PLIEGO DE PRESCRIPCIONES TÉCNICAS QUE HAN DE REGIR EN EL CONTRATO DE UNA LICENCIA, DESPLIEGUE E INTEGRACIÓN DE LA PLATAFORMA DE APRENDIZAJE FEDERADO PARA EL ESPACIO DE DATOS CREADO EN EL MARCO DEL PROYECTO SHARE: SPANISH CANCER DATA SPACE (TSI-100123-2024-3[[1]](#footnote-1)) PARA LA FUNDACIÓN PARA LA INVESTIGACIÓN BIOMÉDICA DEL HOSPITAL UNIVERSITARIO PUERTA DE HIERRO MAJADAHONDA A ADJUDICAR POR PROCEDIMIENTO ABIERTO SIMPLIFICADO

**ÍNDICE**

[1 OBJETO Y FINALIDAD DEL CONTRATO 3](#_Toc204589765)

[2 LEGISLACIÓN 4](#_Toc204589766)

[3 CARACTERÍSTICAS TÉCNICAS DEL PRODUCTO 4](#_Toc204589767)

[3.1. Características de la estructura de la plataforma 5](#_Toc204589768)

[3.2. Funcionalidades 5](#_Toc204589769)

[3.3. Requerimientos técnicos 6](#_Toc204589770)

[3.4. Integración con conectores de espacio de datos 7](#_Toc204589771)

[3.5. Otros requisitos 7](#_Toc204589772)

[4 CONDICIONES Y PLAZOS DE ENTREGA 7](#_Toc204589773)

**PLIEGO DE PRESCRIPCIONES TÉCNICAS QUE HAN DE REGIR EN EL CONTRATO DE UNA LICENCIA, DESPLIEGUE E INTEGRACIÓN DE LA PLATAFORMA DE APRENDIZAJE FEDERADO PARA EL ESPACIO DE DATOS CREADO EN EL MARCO DEL PROYECTO SHARE: SPANISH CANCER DATA SPACE (TSI-100123-2024-3) PARA LA FUNDACIÓN PARA LA INVESTIGACIÓN BIOMÉDICA DEL HOSPITAL UNIVERSITARIO PUERTA DE HIERRO MAJADAHONDA A ADJUDICAR POR PROCEDIMIENTO ABIERTO SIMPLIFICADO**

# OBJETO Y FINALIDAD DEL CONTRATO

El presente contrato tiene por objeto la licencia, despliegue, integración y soporte de una plataforma de aprendizaje federado que permita el desarrollo colaborativo de modelos predictivos en el marco del proyecto SHARE: Spanish Cancer Data Space así como el soporte de incidencias y el servicio de mantenimiento de la licencia del primer año para la Fundación de Investigación Biomédica del Hospital Universitario Puerta de Hierro (en adelante, FIBHUPHM).

Esta plataforma facilitará la federación de datos sensibles sin necesidad de su transferencia, garantizando el cumplimiento normativo en materia de protección de datos (RGPD) y los principios de soberanía digital establecidos por las iniciativas europeas GAIA-X e IDSA.

La licencia incluirá:

* El suministro de una plataforma de aprendizaje federado interoperable.
* El despliegue técnico inicial de dicha plataforma, incluyendo la integración con el conector EDC del espacio de datos, los sistemas de almacenamiento clínico existentes y los módulos de gobernanza.
* El soporte técnico y funcional durante el despliegue y la operación.

Esta plataforma es un componente esencial del ecosistema tecnológico del proyecto SHARE, y debe permitir entrenar modelos de inteligencia artificial de forma distribuida sobre los datos locales alojados en distintos hospitales asociados al Grupo Español de Cáncer de Pulmón (GECP), sin que dichos datos abandonen su ubicación original.

La solución deberá estar basada en tecnologías robustas, escalables, compatibles con arquitecturas cloud y contener mecanismos integrados de preservación de la privacidad, federación segura, trazabilidad y control de acceso, en línea con el estado del arte del aprendizaje federado y las tecnologías de privacidad aplicadas a la inteligencia artificial.

A lo largo del presente pliego se describen concretamente los componentes incluidos dentro del objeto de este contrato, así como el alcance de la licencia propuesta.

