

**PLIEGO DE PRESCRIPCIONES TÉCNICAS PARA LA CONTRATACIÓN DEL
SERVICIO “CONVERSIÓN A OPOP DE LAS BASES DE DATOS
PROVENIENTES DE LOS DISTINTOS BIOBANCOS DE LA PLATAFORMA ISCIII
BIOMODELOS Y BIOBANCOS.”**

Exp. 10/2024



MINISTERIO
DE CIENCIA
E INNOVACIÓN



Instituto
de Salud
Carlos III



UNIÓN EUROPEA

Fondo Europeo de Desarrollo Regional
“Una manera de hacer Europa”

1. OBJETO

El objeto del presente documento es el establecimiento de las prescripciones técnicas que rigen en el procedimiento de contratación destinado a ampliar las prestaciones de servicios de la Plataforma de Biomodelos y Biobancos del Instituto de Salud Carlos III (de ahora en adelante, "P_ISCIII_BB"), mediante un proyecto de conversión a OMOP CDM [del inglés- *Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM)*] de las bases de datos provenientes de los distintos biobancos y unidades que conforman la P_ISCIII_BB , que mejorará la competitividad de la investigación biomédica española, con el fin acelerar la transición del conocimiento a la práctica clínica y cimentar el posicionamiento de la P_ISCIII_BB en el ámbito nacional e internacional de la investigación biomédica, facilitando la labor de todos los investigadores de la organización, haciendo que su trabajo sea más fácil y efectivo y permitiendo, sobre todo, la interoperabilidad de los datos dentro y fuera de la P_ISCIII_BB.

La adquisición englobará las prestaciones de suministro, realización del servicio y formación.

2. LUGAR DE ENTREGA

Fundació Institut de Bioenginyeria de Catalunya
C/ Baldiri Reixac, 15-21
Planta 3 – Lab P3A10

08028 Barcelona

3. PARTES Y DESARROLLO DEL SERVICIO

El alcance requiere en primera instancia, por parte de la empresa adjudicataria, de un periodo inicial destinado al perfilamiento de los datos y la calidad y un segundo periodo destinado a la conversión a OMOP.

3.1. Análisis de los datos por expertos

3.2. Control de Calidad

3.3. Desarrollo y conversión

3.3.1 Metodología *OMOP Agile conversion*, esta metodología se basa en *sprints* de 2 semanas y da soporte a la adaptación y mejora continua durante el proceso de conversión. Durante el *sprint* se llevan a cabo los siguientes pasos:

1. Revisión de las especificaciones de extracción, transformación y carga (ETL)
2. Desarrollo de la documentación y el proceso específico ETL
3. Ejecución del ETL
4. Validación



4. ESPECIFICACIONES TÉCNICAS DEL SERVICIO. ENTREGABLES Y DESCRIPCIÓN DE LOS SERVICIOS INCLUIDOS

A continuación, se detallan los entregables y la descripción de los servicios incluidos:

Descripción de los entregables de empresa adjudicataria para el proyecto:

- Especificaciones de mapeo ETL
- Documento ETL "*Run Process*" con pasos para ejecutar código de conversión y cargar la instancia de CDM (*Common Data Model*) de destino
- Scripts ETL: a fin de extraer, transformar y convertir datos en OMOP CDM
- Asignaciones personalizadas de datos de origen no estándar a conceptos estándar de OMOP (en caso sea necesario)
- *Scripts* de prueba unitaria para realizar pruebas y test de CDM
- Estadísticas de conversión, calidad de datos e informes de pruebas

Descripción de los servicios incluidos en el proyecto:

- Análisis de requerimientos de la P_ISCIII_BB
- Análisis de datos y evaluación de la calidad de los datos
- Especificaciones ETL: para crear/actualizar y código ETL
- Implementación de código ETL para rellenar instancias de CDM
- Planificación y ejecución de QA/QC, incluida la validación y certificación del esquema CDM
- Implementación de producción, por ejemplo, instancia de producción de CDM poblada con datos convertidos

Los entregables y servicios descritos están incluidos para una conversión a OMOP de los datos especificados por la P_ISCIII_BB y de acuerdo al alcance especificado.

El alcance especificado requiere en primera instancia, por parte de la empresa adjudicataria, de un periodo inicial destinado al perfilamiento de los datos y la calidad y un segundo periodo destinado a la conversión a OMOP

5. EMBALAJE Y TRANSPORTE

No requiere

6. CALENDARIO DE IMPLANTACIÓN Y FORMACIÓN

La empresa adjudicataria estará obligada a suministrar el servicio descrito anteriormente, así como los medios humanos y materiales necesarios para llevar a cabo su ejecución, en un plazo máximo de 6 meses desde la formalización del contrato.

