



**RESPUESTAS A LAS ACLARACIONES SOLICITADAS POR EMPRESAS PARA EL EXPEDIENTE A/SER-032533/2025 - DESARROLLO DE UN SISTEMA PARA EL SOPORTE Y AYUDA AL DIAGNÓSTICO DE ENFERMEDADES CARDÍACAS MEDIANTE INTELIGENCIA ARTIFICIAL, CON CARGO AL PLAN DE RECUPERACION TRANSFORMACIÓN Y RESILIENCIA DEL GOBIERNO DE ESPAÑA - FINANCIADO POR LA UNIÓN EUROPEA – NEXTGENERATIONEU**

Pregunta	Respuesta
<p>Sobre el Data LakeHouse corporativo:</p> <p>o ¿Podrían confirmar a qué Data LakeHouse se refiere el pliego?</p> <p>o ¿Será proporcionado por la DGSD (por ejemplo, la plataforma Cloudera Data Platform) o se espera que el contratista lo instale y configure?</p> <p>o Asimismo, se solicita confirmar si esta referencia corresponde a la infraestructura corporativa de datos (Cloudera) o a la base de datos Oracle mencionada en el entorno de persistencia.</p>	<p>El Datalake que se cita en el pliego es Cloudera Data Platform como se indica en la página 17 del PPT.</p> <p>La plataforma Cloudera Data Platform lo proporciona la DGSD.</p> <p>Se confirma lo indicado en las tablas de la página 17. En la arquitectura corporativa, Capa persistencia de datos: Hibernate, JPA, EhCache. En la capa Base de datos Oracle. Referente a aplicaciones web e integraciones. Y por otro lado, para Explotaciones, cuadros de mando e indicadores, Capa Datalake: Cloudera Data Platform.</p>
<p>Sobre el sistema de autenticación y acceso unificado (SSO):</p> <p>o ¿Dispone el Hospital Universitario Ramón y Cajal de un sistema unificado de autenticación (Single Sign-On) integrado en la infraestructura corporativa del SERMAS?</p> <p>o En caso afirmativo, ¿podrían indicar si se basa en OAuth2, SAML2 u otro protocolo, y si el contratista debe integrarse con dicho sistema o implementar su propio mecanismo de autenticación?</p>	<p>El personal del Hospital dispone de un sistema de autenticación y acceso unificado proporcionado por el SERMAS.</p> <p>Como se indica en la página 7 del PPT, el sistema se integrará con los sistemas de información hospitalaria (HIS) del hospital, por lo que deberá hacer uso del mismo sistema de autenticación.</p> <p>La tecnología y estándares del SSO se proporcionará al inicio del proyecto al licitador que resulte adjudicatario.</p>
<p>Sobre los datos clínicos y fuentes ETL</p> <p>o ¿Podrían detallar qué tipos de ficheros no estructurados se espera que formen parte de los procesos ETL (por ejemplo, informes clínicos en texto libre, diagnósticos en texto plano, imágenes médicas u otros)?</p> <p>o ¿El Hospital Universitario Ramón y Cajal dispone de algún dataset estructurado o semiestructurado existente que pueda servir como punto de partida para el entrenamiento y validación del modelo de IA? En caso afirmativo, agradeceríamos información sobre su formato, volumen, grado de anonimización y condiciones de acceso.</p>	<p>Como se indica en la página 7 en el punto 5.3. Análisis descriptivo de las fuentes de datos disponibles del PPT: "Se realizará un estudio exhaustivo de las diversas fuentes de datos disponibles en el Hospital Universitario Ramón y Cajal, tanto estructuradas como no estructuradas, para evaluar su utilidad y calidad. Este análisis incluirá la identificación de datos relevantes para la predicción de IAM e IC y la evaluación de su integridad y consistencia." y, por tanto, de ese análisis, tal como se indica en el punto 5.4. Generación de procesos ETL para datos no esgtrurtados, se detallarán los ficheros para los que: "Se desarrollarán y establecerán procesos de Extracción, Transformación y Carga (ETL) que permitan estructurar la información no estructurada existente en las bases de datos del hospital."</p> <p>Hasta la realización del estudio exhaustivo no se podrá disponer de la información solicitada.</p>
