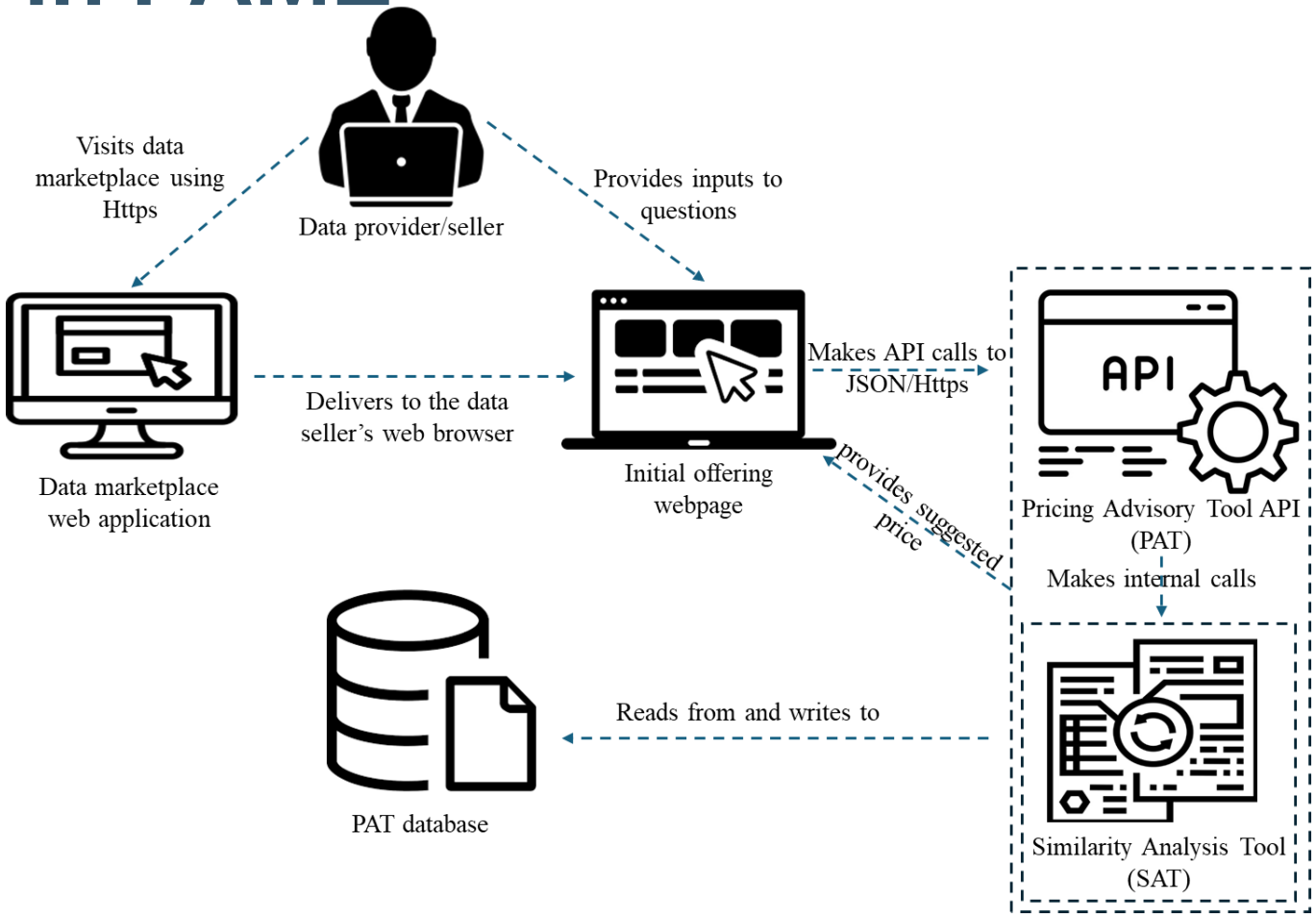
Begin:

(1) For each data asset DA_i , construct its EMST
(2) For each data asset DA_i , compute its class overlap EMST-O according to Definition 1
(3) For each data asset DA_i , merge its EMST-O value with the set of extrinsic characteristics according to Definition 2
(4) Compute the similarity between each pair of data assets using Definition 3
(5) Return and display the top 10 similar data assets in DA in terms of their similarity with focal data asset DA_f

End

This project has received funding from the European Union's Horizon research and innovation programme under Grant Agreement no 101092639

