

Nº Expediente: PA4/2025

PLIEGO DE PRESCRIPCIONES TÉCNICAS DEL CONTRATO DE SUMINISTRO DE LICENCIAS EN SaaS PARA ACTIVACIÓN DE FUNCIONALIDADES DE GOBIERNO DE DATOS Y LINAJE 360, EN EL MARCO DEL PROYECTO “HEALTHDATA@MAD-R&I” ACORDE AL ESPACIO EUROPEO DE DATOS EN SALUD, DE LA CONVOCATORIA DEL MINISTERIO PARA LA TRANSFORMACIÓN DIGITAL Y DE LA FUNCIÓN PÚBLICA EN EL MARCO DEL PLAN DE RECUPERACIÓN, TRANSFORMACIÓN Y RESILIENCIA- NEXT GENERATION EU (PROGRAMA ESPACIOS DE DATOS SECTORIALES), MEDIANTE PROCEDIMIENTO ABIERTO SIMPLIFICADO Y TRAMITACIÓN URGENTE.

Primero. Justificación y Objeto del Contrato

El objeto del contrato será el suministro destinado a la adquisición de licencias de uso en modalidad Software as a Service (SaaS) del software Octopai, propiedad de Cloudera, el cual permite la automatización de procesos de inteligencia de datos mediante funcionalidades como data lineage, data discovery, impact analysis, entre otras.

El adjudicatario se obliga a proporcionar el acceso remoto al software alojado en la nube, así como a garantizar su disponibilidad y correcto funcionamiento durante la vigencia del presente contrato.

Asimismo, el objeto del contrato incluye la provisión de servicios complementarios como soporte técnico, mantenimiento correctivo y evolutivo, así como la realización de actualizaciones y mejoras durante el periodo de vigencia de las licencias.

En concreto, el objetivo de este contrato es activar y desplegar la funcionalidad de linaje 360 de Cloudera Octopai, con integración en la infraestructura Big Data ya existente, sustentada en tecnología Cloudera, y en conexión con herramientas analíticas como Power BI. Esta solución permitirá:

- Visualizar y analizar los flujos de datos a lo largo del ciclo de vida completo del dato
- Evaluar automáticamente el impacto de cambios en las fuentes de datos sobre los procesos analíticos aguas arriba y aguas abajo
- Gobernar el uso de los recursos de información en entornos interconectados y facilitar la toma de decisiones informadas sobre su trazabilidad y transformación.

De esta manera, la plataforma se adapta a los requisitos de interoperabilidad, control y reutilización propios de los espacios europeos de datos, garantizando la eficiencia, la fiabilidad y el cumplimiento de los principios FAIR (Findable, Accessible, Interoperable, Reusable) en el tratamiento del dato sanitario para fines de investigación e innovación.

El contrato propuesto es un contrato de suministro destinado a la adquisición de licencias de uso de software para la ampliación de la plataforma Big Data analytics mediante hibridación en cloud y activación de funcionalidades de gobierno de datos y linaje 360, con el objetivo de desarrollar capacidades adicionales e integrarlas en la plataforma tecnológica del Espacio de Datos Sanitarios de Uso Secundario de la Comunidad de Madrid con fines de investigación e innovación “HealthData@MAD-R&I”, según lo previsto en el artículo 16 de la Ley 9/2017, de 8 de noviembre, de Contratos del Sector Público.

El presente contrato se enmarca en el desarrollo del proyecto Espacio de datos de salud de uso secundario con fines de investigación científica e innovación HealthData@MAD-R&I, de la Fundación para la Investigación e Innovación Biosanitaria de Atención Primaria de la Comunidad

de Madrid (en adelante, FIIBAP o La Fundación) y financiado mediante la Resolución de concesión definitiva de 20 de diciembre de 2024 en el marco del Plan de Recuperación, Transformación y Resiliencia (PRTR), aprobado por la Decisión de Ejecución del Consejo ECOFIN de 13 de julio de 2021 [COM (2021) 322 final].

El proyecto HealthData@MAD-R&I tiene como objetivo general definir, desarrollar, validar y documentar casos de uso relacionados con la investigación y la innovación que permitan validar los servicios del Espacio de datos sanitarios electrónicos de uso secundario. En el marco del mismo, se han previsto cuatro casos de uso en diferentes áreas de investigación que sirven como demostradores del potencial científico-tecnológico y comercial que los datos de salud aportan a la economía del dato en el sector sanitario:

- Caso de uso 1: Optimización de las derivaciones de sujetos con enfermedades reumáticas y musculoesqueléticas entre atención primaria y reumatología.
- Caso de uso 2: Trayectorias asistenciales de las mujeres largas supervivientes de cáncer de mama SURBCANMADRID.
- Caso de uso 3: Hospitalización No Programada (MAD-HNP.eSTRATA).
- Caso de uso 4: Modelo explicativo en vida real sobre la efectividad de las estatinas en la reducción de eventos cardiovasculares y la mortalidad en la población anciana sin antecedentes de enfermedad cardiovascular de la Comunidad de Madrid.

