

**PLIEGO DE PRESCRIPCIONES TÉCNICAS QUE HA DE REGIR EL SUMINISTRO DE LICENCIAS DE PRODUCTOS SOFTWARE, PARA EL PROCESAMIENTO DE LENGUAJE NATURAL E IMAGEN MÉDICA, EN FORMATO DE SUSCRIPCIÓN JUNTO INSTALACIÓN Y SOPORTE, PARA DESARROLLO DE CAPACIDADES ADICIONALES Y INTEGRACIÓN EN LA PLATAFORMA TECNOLÓGICA DEL ESPACIO DE DATOS SOBRE UNA INFRAESTRUCTURA INTERREGIONAL DIGITAL DE IMÁGENES RELACIONADAS CON EL CÁNCER DE PÁNCREAS (EDS-IRCP) «PANDORA», ACORDE AL ESPACIO EUROPEO DE DATOS EN SALUD, DE LA SEGUNDA 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. EXP. PAS 4/2026.**

## **Contenido**

<b>1.- JUSTIFICACIÓN Y OBJETO DEL CONTRATO.....</b>	<b>2</b>
<b>2.- ENTORNO TECNOLÓGICO ACTUAL.....</b>	<b>3</b>
<b>3.- CARACTERÍSTICAS DE LAS PRESTACIONES.....</b>	<b>3</b>
<b>4.- CONDICIONES DE ENTREGA E INSTALACIÓN Y SOPORTE DE LOS PROGRAMAS A SUMINISTRAR. ....</b>	<b>6</b>
<b>5.- COBERTURA DE LA GARANTÍA. ....</b>	<b>7</b>
<b>6.-MANUALES: .....</b>	<b>8</b>
<b>7.- PLAN DE SOPORTE.....</b>	<b>8</b>
<b>8.- PLAZO DE ENTREGA DE LAS LICENCIAS. ....</b>	<b>8</b>
<b>9.- PLAZO DE EJECUCIÓN DEL CONTRATO. ....</b>	<b>9</b>
<b>10.- COORDINACIÓN DEL CONTRATO.....</b>	<b>9</b>
<b>11.- PROTECCION DE DATOS Y DEBER DE CONFIDENCIALIDAD.....</b>	<b>9</b>

## 1.- JUSTIFICACIÓN Y OBJETO DEL CONTRATO.

La Fundación para la Investigación Biomédica del Hospital Universitario de La Princesa (en adelante la FIBHLPR) tiene entre sus fines fundacionales la gestión y promoción de programas y proyectos de investigación científica aplicada a la Biomedicina y a las Ciencias de la Salud. La actividad de investigación de la FIBHLPR se lleva a cabo bajo la estructura del Instituto de Investigación Sanitaria del Hospital de La Princesa (en adelante IIS Princesa).

La presente contratación tiene como objetivo contribuir al fortalecimiento y ampliación de la capacidad de análisis de datos clínicos y radiológicos relacionados con el cáncer de páncreas, mediante el suministro de licencias de software especializado en Procesamiento de Lenguaje Natural (PLN), así como la implementación de los servicios asociados a su instalación y soporte, en el entorno tecnológico de la plataforma EDS-IRCP-PANDORA y su convergencia con el Espacio Europeo de Datos en Salud. El objetivo final es mejorar la capacidad para utilizar datos existentes y nuevos para la detección precoz del cáncer de páncreas, la identificación de biomarcadores radiómicos y el desarrollo de modelos predictivos para la personalización del tratamiento.

El proyecto EDS-IRCP-PANDORA está financiado, en el marco de la segunda convocatoria para la concesión de ayudas en el ámbito de la digitalización para la transformación digital de los sectores productivos estratégicos mediante la creación de demostradores y casos de uso de Espacios de Compartición de Datos, por el Ministerio para la Transformación Digital y de la Función Pública y por la Unión Europea a través de los recursos financieros derivados del Plan de Recuperación, Transformación y Resiliencia – Next Generation EU (Programa Espacios de Datos Sectoriales), con número de expediente TSI-100123-2024-17.

El proyecto EDS-IRCP-PANDORA tiene como objetivo general definir, desarrollar, validar y documentar una infraestructura interregional digital de imágenes y datos clínicos relacionados con el cáncer de páncreas, integrando a hospitales en un ecosistema digital unificado alineado con la iniciativa SEDIA y con el Espacio Europeo de Datos en Salud (EHDS).

Para lo anterior, se precisa ampliar las capacidades de la plataforma EDS-IRCP-PANDORA, por lo que resulta necesaria la adquisición de licencias que cubran todas las necesidades y requerimientos técnicos descritos en el presente pliego.