<p>Sobre la arquitectura tecnológica de referencia</p> <p>o En relación con las tecnologías indicadas en la arquitectura de referencia (Angular, Spring Boot, Oracle/PostgreSQL, Kubernetes, etc.), ¿existe cierto grado de flexibilidad para proponer tecnologías equivalentes o alternativas (por ejemplo: Python) siempre que se garantice la interoperabilidad, compatibilidad y cumplimiento de los estándares corporativos de la DGSD?</p> <p>o En ese caso, ¿sería necesaria la autorización previa de la DGSD para su validación antes del desarrollo o despliegue?</p>	<p>Sí se puede permitir flexibilidad, pero cualquier cambio o propuesta de nueva tecnología deberá ser validada previamente por la DGSD para poder usarse en los trabajos objeto de la licitación.</p> <p>Como se indica en la página 18 en el punto 6.9 de Herramientas: "El contratista podrá proponer modificaciones en las herramientas para obtener una mejor eficiencia y calidad en el servicio, siempre que se asegure la continuidad de los acuerdos de nivel del servicio. Cualquier petición de cambio tendrá que estar documentada previamente para que la DGSD pueda analizar y autorizar la conveniencia de su implantación."</p>
<p>Sobre la referencia al Reglamento (UE) 2024/1183 en materia de inteligencia artificial.</p> <p>En el Pliego de Prescripciones Técnicas se menciona el Reglamento (UE) 2024/1183 como marco normativo aplicable en materia de Inteligencia Artificial. No obstante, entendemos que dicho texto se refiere al ámbito de la responsabilidad civil derivada del uso de sistemas de IA, mientras que el Reglamento (UE) 2024/1689 (AI Act) es el que establece el marco jurídico armonizado para el desarrollo, despliegue y uso de sistemas de IA en la Unión Europea.</p> <p>¿Podrían confirmar si la referencia correcta dentro del pliego debe entenderse como el Reglamento (UE) 2024/1689, relativo a la Ley de Inteligencia Artificial (AI Act), y si será este el marco de cumplimiento obligatorio para el presente contrato?</p>	<p>Confirmamos que la referencia de este Reglamento está incorrecta y el Reglamento correcto es el mencionado (UE) 2024/1689.</p>

¿La integración con los sistemas HIS del Hospital Ramón y Cajal se realizará mediante APIs preexistentes proporcionadas por la DGSD/hospital, o será responsabilidad del contratista el diseño y desarrollo de estas interfaces de integración?. Adicionalmente, ¿existen servicios de interoperabilidad ya implementados sobre las plataformas corporativas Health Share Health Connect u Onesait Healthcare que podamos utilizar?	Como se indica en la página 7 del pliego, el sistema se integrará con los sistemas de información hospitalaria (HIS) del hospital, la DGSD no proporcionará ni realizará trabajo alguno para dicha integración, siendo una de los trabajos que deberá realizar integralmente el adjudicatario. En caso de usarse alguna de las plataformas corporativas, se podrá utilizar la licencia corporativa de la DGSD, pero cualquier desarrollo necesario debe ser a cargo del adjudicatario.
El pliego menciona "integración con HIS" y "sistema de alertas tempranas" pero es ambiguo sobre si estas interfaces deben ser desarrolladas por el contratista o ya existen. La sección 6.5 del pliego especifica la arquitectura tecnológica obligatoria (Angular 18+ para presentación, Spring Boot 6+ para servicios) lo que sugiere que sí hay desarrollo frontend, pero no especifica si es una aplicación completamente nueva o widgets embebidos en sistemas existentes. ¿El consumo de las predicciones del modelo de IA requiere el desarrollo de una aplicación web completa con frontend Angular 18+ y sistema de alertas independiente, o solo debe integrarse mediante APIs/servicios a interfaces ya existentes en el sistema HIS del hospital? Adicionalmente, ¿las alertas tempranas deben desplegarse como módulo visual autónomo o como componente embebido en las aplicaciones clínicas actuales?	Como se indica en la página 7 del PPT: "El sistema se implementará dentro de los procedimientos y flujos de trabajo del servicio de cardiología del Hospital Universitario Ramón y Cajal. Es decir, se integrará con los sistemas de información hospitalarios (HIS) para que el personal sanitario pueda recibir alertas tempranas y actuar rápidamente en la gestión y tratamiento de los casos detectados." Por lo que no se esperan aplicaciones independientes sino que deben formar parte de los procedimientos y flujos de trabajo, dentro de las aplicaciones clínicas actuales. En conclusión el contratista debe ser responsable y hacerse cargo de cualquier integración o interfaz que sea necesaria para la ejecución del proyecto.
