OMICSPACE

Espacio de datos –ómicos ético y seguro

BDIAB

IIS La Fe

Plataforma de Big Data, IA, Bioestadística & Bioinformática

CATÁLOGO DE SERVICIOS



Participantes:

IIS La Fe · Fundación Pública Andaluza Progreso y Salud · Hospital del Mar Barcelona · Hospital Universitario La Paz · Hospital Universitario Ramón y Cajal · iacs · Medicina Xenómica · HARMONY Alliance Foundation

OMICSPACE

Espacio de datos –ómicos ético y seguro

CATÁLOGO DE SERVICIOS



ACCESO y PREPARACIÓN DE DATOS

BDIAB

IIS La Fe

Plataforma de Big Data, IA, Bioestadística & Bioinformática

Participantes:

IIS La Fe · Fundación Pública Andaluza Progreso y Salud · Hospital del Mar Barcelona · Hospital Universitario La Paz · Hospital Universitario Ramón y Cajal · iacs · Medicina Xenómica · HARMONY ALLIANCE FOUNDATION

Contenido

Acceso y Preparación de Datos 3
PTBG-IA-ES06 Extracción variables estructuradas 3
PTBG-IA-ES07 Extracción de variables semiestructuradas o no estructuradas 4
PTBG-IA-ES52a Extracción de una cohorte de pacientes a partir de dato estructurado: TFM o TFGs 5
PTBG-IA-ES52b Extracción de una cohorte de pacientes a partir de dato estructurado 6
PTBG-IA-ES52c Extracción de una cohorte de pacientes a partir de datos estructurado: complejidad media 8
PTBG-IA-ES52d Extracción de una cohorte de pacientes a partir de dato estructurado: complejidad alta 9
PTBG-IA-ES34 Identificación de cohortes digitales (estandarizadas a OMOP) 11
PTBG-IA-ES35 Caracterización de cohortes digitales (formato OMOP) 12
PTBG-IA-ES08 Preparación de datos 13
PTBG-IA-ES55 Uso autónomo de la plataforma BIG DATA por personal con vinculación laboral con el H.U. i P. La Fe o el IIS La Fe 14
PTBG-IA-ES49 Feasability 15
PTBG-IA-ES53 Preparación entorno de trabajo 15
PTBG-IA-ES64 Estandarización de datos 17
PTBG-IA-ES65 · Creación de catálogo de datos 17

Acceso y Preparación de Datos

Esta categoría agrupa servicios que permiten extraer, transformar y preparar información clínica para su uso en investigación, análisis de resultados en salud y desarrollo de modelos. Nuestro objetivo es garantizar que los datos procedentes de sistemas sanitarios se conviertan en datasets estructurados, coherentes y trazables, cumpliendo criterios de calidad y normativa.

Los datos se extraen de la Plataforma de Big Data e IA del IIS La Fe, que se apoya en un Data Warehouse (DW) robusto que integra información clínica, epidemiológica y biológica de más de 1,5 millones de pacientes, con una cohorte poblacional activa de 300.000 personas.

El DW está compuesto por 22 datamarts, 84 tablas, 4.064 columnas y más de 750 millones de registros estructurados, además de 30 millones de registros no estructurados, actualizados de forma continua. Esta información está disponible tanto en formato local como en formato estandarizado OMOP.

PTBG-IA-ES06 Extracción variables estructuradas

Descripción del servicio

Extracción, a partir de los registros almacenados en la Plataforma de Big Data e IA del IIS La Fe, de variables estructuradas (diagnósticos, procedimientos, farmacia/prescripción, visitas, episodios etc) según los requisitos del solicitante, con generación de un dataset listo para su uso en fases posteriores de análisis.

Alcance del servicio

- Definición del caso de uso: objetivos del análisis, criterios de inclusión y exclusión, variables de interés, ventanas temporales y formato de salida establecidos con el solicitante.
- Identificación y documentación de las fuentes: determinación de tablas y dominios clínicos relevantes, verificación de su estructura y aplicación de terminologías normalizadas (ICD-10-ES, SNOMED CT, LOINC, ATC) para garantizar consistencia semántica.
- Diseño y ejecución de la consulta: construcción de consultas sobre la base de datos relacional aplicando los criterios definidos; filtrado por periodos temporales; selección de episodios clínicos relevantes; eliminación de duplicidades y vinculación correcta de identificadores entre tablas, garantizando que la cohorte resultante sea consistente, íntegra y trazable.

- Control de calidad: comprobaciones de completitud, consistencia y coherencia clínica; detección de valores fuera de rango; registro completo del proceso en logs para asegurar reproducibilidad.
- Documentación técnica: elaboración de diccionario de variables e informe con el proceso seguido, los criterios aplicados y las verificaciones realizadas.

