DESAFÍO UNIVERSIDAD - EMPRESA

Esta necesidad tecnológica forma parte del Concurso de Proyectos de I+D+i y/o consultoría en colaboración Universidad – Empresa "Desafío Universidad Empresa" 2026 organizado por la Fundación Universidades y Enseñanzas Superiores de Castilla y León.

TÍTULO DE LA DEMANDA TECNOLÓGICA A RESOLVER

Referencia:

NT80

Título de la demanda tecnológica propuesta

Plataforma completa para crear y evaluar cohortes virtuales de pacientes oncológicos

Acrónimo:

COHORTEV

Áreas de interés de la demanda tecnológica

(Principal) Salud y Atención social
Tecnologías y ciberseguridad

Resumen:

Plataforma completa para crear, evaluar y exportar cohortes virtuales de pacientes oncológicos. El equipo elegirá una patología entre cáncer de colon, páncreas, mama o pulmón, justificando datos, modelo y validación. Deberá generar perfiles coherentes y poblaciones parametrizadas para simulación sin depender de bases hospitalarias privadas. La investigación abordará relaciones entre variables, diversidad, restricciones biológicas e incertidumbre. La entrega incluirá interfaz, generación reproducible, evaluación por subgrupos, trazabilidad, exportación, instalación, código, documentación, pruebas y formación, con derechos acordados y límites de validez explícitos.

PALABRAS CLAVE: Cohortes virtuales, Oncología, Simulación tumoral, Incertidumbre, Validación.

DESCRIPCIÓN DE LA NECESIDAD DEMANDADA

1.- Descripción de la demanda tecnológica.

Se necesita disponer de una plataforma de investigación operativa que permita a personal biomédico definir una población oncológica objetivo, generar cohortes virtuales y examinar qué características y limitaciones tienen antes de utilizarlas. El resultado debe cubrir el flujo completo: configurar, importar referencias, generar, comprobar, comparar, guardar, reproducir y exportar. No bastará una mejora algorítmica aislada.

Una cohorte virtual será un conjunto de individuos computacionales con atributos y procedencia definidos. Se distinguirán los registros sintéticos que reproducen distribuciones de datos y los pacientes parametrizados que producen trayectorias mediante un modelo matemático. No se identificarán automáticamente con gemelos digitales de pacientes reales.

La dificultad investigadora consiste en satisfacer simultáneamente restricciones biológicas oncológicas, dependencias multivariantes, diversidad y ajuste a referencias incompletas, sin ocultar incertidumbre ni eliminar subgrupos al filtrar. Se comparará el método propuesto con líneas base sencillas y componentes reutilizables. Se investigará en qué condiciones mejora y cuándo debe declarar que no puede satisfacer la población solicitada.

La entidad no aporta bases hospitalarias privadas. El mínimo entregable no dependerá de reclutar pacientes, obtener nuevos acuerdos sanitarios ni entrenar un modelo fundacional. Se utilizarán datos públicos con condiciones de uso comprobadas, estadísticas publicadas y modelos reproducibles. Las correlaciones o parámetros no identificables se declararán como supuestos y se analizarán mediante escenarios.

Se entregará una plataforma utilizable, con: (a) interfaz gráfica para definir criterios de inclusión, tamaño, variables, restricciones y escenarios; (b) importación de datos tabulares y referencias con diccionario, unidades y procedencia; (c) generación reproducible de cohortes y, para un modelo dinámico acotado, conjuntos de parámetros y trayectorias; (d) evaluación de plausibilidad, diversidad, relaciones y diferencias por subgrupos; (e) registro de configuración, semillas, versiones, tasas y causas de rechazo, incertidumbre y advertencias; (f) gestión de ejecuciones y exportación de datos, parámetros e informes en formatos abiertos; (g) instalación reproducible, documentación de usuario y técnica, código fuente y dependencias con sus licencias, pruebas automatizadas y formación.

El equipo podrá elegir UNA patología entre cáncer de colon, páncreas, mama o pulmón y justificará subtipo histológico, estadio, población de uso y viabilidad. Ambos demostradores se referirán a esa elección: perfiles clínicos oncológicos de una fuente pública y una población de parámetros con trayectorias de un modelo tumoral acotado. No se exigirá desarrollar las cuatro patologías. Se proponen datos abiertos de GDC/TCGA o estudios de cBioPortal con reutilización permitida; el modelo podrá reutilizarse con licencia compatible o implementarse desde ecuaciones documentadas. No se vincularán perfiles y parámetros sin evidencia. Se declarará qué referencias permiten evaluar las trayectorias y cuáles solo prueban funcionamiento.

2.- Antecedentes.

En oncología, Braniff et al. (10.1002/psp4.13270) estudian calibración conjunta de dinámica tumoral y supervivencia libre de progresión, considerando censura. Los avances metodológicos de 2025–2026 incluyen muestreo adaptativo de poblaciones mecanísticas, aceleración mediante modelos sustitutos y generación de datos con difusión. Schirru et al. (2025; DOI: 10.1007/s11538-025-01532-z) estudian cobertura del espacio paramétrico; Iwata y Saito (2025; 10.1002/psp4.13288), eficiencia de generación. Gadgil et al. (2026; 10.1007/s10928-026-10054-7) combinan covariables y demostraciones farmacocinéticas, con límites de generalización.

La semejanza estadística no garantiza coherencia temporal ni utilidad en subgrupos: Perkonoja et al. (2026; 10.1016/j.jbi.2026.105074) examinan preservación temporal. Masud y Hasan (2026; 10.2196/90590) muestran límites del filtrado de muestras inválidas. Synthea ya genera historias sintéticas y exporta formatos sanitarios; estas capacidades existentes deberán aprovecharse o compararse.

3.- Posibles enfoques del proyecto de investigación.

Se admiten métodos estadísticos, muestreo condicionado, inferencia bayesiana, modelos mecanísticos, aprendizaje generativo y combinaciones justificadas. No se impone una familia algorítmica. Se evaluarán variables y restricciones pertinentes al tumor elegido, separando carga tumoral, progresión, supervivencia y censura. Interesa estudiar la calibración con información limitada, la conservación de dependencias y diversidad, la selección sin sesgos inadvertidos y la detección de objetivos inviables.

Se valorará reutilizar componentes compatibles y evaluar las mejoras con referencias reservadas, repetición de semillas, análisis de sensibilidad, escenarios de ausencia de datos y revisión biomédica documentada. Los demostradores tendrán dimensiones y costes computacionales delimitados. Se priorizará funcionamiento local y bajo coste recurrente, formatos abiertos y mantenimiento asumible. La interoperabilidad sanitaria avanzada podrá ampliarse una vez garantizado el flujo mínimo completo.

4.- Enfoques sin interés.

Revisiones bibliográficas como único resultado; cuadernos o algoritmos sin plataforma instalada; interfaces que solo muestren datos pregenerados; dependencias de bases privadas o permisos futuros para el mínimo; eliminación silenciosa de subgrupos; validación exclusivamente circular presentada como evidencia clínica; prescripción individual, sustitución de ensayos clínicos o garantías de anonimato basadas solo en la etiqueta «sintético». No interesa una entrega que omita código, documentación, pruebas, instalación o derechos de uso y modificación acordados.

Si desea remitir una propuesta de solución tecnológica** (proyecto de investigación y/o consultoría) **deberá enviar el formulario de participación** (ANEXO II), descargable en **www.redtcue.es/desafio** **a una de las direcciones de correo electrónico que se indican en las bases del concurso antes del 18/12/2026

Acceso a información general del concurso