Se precisa el suministro consistente en la adquisición de licencias y la implementación de los servicios asociados en el entorno tecnológico del DataLake de EDS-IRCP-PANDORA para fortalecer y ampliar la capacidad de análisis de datos clínicos y radiológicos, con el objetivo final

de mejorar la capacidad para utilizar datos existentes y nuevos para la detección precoz, la identificación de biomarcadores y el desarrollo de modelos predictivos en el ámbito del cáncer de páncreas. La plataforma EDS-IRCP-PANDORA se alinea con las directrices y especificaciones técnicas para facilitar el uso ininterrumpido de los datos sanitarios en toda Europa en el marco del Espacio Europeo de Datos Sanitarios (EHDS).

En concreto, el objetivo de este contrato es implantar una solución de Procesamiento de Lenguaje Natural para extraer valor de los datos sanitarios existentes en la organización y tomar decisiones informadas en el ámbito del diagnóstico y tratamiento del cáncer de páncreas, incluyendo:

- Facilitar la anonimización de datos de carácter personal en documentación clínica y radiológica
- Desarrollar herramientas de Detección de Entidades (Named Entity Recognition – NER – por sus siglas en inglés) especializadas en oncología pancreática
- Codificar a terminología estándar las entidades reconocidas en los informes clínicos y de imagen

Dicha solución se integrará con la plataforma Cloudera Data Platform, la cual centraliza datos y herramientas para iniciativas de analítica descriptiva y avanzada en el seno del ecosistema tecnológico de EDS-IRCP-PANDORA.

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

## **2.- ENTORNO TECNOLÓGICO ACTUAL.**

EDS-IRCP-PANDORA formará parte del ecosistema digital de datos sanitarios interregional 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.

## **3.- 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.

### **3.1. Requisitos funcionales de los programas a suministrar.**

La Fundación para la Investigación Biomédica del Hospital Universitario La Princesa (FIBHULP), en el marco del proyecto EDS-IRCP-PANDORA, tiene en curso varias iniciativas que persiguen la definición, diseño y desarrollo de nuevos sistemas analíticos con modelos predictivos que permitan extraer valor de los datos clínicos y radiológicos existentes. Entre las aplicaciones previstas se encuentran: la optimización del diagnóstico precoz del cáncer de páncreas, la identificación de biomarcadores radiómicos, la predicción de respuesta al tratamiento y la mejora de la planificación terapéutica mediante modelos de inteligencia artificial entrenados sobre datos interoperables de múltiples centros hospitalarios.

Actualmente, y dentro del proyecto Espacio de datos sobre una infraestructura interregional digital de imágenes relacionadas con el cáncer de páncreas – EDS-IRCP, se precisa gestionar y procesar un gran volumen de datos clínicos en texto libre (informes radiológicos, notas de evolución, informes anatomopatológicos) así como datos estructurados y metadatos de imagen (estudios de TC, RM y ecografía abdominal).

Actualmente se dispone de un repositorio central, que se sustenta en una infraestructura de Big Data basada en Cloudera (Data Platform private Cloud – CDP) para el tratamiento de información. 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 en los hospitales participantes en el proyecto. 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 de los distintos roles implementados en nodos diferenciados.

#### **Nodos de proceso para el plano de Control**

Estos nodos tienen implementados los roles del servicio de Kubernetes y el Cloudera Experiences (Cloudera Data Engineering (CDE) / Cloudera Machine Learning (CML) / Cloudera Data Warehouse (CDW)), el plano de control y el catálogo de datos de Cloudera.

#### **Nodos Master de Cloudera o equivalente**

Estos nodos tienen implementados los roles HDFS, YARN, HBASE, Sentry Server, StateStore Server y servidor de catálogo, IMPALA y Cloudera Manager.

### **Nodos Worker y/o Almacenamiento**

Estos nodos tienen implementados para el almacenamiento los roles HDFS Data Node, YARN Node Manager, HBASE Region Server, IMPALA Daemon, Search Worker Daemons y Kudu Tablet.

La solución objeto de este contrato debe permitir el uso de técnicas avanzadas de Procesamiento de Lenguaje Natural para identificar y extraer entidades relevantes en el contexto oncológico, anonimizar datos personales procedentes de Historias Clínicas Electrónicas e informes radiológicos, y codificar la información a terminologías sanitarias estándar. Asimismo, se integrará de manera transparente con el clúster de Cloudera, aprovechando las capacidades de la plataforma para la ingesta, procesamiento, análisis y disponibilización de grandes volúmenes de datos clínicos, garantizando así la escalabilidad y el rendimiento necesarios para cubrir el proyecto EDS-IRCP-PANDORA y la interoperabilidad entre los participantes.