No incluye

- Estandarización a modelos OMOP o FHIR y mapeo terminológico avanzado; véase PTBG-IA-ES64.
- Análisis estadístico, cuadros de mando o modelización; véase el bloque Análisis y Modelización.
- Rediseños mayores del procedimiento de extracción y preparación, con nuevas reglas de filtrado, mapeo y vinculación entre tablas, o múltiples ciclos de iteración tras la validación final.

PTBG-IA-ES07 Extracción de variables semiestructuradas o no estructuradas

Descripción del servicio

Servicio basado en procesamiento de lenguaje natural para identificar y estructurar información relevante contenida en notas clínicas, como informes de anatomía patológica o radiológicos. Orientado a convertir texto libre en variables útiles y trazables para investigación, estudios RWD/RWE y desarrollo de modelos.

Alcance del servicio

- Definición de requisitos: objetivos, criterios de inclusión y exclusión, variables, fuentes, periodo y formato de salida definidos con el solicitante.
- Preprocesamiento y segmentación del texto: limpieza, normalización y organización por secciones para facilitar la extracción.
- Reconocimiento de entidades clínicas: identificación de diagnósticos, hallazgos, procedimientos, biomarcadores, tratamientos y eventos mediante NER.
- Detección de negaciones y referencias temporales: análisis del texto para identificar negaciones, incertidumbre y temporalidad.
- Captura de atributos cuantitativos: extracción de valores numéricos con unidades y rangos.
- Normalización terminológica: aplicación de estándares como SNOMED CT, ICD-10-ES, LOINC y ATC/RxNorm.

- Generación del dataset estructurado: creación del conjunto de datos en CSV, Parquet o SQL, con diccionario y logs.
- Control de calidad: muestreo, métricas (precisión, recall, F1) e informe con ejemplos y recomendaciones.

No incluye

- Estandarización a modelos OMOP o FHIR y mapeo terminológico avanzado; véase PTBG-IA-ES64.
- Análisis estadístico, cuadros de mando o modelización; véase el bloque Análisis y Modelización.
- Integración en historia clínica electrónica o entornos de producción, acceso directo a sistemas, iteraciones múltiples o cambios de alcance.
- Rediseños mayores del proceso de extracción a partir de texto, con nuevas entidades, reglas de negación y temporalidad o cambios sustanciales de normalización terminológica, o múltiples ciclos de iteración tras la validación final.

PTBG-IA-ES52a Extracción de una cohorte de pacientes a partir de dato estructurado: TFM o TFGs

Descripción del servicio

Extracción de una cohorte básica desde sistemas clínicos estructurados para trabajos académicos (Trabajo Fin de Grado o Fin de Máster), a partir de los criterios proporcionados por el solicitante. Enfoque centrado en identificación y caracterización de la cohorte, con dataset trazable y listo para uso académico y cumplimiento de la normativa aplicable.

Alcance del servicio

- Identificación de la cohorte según los criterios proporcionados, con documentación de la especificación de la extracción y del formato de salida.
- Identificación de las fuentes de datos y variables de interés: localización de sistemas y repositorios relevantes, tablas y campos que contienen la información requerida y verificación de codificación y terminología empleada (ICD-10-ES, SNOMED CT, ATC o DDD, LOINC) para asegurar coherencia y trazabilidad.
- Construcción y ejecución de la extracción: acotación del periodo de observación y de la unidad de análisis por episodio o visita; depuración de duplicados; armonización de claves internas; reconciliación de identificadores de persona, episodio y visita entre fuentes; generación de resúmenes descriptivos esenciales con conteos, primeras y últimas fechas y duraciones.

- Generación de la cohorte: listado seudonimizado de pacientes o tabla de episodios asociada; entrega del diccionario de variables y del registro detallado de la extracción.
- Caracterización y control de calidad: descripción de características básicas de la cohorte, incluida la distribución por sexo y edad; comprobaciones de completitud, consistencia e integridad referencial; informe conciso con resultados y recomendaciones.

No incluye

- Extracción de texto no estructurado ni procesamiento de lenguaje natural; véase PTBG-IA-ES07.
- Estandarización a modelos OMOP o FHIR y mapeo terminológico avanzado; véase PTBG-IA-ES64.
- Análisis estadístico, cuadros de mando o modelización; véase el bloque Análisis y Modelización.
- Integración en historia clínica electrónica o entornos de producción, acceso directo a sistemas, iteraciones múltiples o cambios de alcance.
- Cohortes múltiples, comparativas entre grupos o uso multicéntrico; véase PTBG-IA-ES52b–d.
- Rediseños mayores de la definición de cohorte y de la lógica temporal, con nuevas reglas, secuencias y ventanas, o múltiples ciclos de iteración tras la validación final.

PTBG-IA-ES52b Extracción de una cohorte de pacientes a partir de dato estructurado

Descripción del servicio

Definición y extracción de una cohorte desde datos clínicos estructurados mediante criterios simples, con alcance limitado a hasta tres dominios y sin lógica temporal compleja. Adecuado para la obtención rápida de una cohorte con criterios sencillos cuando se requiere un dataset inicial para exploración descriptiva y el alcance se limita a hasta tres dominios sin ventanas temporales complejas ni reglas de eventos avanzadas.

