Informe de progreso

MLEDGE - Aprendizaje automático en la nube y en el borde

(Cloud and Edge Machine Learning)

Diciembre de 2023

Información sobre el entregable

Nombre del documento:

Plan estratégico para el control de calidad y gestión de riesgos

Versión actual: 1.0

Proyecto: MLEDGE - Aprendizaje automático en la nube y en el borde (Cloud and Edge Machine Learning)

Paquete de trabajo: P0 - Gestión del Proyecto

Tarea: A0.1: Gestión del proyecto

Entregable: E0.2 - M12 - Informe de Progreso

Autores: Santiago Andrés (IMDEA)

Revisores: Nikolaos Laoutaris (IMDEA)

Historial de Versiones

Versión	Fecha	Resumen de modificaciones
Version 1.0	22-12-2023	Versión inicial del documento

Índice

1. Introducción

Este Plan de Calidad y Control de Riesgos, se puede considerar como un documento introductorio del proyecto y como fuente única de información de los principales trámites a cumplir para la implementación exitosa del proyecto (tanto técnica como financiera). Además, es un instrumento operativo destinado a apoyar a los socios del consorcio en la ejecución de las actividades del proyecto, y está dirigido principalmente a puntos de contacto administrativo. Este documento contiene un conjunto de directrices para garantizar:

- Una coordinación estratégica.
- La adecuada comunicación entre socios y el intercambio de información.
- El respeto de las normas, especialmente sobre el desempeño de las actividades, informes, entregables, entrega de productos según lo previsto, etc.
- Controlar la calidad de la ejecución y de los entregables.
- La gestión financiera adecuada del proyecto.
- Los derechos de la propiedad intelectual.
- Las medidas más efectivas de cara a riesgos potenciales.

2. Memoria de actividades

2.1. Objetivo del proyecto

El objetivo del proyecto MLEDGE es impulsar la implementación de FL como una capa intersectorial independiente pero optimizada sobre CloudEdge, utilizando aplicaciones y datos del mundo real para demostrar que esta sinergia puede producir grandes beneficios para todos. Con ello se podrá habilitar un ecosistema próspero de servicios FL en el borde seguros y eficientes capaces de facilitar el uso de datos personales y B2B confidenciales para entrenar modelos de ML para consumidores mientras se protege la privacidad de los datos y de sus propietarios.

Para allanar el camino a la adopción del FL en el borde de la red para un creciente número de aplicaciones que empleen modelos de ML, MLEDGE persigue el desarrollo de técnicas, librerías y componentes que permitan poner en marcha más ágilmente estos servicios. La Figura 1 resume la arquitectura de MLEDGE y los bloques del proyecto.