Se precisa que las librerías puedan ejecutarse sobre Spark 3.X y se permita la reutilización, el entrenamiento y la combinación de modelos de IA para tareas como el reconocimiento de entidades, la clasificación de texto, la corrección ortográfica y gramatical, la respuesta a preguntas y la extracción de conocimiento.

Se debe permitir la preparación de datos para las soluciones RAG LLM, incluyendo la división de documentos, la limpieza, el enriquecimiento de metadatos, el resumen y el cálculo de incrustaciones.

Se solicita que el producto licenciado cumpla con:

- Extracción de información
- Clasificación de documentos
- Desambiguación de entidades
- Análisis contextual
- Puntuación del riesgo del paciente
- Ofuscación de datos
- Coherencia de nombres
- Coherencia de género
- Grupo de edad
- Formato
- Coherencia gramatical clínica

- Detector profundo de frases
- Corrección ortográfica médica
- Mapeo terminológico
- Aprendizaje Zero-Shot
- Entidades por pregunta
- Relaciones por pregunta
- Clasificación por Prompt
- Extracción de datos relativos

LA TOTALIDAD DE CONDICIONES ENUMERADAS EN ESTE APARTADO SON DE CARÁCTER ESENCIAL AL OBJETO DEL CONTRATO. LA FALTA DE CUMPLIMIENTO DE CUALQUIERA DE ELLAS SE CONSIDERARÁ CAUSA DE EXCLUSIÓN DE LA OFERTA.

#### **4.- CONDICIONES DE ENTREGA E INSTALACIÓN Y SOPORTE DE LOS PROGRAMAS A SUMINISTRAR.**

El adjudicatario realizará todas las tareas de instalación de las librerías en la plataforma EDS-IRCP-PANDORA y las tareas de soporte necesarias para garantizar el correcto funcionamiento de la herramienta y de la actividad diaria de la plataforma que da soporte al espacio de datos, 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 por la Fundación para la Investigación Biomédica del Hospital Universitario La Princesa (FIBHULP) y 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.

Los recursos humanos a incorporar al contrato, aunque los licitadores podrán incorporar un número mayor de recursos si lo consideran conveniente, son los siguientes. A continuación, se presenta la estimación de los recursos humanos necesarios para el desarrollo del contrato, de forma que los operadores económicos puedan disponer de margen suficiente para configurar su organización con la mayor eficiencia y ajustar de forma óptima su oferta.

Servicios / tareas a desarrollar	Perfil necesario	Número de recursos necesarios
Soporte extendido sobre licencias de la plataforma	Científico/a de Datos	1
Servicio de instalación y despliegue	Ingeniero/a de Datos	1

Ambos perfiles (Ingeniero/a de Datos y Científico/a de Datos) deben cumplir con dos tipos de requisitos:

### 1. Titulación académica:

- Ingeniero/a de Datos: Grado universitario de Nivel 2 (MECES 2) en Ingeniería o titulaciones técnicas.
- Científico/a de Datos: Grado universitario de Nivel 3 (MECES 3) en Biotecnología, Bioinformática, Ingeniería u otras titulaciones técnicas con especialización en análisis de datos biomédicos.

### 2. Certificaciones técnicas del fabricante:

- Para el perfil de Científico/a de Datos se requieren certificaciones del fabricante que acrediten formación técnica en el producto:
  - Generative AI for Data Scientists
  - Medical Language Models for Data Scientists

### 3. Cobertura de la garantía extendida del adjudicatario

La garantía extendida que debe prestar el adjudicatario durante todo el periodo de vigencia de las licencias se rige por:

### 5.- COBERTURA DE LA GARANTÍA.

La garantía extendida que debe prestar el adjudicatario durante todo el periodo de vigencia de las licencias se rige por:

- Soporte de nivel 1 y nivel 2 prestado por el adjudicatario a petición del organismo destinatario, en los términos descritos en el presente PPT.
- Soporte del adjudicatario al organismo para el acceso a la garantía del fabricante (acceso al soporte de nivel 3).
- Soporte a la instalación de actualizaciones.

Horario de contacto: Asistencia Empresarial durante 8 horas, 5 días a la semana (8x5).