Para lo anterior, se precisa ampliar las capacidades de la Plataforma HealthData@MAD-R&I que dará soporte a los casos de uso, por lo que resulta necesario la adquisición de licencias que cubran todas las necesidades y requerimiento técnicos.

Las características técnicas principales de los programas y prestaciones a contratar se especifican en el punto tercero del presente pliego.

Segundo. Entorno Tecnológico Actual

HealthDataMADR&I formará parte del Espacio de Datos de la Comunidad de Madrid, que actualmente utiliza la plataforma *Cloudera Data Platform*, con los componentes que se describen en el punto tercero del presente pliego. Dicha plataforma centraliza datos y herramientas para iniciativas de analítica descriptiva y avanzada.

Tercero. Características de las prestaciones

Con respecto a las licencias objeto del contrato, se admiten programas:

- Puestos a disposición en modalidad de nube.
- Para su instalación en infraestructura local.
- En cualquier modalidad de puesta a disposición.

1. Requisitos funcionales de los programas a suministrar

La Comunidad de Madrid, dentro del Plan de Salud Digital de la Consejería de Digitalización, tiene en curso varias iniciativas que persiguen la definición, diseño y desarrollo de nuevos sistemas analíticos con modelos predictivos que permita extraer valor de los datos existentes, Sanitarios y no Sanitarios del Servicio Madrileño de Salud, por ejemplo, optimizar la gestión global y local de la

capacidad, alerta temprana de variabilidad en la práctica clínica, predicción, seguimiento y actuación en el ámbito de la cronicidad, detección temprana de desajustes y alertas epidemiológicas ... y aplicarlos a la toma de decisiones.

Esencialmente, para gestionar y procesar volúmenes ingentes de datos y para detectar patrones en los mismos, se utilizan tecnologías denominadas como "Big Data" las cuales son capaces de utilizar dichos datos para extraer conclusiones de valor.

Actualmente se dispone de un repositorio central a la fecha, con más de 13.000 millones de registros (Data Lake) que se sustenta en una infraestructura de Big Data basada en Cloudera on Premise para el tratamiento de información, en el CPD Central de la DGSD Athene@ en el Hospital 12 de octubre y que está estructurado modularmente para que los procesos de alimentación e ingesta, así como de explotación de datos, permitan Interconexión a otros subsistemas de análisis de datos.

Además, este data-lake permitirá la interoperabilidad de informaciones con la Historia Clínica del Sistema Nacional de Salud, proyecto que promueve el Ministerio de Sanidad para mejorar el acceso a la información. Por tanto, la actual infraestructura está compuesta de varios nodos que disponen implementados los distintos roles en nodos diferenciados.

Debido a las necesidades de crecimiento en función de los organismos y departamentos que consuman y realicen modelos de IA avanzada en la plataforma, se ve necesario aumentar el control del linaje y uso de los datos que están entrando en la plataforma y se están consumiendo por la misma mediante Power BI. Para ello, es necesario activar la licencia de linaje 360 que cubre el descubrimiento, linaje, búsqueda, y catálogo de todo metadato que interaccione con la plataforma, a través de fuentes origen o herramientas de BI.

Para ello, se van a desplegar las siguientes características de las licencias de Cloudera

SKU	Unidades	Descripción
OCT-PLAT-ENT-BUS	1	Octopai by Cloudera Platform - Enterprise Subscription - Business-Level Support. Includes one Octopai Environment, 5 Connectors, 10 Users, 25 Business Users and 550 Activities per month
OCT-PLAT-CON5-BUS	1	Octopai by Cloudera Platform - Subscription for 5 Connectors - Business-Level Support.

De esta manera se dispondrá del servicio SaaS de Cloudera Octopai y las siguientes características:

- 5 + 5 conectores a distintas fuentes de datos (Power BI, Servidores Oracle, etc. hasta 10) dentro de la lista de servidores indicada a continuación.

- 10 usuarios de acceso a linaje
- 25 usuarios de acceso a catálogo y Knowledge hub
- 550 Actividades de IA al mes para obtener información adicional de los recursos mediante motores de IA generativa.

Los conectores compatibles son:

Source Type	Connector Name	Connector Type
ETL	Microsoft ADF - Azure Data Factory	Azure Data Factory Name
	Informatica Power Center	Informatica Repository
	Informatica - Cloud IDMC	Account (URL)
	IBM - DataStage	Project
	Microsoft - SSIS	SQL Server Database or Root Folder
	Oracle - ODI (on-prem)	ODI Repository
	Oracle - OWB	Project
	Microsoft Azure Databricks	Cluster
	Qlik (Sense + View) - ETL	Server
	Spark	Cluster
	DBT	File Server
	Talend (*)	Project
	AirFlow (*)	Account (URL)
	Aws Glue (*)	clientID and Region
	Stored Procedure - Oracle	Database
	Stored Procedure - Teradata	Database
	Stored Procedure - Netezza	Database
	Stored Procedure - Vertica	Database
	Stored Procedure - Snowflake	Account (URL)
	Stored Procedure - Microsoft SQL Server/Azure SQL	SQL Server