¿Quién será el responsable de proporcionar los datos para el entrenamiento y validación de los modelos: la Comunidad de Madrid, el Hospital Ramón y Cajal u otra entidad?	Como se indica en la página 8 del PPT, "Se desarrollarán y establecerán procesos de Extracción, Transformación y Carga (ETL) que permitan estructurar la información no estructurada existente en las bases de datos del hospital. Esto incluirá la transformación de datos textuales y otros formatos no estructurados en datos estructurados, adecuados para el análisis y el entrenamiento del modelo predictivo". Los datos provendrán de las diversas fuentes de datos disponibles en el Hospital y que deben ser analizadas por el adjudicatario.
¿Podrían detallar, para cada modelo (IAM e IC), el tipo, volumen y estructura de los datos disponibles?	Como se indica en la página 7 del PPT: "Se realizará un estudio exhaustivo de las diversas fuentes de datos disponibles en el Hospital Universitario Ramón y Cajal, tanto estructuradas como no estructuradas, para evaluar su utilidad y calidad. Este análisis incluirá la identificación de datos relevantes para la predicción de IAM e IC y la evaluación de su integridad y consistencia."
¿Qué subtipos o entidades clínicas se espera que aborden cada uno de los modelos (IAM e IC)?	Como se indica en la página 7 del PPT: "Se realizará un estudio exhaustivo de las diversas fuentes de datos disponibles en el Hospital Universitario Ramón y Cajal, tanto estructuradas como no estructuradas, para evaluar su utilidad y calidad. Este análisis incluirá la identificación de datos relevantes para la predicción de IAM e IC y la evaluación de su integridad y consistencia."
¿Se permitirá combinar los datos proporcionados con otros conjuntos externos o propios para mejorar el rendimiento del modelo?	Siempre que se cumpla con la regulación de los datos, tanto en su origen y como en el uso que se de en el resultado podría ser posible, pero se deberá hacer un análisis exhaustivo de todos los aspectos asociados para poder aprobarse el uso de fuentes externas.
¿Existen objetivos mínimos de rendimiento definidos (por ejemplo, AUC, sensibilidad o especificidad esperadas)?	Al tratarse de un piloto no existe un umbral mínimo, pero los resultados esperados deberían ser acordes a los encontrados en pruebas de concepto o pilotos similares en el sector. De los resultados obtenidos se valorará la viabilidad del modelo a futuro.
¿Qué formato y estructura de entrega específica requieren para el código fuente? ¿Qué nivel de documentación técnica requieren que se entregue junto con el código fuente para que sea operable y mantenible por un tercero?	Tanto el formato y estructura de entrega del código fuente, como el nivel de documentación mínimo serán informados por la DGSD tras la adjudicación. En cuanto al nivel de documentación técnica, el licitador podrá hacer su propuesta dentro del plan de transferencia en la memoria técnica.