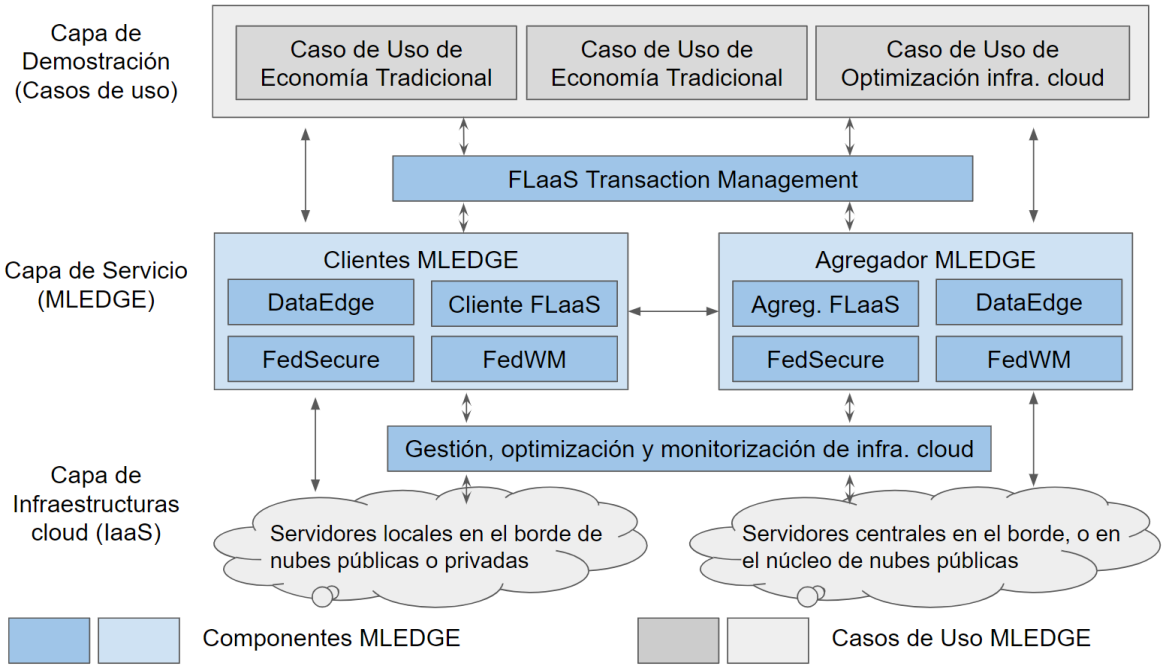

Figura 1. Diagrama de bloques de MLEDGE

2.2. Estructura de ejecución

. En la Figura 2 se proporciona un esquema de los paquetes de trabajo del proyecto y las relaciones entre ellos. El proyecto se estructura en 7 paquetes de actividades (P0-P6). P0 cubre la gestión del proyecto, y P1 tiene como objetivo definir los requisitos de los casos de uso y el diseño de la arquitectura del proyecto. Adicionalmente, se prevén 4 paquetes de trabajo técnicos (P2-P5), y un último paquete (P6) orientado a mostrar las pruebas de

concepto, y a la diseminación, explotación y comunicación de los resultados del proyecto.