# LEGISLACIÓN

Todos los productos presentados a este procedimiento, deberán cumplir la legislación vigente que sea de aplicación, incluyendo:

* El Reglamento (UE) 2022/868, del Parlamento Europeo y del Consejo, de 30 de mayo de 2022, relativo a la Gobernanza Europea de los Datos (Reglamento Europeo de Gobernanza de Datos)
* Las disposiciones específicas relativas a la protección de datos y los espacios de compartición de datos en el marco de la Estrategia Europea del Dato
* Respetar la normativa medioambiental aplicable (Principio de no causar un perjuicio significativo al medioambiente o «Do No Significant Harm» –DNSH–), de acuerdo a lo establecido en el Reglamento (UE) número 2020/852, del Parlamento Europeo y del Consejo, de 18 de junio de 2020,
* Reglamento (UE) 2016/679, del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos, junto a la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales.
* Reglamento (UE) 2021/694, del Parlamento Europeo y del Consejo, de 29 de abril de 2021, por el que se establece el Programa Europa Digital 0.
* Real Decreto 1494/2007, de 12 de noviembre, por el que se aprueba el Reglamento sobre las condiciones básicas para el acceso de las personas con discapacidad a las tecnologías, productos y servicios relacionados con la Sociedad de la Información y medios de comunicación social.

El contratista deberá respetar el carácter confidencia de aquella información a la que tenga acceso con ocasión de la ejecución del contrato a la que se le hubiese dado el referido carácter en los pliegos o en el contrato, o que por su propia naturaleza deba ser tratada como tal, quedando el contratista sometido a la normativa nacional y europea en materia de protección de datos, siendo ésta, a su vez, una obligación contractual esencial (211.1.f LCSP).

La licencia deberá estar conforme y actualizada en el momento en el que se realice su suministro y **ser compatibles con los estándares de interoperabilidad establecidos por los espacios de datos europeos**.

# CARACTERÍSTICAS TÉCNICAS DEL PRODUCTO

El objeto de esta contratación comprende la licencia, despliegue, configuración e integración de una plataforma de aprendizaje federado, alineada con los estándares técnicos y regulatorios aplicables a espacios de datos clínicos europeos. La solución deberá cumplir las siguientes características:

## Características de la estructura de la plataforma

* Tipo de licencia: La solución se entregará bajo licencia anual, con posibilidad de escalar horizontalmente a nuevos nodos.
* Arquitectura modular: La plataforma deberá tener una arquitectura basada en microservicios o contenedores, con soporte para nodos heterogéneos (edge, cloud, on-premise).
* Interoperabilidad: Deberá integrarse con los conectores EDC del espacio de datos (conforme a la arquitectura IDS-RAM y Gaia-X) y soportar el estándar OMOP CDM para la representación de datos clínicos.
* Portabilidad: La solución deberá poder desplegarse tanto en infraestructuras locales (hospitalarias) como en nubes públicas o privadas, garantizando la independencia del proveedor de infraestructura.
* Estabilidad y escalabilidad: Se exigirá tolerancia a fallos, escalado automático y capacidad para gestionar múltiples experimentos federados concurrentemente.
* Gestión de identidades: Integración con sistemas de identidad federada (eIDAS/eID, cartera digital, DID) y soporte para autenticación robusta con control de acceso basado en roles (RBAC).

El licitador debe contar con un sistema profesional de soporte que permita el seguimiento de las incidencias y soportes solicitados.

## Funcionalidades

La plataforma deberá proporcionar un conjunto completo de funcionalidades que permitan operar un sistema de aprendizaje federado interoperable, seguro, extensible y conforme a las exigencias normativas y técnicas del ecosistema europeo de espacios de datos. En concreto, deberá incluir:

* + Orquestación centralizada y descentralizada de experimentos federados, con soporte para ejecución síncrona o asíncrona, coordinada desde un nodo central o desde una lógica peer-to-peer.
  + Diseño modular de flujos de entrenamiento, que permita seleccionar algoritmos, definir ciclos de entrenamiento, condiciones de parada, métricas de evaluación y validaciones cruzadas, sin requerir programación explícita por el usuario.
  + Gestión dinámica de participantes, permitiendo alta, baja, validación e inclusión temporal de nodos de entrenamiento con control de identidad, rol y ámbito de participación (full-node, edge-node, observer).
  + Configuración de estrategias de agregación personalizadas, incluyendo algoritmos clásicos (FedAvg, Split) y robustos, con posibilidad de incluir extensiones propias.
  + Gestión de políticas de entrenamiento y privacidad, incluyendo la posibilidad de fijar niveles de privacidad diferencial, anonimización por ruido, límites de varianza o thresholds de confianza en cada nodo.
  + Visualización de métricas agregadas y locales, incluyendo precisión, pérdida, AUC y otras medidas clínicas relevantes, tanto en cada ciclo como a lo largo del tiempo.
  + Control de versiones de modelos, incluyendo versión del modelo global, historiales de entrenamiento locales y comparación entre versiones.
  + Gestión de incidentes y errores federados, incluyendo alertas, dashboards de salud de nodos, notificaciones automáticas y políticas de reintento o aislamiento.
  + Auditoría completa del ciclo de vida de los experimentos, incluyendo logs, trazabilidad de modelos y registros de metadatos asociados a cada entrenamiento.
  + Exportación e interoperabilidad de modelos entrenados, permitiendo la conversión de los modelos finales a formatos estándar como ONNX, PMML, o TensorFlow SavedModel para integración en plataformas clínicas.
  + Interfaz gráfica para diseño de experimentos, para usuarios no técnicos, incluyendo asistentes para la selección de nodos, algoritmos y parámetros.
  + Integración nativa con el conector EDC del espacio de datos, como plano de datos federado, cumpliendo con los estándares de DCAT y ODRL para la definición de contratos y políticas de uso.
  + Compatibilidad con técnicas de privacidad avanzada, tales como privacidad diferencial.
  + APIs RESTful que permitan programar, consultar, modificar y automatizar entrenamientos y análisis desde sistemas externos.
  + Soporte para aprendizaje federado horizontal y vertical, permitiendo escenarios clínicos heterogéneos con diferentes tipos de datos distribuidos (e.g. imagen, texto, registros estructurados).
  + Gestión granular de permisos y control de acceso, por roles, centros, proyectos y tipos de datos/modelos, integrable con sistemas corporativos de identidad como LDAP, SAML o eIDAS Wallet.
  + Simulación y testeo de escenarios federados, en entorno de sandbox, para permitir la validación previa de configuraciones sin impacto sobre entornos reales o clínicos.

## Requerimientos técnicos

La licencia previamente descrita deberá incluir los siguientes requerimientos:

* + La plataforma debe ser compatible con sistemas de bases de datos PostgreSQL, Redis u otros recomendados en entornos federados distribuidos.
  + Deberá ejecutarse en entornos Linux modernos, en contenedores Docker/Podman y orquestada con Kubernetes/Docker Compose.
  + Debe permitir la instalación tanto en modalidad SaaS, como on-premise o edge según el tipo de nodo.
  + Deberá soportar comunicaciones seguras TLS 1.3, autenticación mTLS y cifrado de extremo a extremo en canales de federación.

## Integración con conectores de espacio de datos

La plataforma deberá incluir un módulo de integración nativo o adaptador que permita su funcionamiento como plano de datos federado dentro del conector EDC, incluyendo:

* + Interoperabilidad semántica (con soporte DCAT, ODRL, vocabularios).
  + Compatibilidad con contratos de uso y políticas dinámicas.
  + Capacidad de descubrimiento de nodos participantes a través del Catálogo Federado

## Otros requisitos

La utilización de la nueva licencia deberá incluir un servicio de atención de los usuarios para resolver dudas e incidencias.

Para llevar a cabo el objeto de esta licitación es necesario un equipo técnico funcional capaz de desplegar esta licencia dentro de los plazos establecidos.

Durante la ejecución del contrato la empresa adjudicataria deberá ofrecer un servicio de mantenimiento de la licencia contratada.

# CONDICIONES Y PLAZOS DE ENTREGA

La entrega de la licencia y los servicios asociados al mismo objeto de la presente licitación habrán de realizarse en un plazo de un (1) mes a partir de la formalización del contrato.

POR EL ÓRGANO DE CONTRATACIÓN,

Presidente del Patronato

CONFORME

EL ADJUDICATARIO

1. Proyecto concedido por la Convocatoria de ayudas del programa de espacios de datos sectoriales para la Transformación Digital de los Sectores Productivos Estratégicos mediante la creación de demostradores y casos de uso de espacios de compartición de datos, financiado por los Fondos Next Generation [↑](#footnote-ref-1)