<ul style="list-style-type: none"> <li>• ¿Están pidiendo una cesión total de los Derechos de Propiedad Intelectual del software, o una licencia exclusiva e irrevocable para su uso? En su caso se refiere a la cesión de la totalidad de los derechos de explotación (reproducción, distribución, modificación, etc.) del software, documentación, estudios y cualquier otro entregable. Por favor confirmar.</li> <li>• En caso de que pidan cesión total de derechos: <ul style="list-style-type: none"> <li>o ¿La cesión de la IP aplica a los elementos de diseño, las bases de datos y la documentación, o solo al código fuente?</li> <li>o Reserva de Derechos sobre Componentes Preexistentes: Es posible que se utilicen componentes de software, infraestructura, modelos de machine learning, procesos, algoritmos de nuestra propiedad, etc que componen la IP y Background propietario desarrollados con anterioridad al presente contrato. Estos componentes pueden ser usados para la eficiencia en el desarrollo. Para proteger la IP preexistente ¿es posible añadir al contrato una cláusula de reserva de derechos excluyendo expresamente de la Cesión de Propiedad Intelectual del Background propietario preexistente</li> </ul> </li> </ul>	<p>Todos los resultados de los servicios realizados para cumplir con el objeto de la licitación son propiedad de la administración pública contratante, en este caso la DGSD. Como se indica en la página 34-35 del PPT: "El contratista acepta expresamente que todos los derechos de propiedad intelectual sobre las configuraciones, parametrizaciones, adaptaciones, implementaciones complementarias, estudios, documentos, productos, subproductos, etc., generados al amparo del presente contrato, corresponden únicamente a la DGSD, con exclusividad y a todos los efectos, quien podrá reproducirlos, publicarlos y divulgarlos, total o parcialmente, sin que pueda oponerse a ello el contratista autor material de los trabajos.</p> <p>Así, podrán ser reutilizados sin coste en cualquier otra implantación en el ámbito del SERMAS o del SNS. No se incluye en el anterior apartado los derechos de uso sobre los productos protegidos con propiedad intelectual y que se adquieran para la puesta en marcha de los sistemas citados como complemento a esta contratación."</p> <p>Si el contratista va a suministrar algún componente sobre los que exista algún tipo de propiedad intelectual, deberá ceder sus derechos para su uso por la DGSD en el SERMAS para poder cumplir con los trabajos. Como se indica en la página 34 del PPT: "El uso de estas herramientas adicionales no puede poner en riesgo la continuidad del servicio después de la finalización de la relación contractual".</p>
<p>Fuentes de datos:</p> <ul style="list-style-type: none"> <li>o ¿Qué datos están disponibles? Imágenes, texto, marcadores?</li> <li>o Los casos están etiquetados?</li> <li>o ¿Que volumen de datos disponibles? De que fechas?</li> <li>o Cómo y dónde está almacenada la información?</li> <li>o Los datos disponibles son datos brutos o son datos de calidad?</li> <li>o ¿Qué formatos?</li> <li>o Cuales son las volumetrías de datos disponibles?</li> </ul>	Como se indica en la página 7 del PPT: "Se realizará un estudio exhaustivo de las diversas fuentes de datos disponibles en el Hospital Universitario Ramón y Cajal, tanto estructuradas como no estructuradas, para evaluar su utilidad y calidad. Este análisis incluirá la identificación de datos relevantes para la predicción de IAM e IC y la evaluación de su integridad y consistencia." Hasta la realización del estudio exhaustivo no se podrá disponer de la información solicitada.

<p>Modelo:</p> <p>o El horizonte de predicción del modelo es cercano (días) o a medio plazo (semanas/meses)</p> <p>o Puede considerarse procesamiento en NRT o también en batch para los modelos a desarrollar?</p>	<p>Uno de los objetivos del modelo es agilizar la toma de decisiones médicas, mejorando los tiempos de respuesta, por lo tanto, la predicción será la más cercana posible. Tal como se indica en la página 7 del PPT, "se creará y entrenará un modelo avanzado de IA que pueda analizar los datos de los pacientes y precedir tempranamente la probabilidad de IAM e IC". Se espera que las predicciones del modelo puedan ser consultadas al consultarse la ficha del paciente, por lo que se desea una solución en tiempo real o como mínimo NRT.</p>