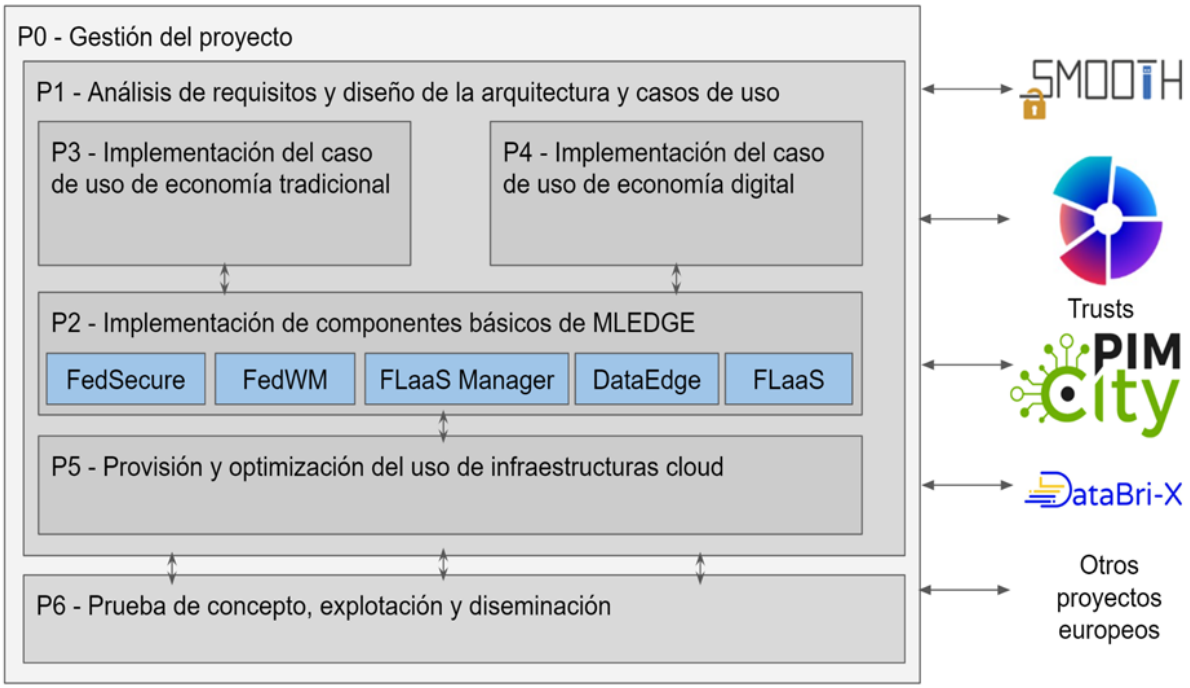

Figura 2. Paquetes de trabajo de MLEDGE

A continuación se ofrece un resumen de los avances y actividades por paquete de trabajo.

2.3. Actividad por paquete de trabajo

2.3.1. P0 - Gestión del Proyecto

Como parte de este paquete de trabajo se han realizado las siguientes actividades:

- Configurar y mantener la infraestructura tecnológica, incluidos sitio web y listas de correo.
- Organizar reuniones periódicas de seguimiento.
- Gestionar la coordinación estratégica mediante revisión periódica de la visión del proyecto, análisis y solución de problemas en la implementación del plan de trabajo.
- Controlar la calidad de la ejecución y los entregables del proyecto
- Elaborar una presentación oficial del proyecto y presentarla al equipo de trabajo
- Coordinar los diferentes componentes técnicos
- Supervisar el proceso de licitación de los componentes técnicos del proyecto
- Elaborar resúmenes de reuniones periódicas de equipos de investigación
- Elaborar el presente informe de actividad

2.3.2. P1 - Análisis de requisitos y diseño de la arquitectura y casos de uso

Como parte de este paquete de trabajo se han realizado las siguientes actividades:

- Investigar el estado del arte de los diferentes componentes técnicos del proyecto

- Definir un conjunto de escenarios iniciales y obtener para cada caso de uso los requisitos de rendimiento.
- Definir los requisitos y elaboración de los pliegos para la subcontratación de las empresas encargadas de los paquetes técnicos.
- Supervisar desde un punto de vista técnico el proceso de subcontratación, que se desarrolló en dos partes debido a que la primera licitación se declaró desierta.
- Revisar y valorar técnicamente las diferentes ofertas recibidas en respuesta a los pliegos
- Elaborar los siguientes entregables:
    - E1.1 – M6 – Requisitos y diseño de la arquitectura y casos de uso (preliminar) (pdf)
    - E1.2- M6 – Pliegos del proceso de licitación (enlace al proceso de licitación)
    - E1.2 – M9 – Pliegos del segundo proceso de licitación (enlace al proceso de licitación). ESTADO: Pendiente de adjudicación.
    - E1.3 - M12 - Requisitos y diseño de la arquitectura y casos de uso (final).

- Reporte a la gestión del proyecto

2.3.3. P2 - Implementación de componentes básicos de MLEDGE

Como parte de este paquete de trabajo se han realizado las siguientes actividades:

- Actividades propias de investigación en los correspondientes campos
- Desarrollos vinculados a los diferentes componentes
- Documentación de la actividad investigadora
- Elaboración de artículos científicos - ver detalle en P6 y en entregable 2.1 del proyecto.
- Elaboración de material para charlas y diseminación
- Elaboración del E2.1 - M12 - Versión preliminar de los componentes básicos de MLEDGE, que ofrece un mayor detalle de los avances en el paquete 2 de trabajo.

2.3.4. P3 - Implementación del caso de uso de economía tradicional

Este paquete de trabajo comenzará su actividad a partir del mes 13 de proyecto.

2.3.5. P4 - Implementación del caso de uso de economía digital

Este paquete de trabajo comenzará su actividad a partir del mes 13 de proyecto.

2.3.6. P5 - Provisión y optimización de infraestructuras cloud

Este paquete de trabajo comenzará su actividad a partir del mes 13 de proyecto.