## **6.-MANUALES:**

Se entregarán con la licencia todos los manuales y documentación correspondientes a la descripción, funcionamiento y operatividad de la solución, que deberán ser, como mínimo, los siguientes:

- MANUAL DE INSTALACIÓN, que deberá incluir los requisitos necesarios para la correcta implantación de la solución, así como los procedimientos de instalación, configuración y puesta en marcha.
- MANUAL DE USO, que deberá incluir una explicación de las funcionalidades del sistema, los modos de operación, las instrucciones de uso, así como las operaciones básicas para la verificación del correcto funcionamiento.
- DOCUMENTACIÓN TÉCNICA DE LA SOLUCIÓN, que deberá incluir una descripción general del sistema, sus componentes, estructura y características técnicas relevantes para su correcta utilización e integración.

Toda la documentación detallada en el párrafo anterior se deberá entregar en formato electrónico.

## **7.- PLAN DE SOPORTE.**

La oferta que resulte adjudicataria incluirá un plan de soporte que deberá incluir al menos:

- Soporte para el trabajo de ciencia de datos y anotación relacionado con modelos en español de Reconocimiento de Entidades Nombradas (NER), extracción de relaciones y modelos de aserción, incluyendo el desarrollo, entrenamiento y despliegue de modelos.
- Soporte para la adaptación de modelos y el desarrollo de pipelines.
- Revisión técnica y resolución de problemas.
- Orientación al equipo de anotación sobre metodología de anotación y control de calidad.
- Formación para los miembros relevantes del equipo.

Todas las actuaciones descritas anteriormente se considerarán incluidas en el objeto del contrato y, por tanto, en el alcance de la licencia suministrada, entendiéndose que su prestación forma parte inseparable de la misma.

## **8.- PLAZO DE ENTREGA DE LAS LICENCIAS.**

El plazo máximo de entrega será de 5 días naturales contados a partir de la fecha de inicio de ejecución del contrato. No se admiten entregas parciales.



## **9.- PLAZO DE EJECUCIÓN DEL CONTRATO.**

El contrato tendrá una duración de doce (12) meses desde su formalización. No obstante, el sistema deberá encontrarse plenamente desplegado y en funcionamiento efectivo, conforme a los hitos establecidos, antes del 30 de junio de 2026.

A estos efectos, la duración anual del contrato responde a la naturaleza de licencia del servicio contratado, manteniéndose dicha vigencia con independencia de la fecha en que se alcance el despliegue efectivo. En consecuencia, el eventual acortamiento del plazo necesario para la consecución de los hitos no implicará reducción alguna del importe de adjudicación, al haberse determinado el presupuesto en función del cumplimiento de los resultados previstos y de la disponibilidad de la licencia durante todo el periodo contractual.

## **10.- COORDINACIÓN DEL CONTRATO.**

El adjudicatario designará a un técnico competente que actuará como responsable del contrato y que será considerado interlocutor válido ante la Fundación. La persona designada será responsable de velar por el adecuado cumplimiento de lo acordado en el contrato y por la resolución de incidencias en la fase de ejecución.

A efectos de lograr una correcta ejecución del contrato, se designa como coordinador del mismo a la persona responsable del proyecto EDS-IRCP en la Fundación para la Investigación Biomédica del Hospital Universitario La Princesa (FIBHULP), entidad beneficiaria de la subvención TSI-100123-2024-17 otorgada en el marco de la Segunda Convocatoria para la Concesión de Ayudas en el Ámbito de la Digitalización para la Transformación Digital de los Sectores Productivos Estratégicos mediante Creación de Demostradores y Casos de Uso de Espacios de Compartición de Datos

## **11.- PROTECCION DE DATOS Y DEBER DE CONFIDENCIALIDAD**

La empresa seleccionada se compromete a guardar confidencialidad sobre las informaciones que reciba en el desarrollo de su trabajo, así mismo se compromete al cumplimiento de la normativa vigente en materia de protección de datos.

En Madrid,

Los efectos del presente documento se considerarán desplegados a partir de la fecha en la que haya firmado el último de los integrantes del Órgano de Contratación.

POR EL ÓRGANO DE CONTRATACIÓN,

<b>ÓRGANO DE CONTRACIÓN</b>	
<b>FIBHLPR</b>	
Dña. Rosario Ortiz de Urbina Barba	D. Isidoro González Álvaro

NOTA: Por la presente se hace saber que el presente documento asociado al expediente referenciado, y que se corresponde con el que ha sido publicado en el Portal de la Contratación de la Comunidad de Madrid, ha sido efectivamente firmado se encuentra dentro del expediente sito en las instalaciones de la Fundación para la Investigación Biomédica del Hospital Universitario La Princesa.