<p>Integración:</p> <p>o Está determinado el número de peticiones al día? Y cuál sería su nivel de concurrencia?</p> <p>o ¿la gestión de despliegues sobre Kubernetes y el mantenimiento del mismo lo hace el SERMAS (o entidad delegada) o el licitador?</p>	<p>No está determinado.</p> <p>Como se indica en la página 27 del PPT, el ingeniero del proyecto tiene que "Gestionar el despliegue, integración y monitorización de los modelos en entornos de producción" e "Implementar prácticas de integración y despliegue continuos (CI/CD)".</p>
<p>Hardware:</p> <p>o El SERMAS va a proporcionar GPUs para desarrollo, entrenamiento y producción o las ha de proporcionar el licitador?</p>	<p>La DGSD proporcionará el hardware que se considere mínimo necesario para la correcta ejecución del proyecto.</p>
<p>Según se recoge en el PCAP, si el contrato tiene por objeto el desarrollo de un producto protegido por propiedad intelectual, éste debe ser cedido a la Administración.</p> <p>Leyendo esa cláusula y la del PPT, entendemos que todo lo que se genere y desarrolle en el marco del contrato es de la Administración, mientras que los derechos de propiedad intelectual preexistentes sobre productos que suministremos no se tienen que ceder a la Administración.</p> <p>Solicitamos que nos confirmen que esta interpretación es correcta, por favor.</p>	<p>Todos los resultados de los servicios realizados para cumplir con el objeto de la licitación son propiedad de la administración pública contratante, en este caso la DGSD. Como se indica en la página 30 del PPT: "El contratista acepta expresamente que todos los derechos de propiedad intelectual sobre las configuraciones, parametrizaciones, adaptaciones, implementaciones complementarias, estudios, documentos, productos, subproductos, etc., generados al amparo del presente contrato, corresponden únicamente a la DGSD, con exclusividad y a todos los efectos, quien podrá reproducirlos, publicarlos y divulgarlos, total o parcialmente, sin que pueda oponerse a ello el contratista autor material de los trabajos.</p> <p>Así, podrán ser reutilizados sin coste en cualquier otra implantación en el ámbito del SERMAS o del SNS.</p> <p>No se incluye en el anterior apartado los derechos de uso sobre los productos protegidos con propiedad intelectual y que se adquieran para la puesta en marcha de los sistemas citados como complemento a esta contratación."</p> <p>Si el contratista va a suministrar algún componente sobre los que exista algún tipo de propiedad intelectual, deberá ceder sus derechos para su uso por la DGSD en el SERMAS para poder cumplir con los trabajos. Como se indica en la página 18 del PPT: "El uso de estas herramientas adicionales no puede poner en riesgo la continuidad del servicio después de la finalización de la relación contractual".</p>
<p>Podrían confirmar el presupuesto base de licitación (sin impuestos)?. En el anuncio aparece 749.999,00 €, pero en el PCAP es diferente: 794.999,00 €</p>	<p>Efectivamente, el importe sin IVA del anuncio es erróneo, la cifra a tener en cuenta es 794,999,00 €</p>
<p>Disponen de herramientas corporativas de gestión del ciclo de vida de proyectos (por ejemplo, Jira, Confluence, etc.) que pueden ser utilizadas para este proyecto, o debemos proporcionar nosotros las herramientas?</p>	<p>Sí, se dispone de la herramienta Jira, al menos para la interlocución con la DGSD. Para su gestión interna podrán usar las herramientas que consideren.</p>
<p>En el Pliego se menciona que el modelo se evaluará en un conjunto de pacientes con monitorización de las constantes hospitalizados a domicilio. ¿Qué tipo de constantes y señales son monitorizadas por protocolo actual?</p>	<p>Las constantes vendrán determinadas por el estudio exhaustivo de las diversas fuentes de datos disponibles en el Hospital Universitario Ramón y Cajal, tanto estructuradas como no estructuradas, para evaluar su utilidad y calidad. Este análisis incluirá la identificación de datos relevantes para la predicción de IAM e IC y la evaluación de su integridad y consistencia. Por lo que las constantes se determinarán según se vea su relevancia en el estudio que hay que realizar.</p>
<p>¿Qué tipos de datos clínicos estarán disponibles para el entrenamiento y validación del modelo? (EHR estructurados, ECG, ecocardiogramas, biomarcadores, datos demográficos, hábitos, etc.)</p>	<p>Como se indica en la página 8 del PPT, "Se desarrollarán y establecerán procesos de Extracción, Transformación y Carga (ETL) que permitan estructurar la información no estructurada existente en las bases de datos del hospital. Esto incluirá la transformación de datos textuales y otros formatos no estructurados en datos estructurados, adecuados para el análisis y el entrenamiento del modelo predictivo". Los datos provendrán de las diversas fuentes de datos disponibles en el Hospital y que deben ser analizadas por el adjudicatario.</p>