	SQL Scripts Files (PL/SQL, T-SQL, BTEQ SQL, Teradata SQL, Netezza SQL, Vertica SQL)	Root Folder
Database/DWH	Oracle DB	Database
	Hive	Server
	Impala	Server
	Kafka (*)	Broker
	Oracle Exadata	Database
	Oracle Autonomous Database	Database
	Microsoft SQL Server	SQL Server
	Netezza	Database
	Vertica	Database
	Teradata on-prem DB	Server Endpoint
	Teradata Vantage	Server Endpoint
	Textual Files	File Server
	Microsoft Azure SQL Data Warehouse	Server Instance
	Azure Synapse Analytics	Server Instance
	Snowflake	Account (URL)
	Microsoft SQL Azure	SQL Server
	Amazon Red Shift	Server
	MySQL - On prem/Cloud	Server
	PostgreSQL - On prem/Cloud	Server
	GreenPlum - On prem/Cloud	Server
	MariaDB - On prem/Cloud	Server
	Google BigQuery	Project ID
	DB2 iSeries / LUW	Server
Analysis Services	Microsoft SSAS - Tabular	SQL Server

Reports & Analytics	Microsoft SSAS - OLAP	SQL Server
	Microsoft SSAS - Azure Tabular	SQL Server
	IBM Cognos Report Studio	Server
	SAP - Business Objects (BO)	Server
	Oracle - OBIEE/OAS	Server
	Microsoft SSRS	Site (URL)
	QlikView	Root Folder
	QlikSense	Server (API) / URL
	MicroStrategy (MSTR)	Project
	Tableau Server / Tableau Server on Cloud	Site (All projects inside)
	Microsoft PowerBI	Site / Instance (URL) (Octopai is pricing by instance of PowerBI with all folders --> workspaces --> Reports/Datasets in that instance - as 1 connection)
	QlikView - Report	Server
	QlikSense - Report	Server
	Google Looker	Account
(*) roadmap/coming soon		

Para acceder a estas fuentes de datos, es necesario que se disponga de un equipo con las siguientes características con visibilidad a esas fuentes de datos, no incluido en esta propuesta, donde se instalará el Agente de la herramienta (que si está incluido)

Windows Server

- 2016 OS and above on the organization Domain
- .NET Framework 6 and above (offline installer)
- .NET Core 6 Runtime and above (offline installer)
- 8 Cores (CPU); 16 GB – RAM; 250 GB - Hard Disk
- Reserved ports for local use 5656 & 5657
- cURL installed

Linux Server

- Please get in touch with the Octopai Support team at support@octopai.com for more details and instructions.
- Linux - Distro: RHEL 7.9 and above | Oracle 8 and above
- .NET Core 6 Runtime and above <https://dotnet.microsoft.com/download>
- 8 Cores (CPU); 16 GB – RAM; 250 GB - Hard Disk
- Reserved ports for local use 5656 & 5657
- cURL installed

2. Servicios de instalación y soporte de los programas a suministrar

El adjudicatario realizará todas las tareas de instalación de las librerías en la plataforma y las tareas de soporte necesarias para garantizar el correcto funcionamiento de la herramienta y de la actividad diaria de la plataforma HealthData@MAD-R&I, garantizando los objetivos establecidos en el presente pliego.

El adjudicatario ejecutará el contrato de forma remota desde sus propias oficinas, de acuerdo con las políticas de seguridad establecidas por el órgano contratante, así como el órgano competente en materia de Digitalización de la Comunidad de Madrid.

El adjudicatario deberá tener el aval del fabricante de software para comercializar el producto y la gestión del soporte y mantenimiento especializado.

Quinto. Plazo de entrega de las licencias

El plazo máximo de entrega será de un (1) mes contado a partir de la fecha de inicio de ejecución del contrato. No se admiten entregas parciales.

La vigencia de las licencias será hasta el **15 de junio de 2026**, con independencia de la fecha de su activación. Durante todo ese periodo deberá prestarse el soporte técnico asociado a las mismas.

Sexto. Plazo de ejecución del contrato

La duración del contrato se extenderá hasta el 15 de junio de 2026. No obstante, el adjudicatario deberá realizar la entrega de las licencias objeto del suministro, así como, en su caso, su instalación y puesta en funcionamiento, en el plazo máximo de un (1) mes contado a partir del día siguiente a la formalización del contrato.

El cumplimiento del plazo de entrega tendrá la consideración de obligación esencial del contrato, sin perjuicio de posibles ampliaciones por causa debidamente justificada y aceptada por el órgano de contratación.

Séptimo. Coordinación del contrato

A efectos de lograr una correcta ejecución del contrato se designa como coordinador del mismo a la responsable del proyecto de HealthData@MAD-R&I en la Fundación para la Investigación e Innovación Biosanitaria de Atención Primaria (FIIBAP).

En Madrid, a fecha de última firma electrónica
La Presidencia del Patronato

Fdo.: Ana Isabel González González
Jefe de Proyecto.