2.3.7. P6 - Prueba de concepto, explotación y diseminación

Como parte de este bloque de trabajo se han realizado las siguientes actividades:

- Establecimiento de una página web del proyecto (https://mledge.networks.imdea.org/)
- Establecimiento de una página en LinkedIn (https://www.linkedin.com/company/mledge-project/)
- Elaboración de material y diseminación de trabajos en página web y redes sociales
- Presentaciones y charlas de diseminación del trabajo en el proyecto.

A continuación se resumen las publicaciones científicas que se han dado como resultado de los desarrollos del proyecto hasta la fecha:

- Santiago Andres Azcoitia, Costas Iordanou, y Nikolaos Laoutaris. (2023) Understanding the Price of Data in Commercial Data Marketplaces. IEEE International Conference on Data Engineering ICDE. Los Angeles, California, USA. April 2023.

- Tianyue Chu, Alvaro Garcia-Recuero, Costas Iordanou, Georgios Smaragdakis, y Nikolaos Laoutaris Securing Federated Sensitive Topic Classification against Poisoning Attacks. Network and Distributed System Security (NDSS) Symposium. 2023.

- Devriş Işler, Elisa Cabana, Álvaro Garcia-Recuero, Gieorgia Koutrika, Nikolaos Laoutaris, FreqyWM: Frequency Watermarking for the New Data Economy, aceptado para publicación en IEEE International Conference on Data Engineering ICDE'24.

- Devriş Işler, Seoyeon Hwang, Yoshimichi Nakatsuka, Nikolaos Laoutaris, y Gene Tsudik. "Puppy: A publicly Verifiable Watermarking Protocol". En preimpresión.

Estos artículos se incluyen como parte del entregable 2.1 del paquete de trabajo 2 del proyecto. Adicionalmente, se han realizado las siguientes publicaciones y charlas:

- Santiago Andres Azcoitia. Presentación del paper Understanding the Price of Data in Commercial Data Marketplaces. en el IEEE International Conference on Data Engineering ICDE. Los Angeles, California, USA. April 2023. (enlace al Video).

- Tianyue Chu. Presentación del paper Securing Federated Sensitive Topic Classification against Poisoning Attacks. Network and Distributed System Security (NDSS) Symposium. 2023.

- Santiago Andrés Azcoitia y Alba Ribera Martínez Data Marketplaces and the Data Governance Act: A Business Model Perspective. September 2023. Kluwer Competition Law Blog.

- Santiago Andrés Azcoitia y Alba Ribera Martínez Data Marketplaces and the Data Governance Act: A Business Model Perspective. Noviembre 2023. Charla PLAMADISO (Platforms, Markets, and the Digital Society) en el Weizenbaum Institute for the Networked Society.
- Santiago Andrés Azcotiia, Charla: Towards a Human-centrc data economy. Escuela Técnica Superior de Ingenieros de Telecomunicación, Universidad Politécnica de Madrid.

3. Siguientes pasos

En los primeros meses de 2024 se iniciaran las actividades con las empresas adjudicatarias de los paquetes de trabajo 3, 4 y 5. En paralelo, se seguirá avanzando con las actividades de investigación de acuerdo con el plan de trabajo, que se irán progresivamente integrando en los casos de uso según se trabaje con estas compañías. El objetivo de los primeros seis meses de 2024 es precisamente definir estos casos de uso y las demostraciones que se harán en los mismos de los diferentes componentes técnicos de MLEDGE.

Respecto a la actividad de publicaciones, se estarán desarrollando los siguientes artículos científicos:

- Un modelo de aprendizaje federado de precios de los datos en mercados de datos comerciales
- Un modelo de mercado de datos federado y algoritmos de optimización de adquisición de datos por parte de los compradores
- Un modelo de aprendizaje federado utilizando BitTorrent para la comunicación con garantías de privacidad entre los diferentes participantes

Adicionalmente, y debido al avance de los componentes técnicos se pondrá más énfasis en las actividades de diseminación de los trabajos realizados y en proveer de contenido a la página web y las redes sociales sobre el proyecto.