6.1. PERIODO 1: EVALUACIÓN INICIAL DE LOS DATOS.

Durante el *periodo 1* se llevará a cabo el perfilado de los datos de origen, se prepararán las cargas, se crearán los esquemas y tablas y se cargarán los archivos de vocabulario OHDSI. Se detallan aquí los distintos procesos:

- Source Data Profiling para determinar las necesidades de ingestión de datos de origen, mediante, *a priori*:
 - Identificación de ontologías médicas controladas y códigos
 - Estimación de mapeo de conceptos estándar OMOP
 - Estimación mapeo de conceptos personalizados OMOP
 - Recomendación para el producto mínimo viable del modelo de datos OMOP CDM (modelo de datos común OMOP)
- Carga de datos de origen e inicialización de la base de datos
- Creación de esquemas y tablas
- Carga de archivos de vocabulario OHDSI
- Recomendaciones para alojamiento del modelo OMOP CDM

6.2. PERIODO 2: CONVERSION A OMOP

- Especificaciones iterativas de ETL: crear/actualizar código ETL
- Despliegue del código ETL para poblar instancias del Modelo de Datos Común OMOP (OMOP CDM)
- Planificación y ejecución de control de calidad (QA/QC), incluyendo validación y certificación de esquemas del CDM
- Despliegue en producción del Modelo de Datos Común OMOP (OMOP CDM) final
- Especificaciones de mapeo ETL
- Documento de "Proceso de Ejecución" de ETL con pasos para ejecutar el código de conversión y cargar las instancias de destino del Modelo de Datos Común OMOP (OMOP CDM)
- Scripts de ETL: extraer, transformar y convertir datos de origen en OMOP CDM o Mapeo personalizado de datos de origen no estándar a conceptos estándar de OMOP



- Scripts de pruebas unitarias o Estadísticas de conversión e informes de calidad de datos

De acuerdo con el alcance especificado, fuentes de datos, los distintos formatos y la complejidad de los sets de datos, el servicio se realizará en 6 meses. El plazo se computará a partir del día siguiente al de la formalización del contrato, excepto otra notificación por parte del IBEC. En el transcurso de estos 6 meses se llevará a cabo un minucioso análisis de la estructura y contenido de cada fuente de datos, se identificarán anomalías y se diseñarán los mapeos de código a los estándares para llevar a cabo la conversión a OMOP CDM.

Se impartirá una formación a los potenciales usuarios del sistema con una duración mínima de 2 horas en una fecha a convenir, con el fin de instruirlos en OMOP como modelo de datos común. Dicha formación irá a cargo de la empresa adjudicataria y se realizará de forma telemática. Las sesiones de formación se grabarán y la empresa adjudicataria presentará así informes en formato de vídeo y documentación (formato .doc o .pdf) que recogerá las sesiones de formación impartidas en las diferentes sesiones. Los informes compilarán información para posterior revisión por parte de los biobancos y unidades de la P_ISCIII_BB.

Una vez finalizado el servicio contratado y comprobada su adecuación a las previsiones del presente Pliego de Prescripciones Técnicas se levantará la correspondiente acta de recepción, momento a partir del cual empezará a contar el periodo de garantía.

8. GARANTÍA MÍNIMA Y SERVICIO TÉCNICO POSTVENTA

La empresa adjudicataria garantizará la conversión a OMOP de data set centralizado que proviene de todos los biobancos y unidades de biomodelos de la P_ISCIII_BB, que a fecha 28 de junio de 2024 son 64, a título informativo, sin perjuicio de que esta cifra pueda estar sujeta a cambios. La conversión requerirá una exhaustiva revisión de cada uno de los data sets, asegurando el contenido adecuado y el mapeo de vocabulario necesario que deberá formar parte del set de datos centralizado de todos los biobancos y unidades de biomodelos.

Los plazos de garantía mínima indicados podrán ser objeto de ampliación según lo previsto en el Anexo 4 del Pliego de Cláusulas Particulares.

Barcelona, a 28 de junio de 2024

Prof. Núria Montserrat Pulido

Coordinadora de la Plataforma de Biomodelos y Biobancos del Instituto de Salud Carlos III.