<p>¿Los datos estarán anonimizados o pseudoanonimizados antes de su entrega, o será responsale del adjudicatario el proceder con su anonimización?</p>	<p>El adjudicatario es el responsable de realizar cualquier tratamiento que sea necesario realizar sobre los datos.</p>
<p>¿Cuál será la periodicidad de actualización de los datos (batch, streaming, diario)?</p>	<p>Depende de la fuente, que como se ha indicado en la página 7 del PPT: "Se realizará un estudio exhaustivo de las diversas fuentes de datos disponibles en el Hospital Universitario Ramón y Cajal".</p>
<p>¿Qué recursos de hardware estarán disponibles (GPU, CPU, RAM, almacenamiento)? ¿Existen entornos GPU autorizados en la infraestructura corporativa para entrenamiento?</p>	<p>Se dispondrá de los recursos hardware mínimos necesarios para la explotación de los algoritmos. En el momento de esta licitación no se disponen de GPUs lo que no impide que durante la ejecución del proyecto se considere necesario se pudieran emplear.</p>
<p>¿Se permitirá el uso de componentes open-source en los pipelines de IA, o se requiere software licenciado corporativo?</p>	<p>Se puede usar siempre y cuando previamente se haya informado a la DGSD de su uso detallado y esta haya dado su consentimiento.</p>
<p>¿Existen requisitos explícitos de anonimización/pseudonimización para los datos usados en el entrenamiento?</p>	<p>No más de los que pueda exigir la normativa y la regulación de aplicación en cada momento.</p>
<p>¿Debe la solución integrarse en Onesait Healthcare / HealthShare Health Connect, o simplemente interoperar mediante APIs?</p>	<p>Si es necesario el uso de este tipo de soluciones se deberán usar las indicadas en los pliegos, en caso contrario, siempre que la solución sea acorde a la arquitectura propuesta y no se uso ninguna solución que no esté autorizada por la DGSD no hay inconveniente.</p>
<p>Se espera que el contratista diseñe y construya ETL/ELT para datos no estructurados (5.4) ¿Se prevee que también sean necesarias pipelines para datos estructurados?</p>	<p>Dependerá de los resultados de la actividad de análisis de las fuentes disponibles.</p>
<p>¿LA tecnología del DataLakehouse actual es cloudera? ¿Onesait Healthcare?¿qué rol o funcionalidad tiene cada uno en el global de la arquitectura? ¿Con qué tecnología se realizan actualmente las ETL/ELT?</p>	<p>El datalake es cloudera. La suite Onesait Healthcare es una plataforma corporativas para el desarrollo e integración de aplicaciones con el estándar HL7, FHIR e IHE.</p>

Entendemos que la respuesta definitiva saldrá del análisis inicial de datos necesarios. Sin embargo, de cara a la estimación de costes, nos ayudaría saber cuántos orígenes de datos se prevé integrar en la plataforma (aproximadamente)?	No hay una previsión, todo dependerá del valor que durante el estudio se pueda dar o no a una fuente de datos.
¿Qué mecanismos de acceso o exportación estarán disponibles para las fuentes clínicas (API FHIR, SQL, HL7, ficheros, etc.)?	Dependerá del proveedor de la fuente pueden usarse unas u otras, el proveedor deberá adaptarse a la que este dispnible en cada caso.
¿Qué sistemas de información hospitalarios están disponibles para la extracción de datos (HIS, EHR, PACS, LIS, etc.)?	Todas las fuentes de información disponibles en el Hospital Universitario Ramón y Cajal estarán disponibles para su análisis. Como se indica en la página 7 del PPT: "Se realizará un estudio exhaustivo de las diversas fuentes de datos disponibles en el Hospital Universitario Ramón y Cajal, tanto estructuradas como no estructuradas, para evaluar su utilidad y calidad. Este análisis incluirá la identificación de datos relevantes para la predicción de IAM e IC y la evaluación de su integridad y consistencia."
¿Qué datos necesarios para el entrenamiento del modelo ya se encuentran cargados en el Data Lake?	Hasta la realización del estudio de las fuentes, no se sabe cuáles son los datos necesarios y si están o no disponibles en el datalake.
¿Existen restricciones o procesos de consolidación en las fuentes de datos? Si es así, ¿en qué momento se encuentran disponibles para su extracción?	Con la realización del estudio exhaustivo se deberá abordar estos aspectos en las fuentes.