Component Level Architecture of pricing advisory in FAME



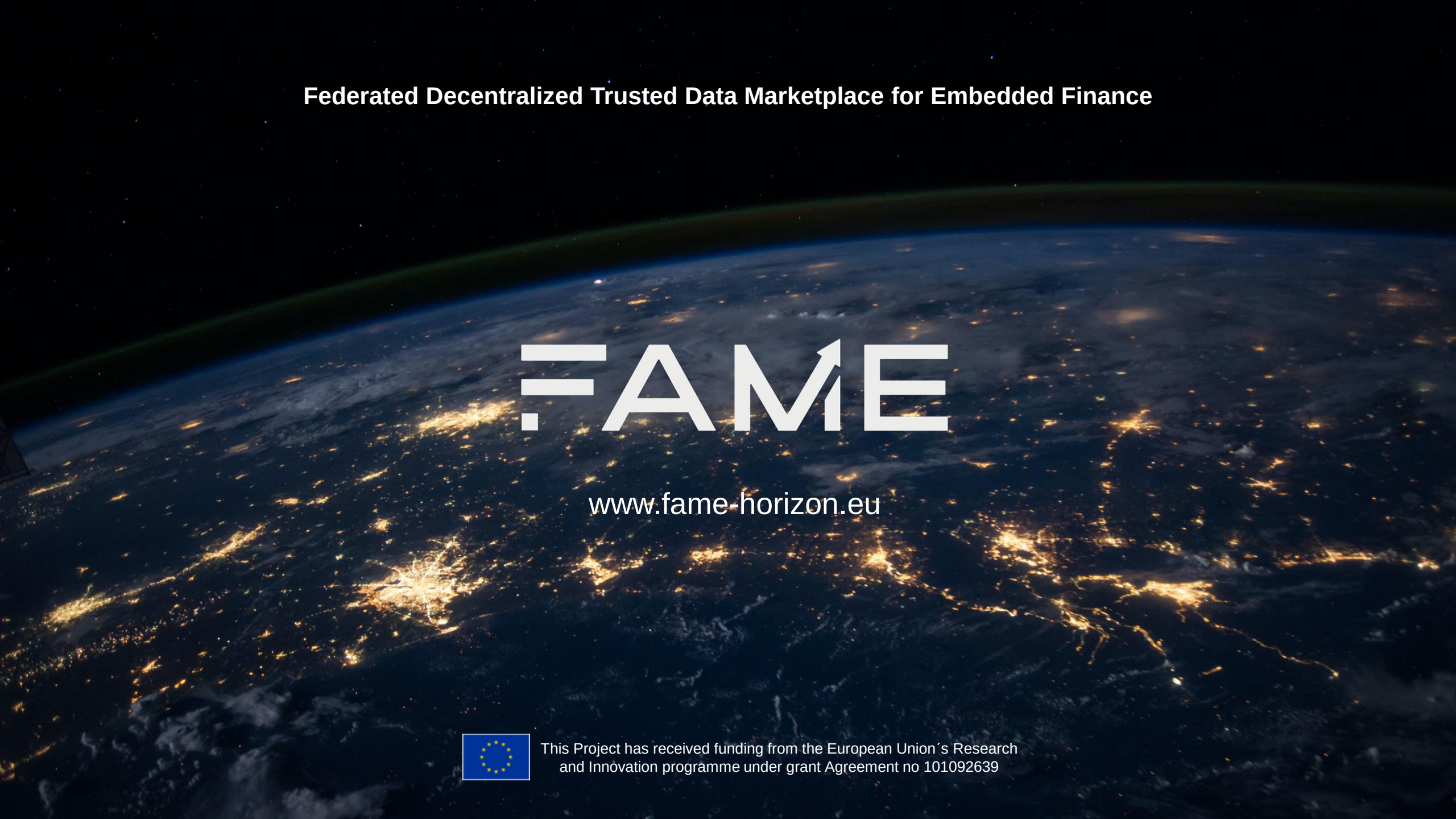
This project has received funding from the European Union's Horizon research and innovation programme under Grant Agreement no 101092639

"Data is the new oil of the internet and new currency of the digital world."

Meglena Kuneva, EU Commissioner

This project has received funding from the European Union's Horizon research and innovation programme under Grant Agreement no 101092639

Federated Decentralized Trusted Data Marketplace for Embedded Finance



www.fame-horizon.eu

This Project has received funding from the European Union´s Research and Innovation programme under grant Agreement no 101092639