Alcance del servicio

- Identificación de la cohorte según los criterios proporcionados, con documentación de la especificación de la extracción y del formato de salida.
- Identificación de las fuentes de datos y de las variables de interés: localización de sistemas y repositorios relevantes, tablas y campos que contienen la información
-

- requerida dentro de hasta tres dominios (por ejemplo, diagnósticos, procedimientos, farmacia).
- Verificación terminológica y de codificación: comprobación de coherencia y trazabilidad de la codificación y la terminología empleadas (ICD-10-ES, SNOMED CT, ATC, DDD, LOINC).
- Diseño y ejecución de la extracción: delimitación del periodo de observación y de la unidad de análisis por episodio o por visita; depuración de duplicados; armonización de claves internas; reconciliación de identificadores de persona, episodio y visita entre fuentes; generación de resúmenes descriptivos esenciales con conteos, primeras y últimas fechas y duraciones.
- Generación del dataset y documentación: listado seudonimizado de pacientes o tabla de episodios asociada; entrega del diccionario de variables y del registro detallado de la extracción.
- Caracterización y control de calidad básico: descripción de características fundamentales de la cohorte (incluida la distribución por sexo y edad) y comprobaciones de completitud, consistencia e integridad referencial, con informe conciso de resultados y recomendaciones.

No incluye

- Extracción de texto no estructurado ni procesamiento de lenguaje natural; véase PTBG-IA-ES07.
- Estandarización a modelos OMOP o FHIR y mapeo terminológico avanzado; véase PTBG-IA-ES64.
- Análisis estadístico, cuadros de mando o modelización; véase el bloque Análisis y Modelización.
- Comparativas entre cohortes, uso multicéntrico o lógica temporal avanzada; véase PTBG-IA-ES52c–d.
- Acceso directo a sistemas, integración en historia clínica electrónica, iteraciones múltiples o cambios de alcance.
- Rediseños mayores de la definición de cohorte y de la lógica temporal, con nuevas reglas, secuencias y ventanas, o múltiples ciclos de iteración tras la validación final.

PTBG-IA-ES52c Extracción de una cohorte de pacientes a partir de datos estructurado: complejidad media

Descripción del servicio

Extracción de una cohorte desde datos clínicos estructurados aplicando criterios de complejidad intermedia. Se incluyen múltiples dominios, reglas combinadas y lógica temporal básica a intermedia (ventanas temporales, periodos de retrospectiva, secuencias simples de eventos y agrupación de episodios). Adecuado cuando se requiere una cohorte con reglas combinadas y ventanas temporales sencillas, que abarque varios dominios y precise una caracterización descriptiva más completa que la de un alcance básico.

Alcance del servicio

- Identificación de la cohorte según los criterios proporcionados, con criterios combinados y jerárquicos definidos por el solicitante y documentación de la especificación de la extracción y del formato de salida.
- Identificación de fuentes de datos y variables de interés: localización de sistemas y repositorios relevantes, tablas y campos necesarios, con hasta seis dominios (diagnósticos, procedimientos, farmacia, laboratorio, hospitalizaciones, urgencias).
- Verificación terminológica y de codificación: comprobación de coherencia y trazabilidad de terminologías y códigos empleados (ICD-10-ES, SNOMED CT, ATC, DDD, LOINC, OPS, UCUM).
- Diseño y ejecución de la extracción: delimitación del periodo de observación y de la unidad de análisis; aplicación de reglas temporales básicas e intermedias (A antes de B, A dentro de X días tras B); ventanas de exposición sencillas; periodos de retrospectiva de 6 a 12 meses; construcción de evento índice; episodización y consolidación simple de estancias; combinación de dominios mediante uniones e intersecciones; depuración de duplicados; armonización de claves internas; reconciliación de identificadores de persona, episodio y visita entre fuentes; generación de resúmenes descriptivos ampliados con conteos, primeras y últimas fechas y duraciones.
- Generación del dataset y documentación: listado seudonimizado de pacientes o tabla de episodios asociada; entrega del diccionario de variables y del registro detallado de la extracción.
- Caracterización y control de calidad: descripción de características por dominios relevantes (incluida la distribución por sexo y edad); verificación de completitud, consistencia temporal, coherencia clínica e integridad referencial; entrega de informe técnico con criterios aplicados y recomendaciones.

No incluye

- Extracción de texto no estructurado ni procesamiento de lenguaje natural; véase PTBG-IA-ES07.
- Estandarización a modelos OMOP o FHIR y mapeo terminológico avanzado; véase PTBG-IA-ES64.
- Análisis estadístico, cuadros de mando o modelización; véase el bloque Análisis y Modelización.

- Comparativas multicentro o armonización entre distintas bases de datos; véase PTBG-IA-ES52d.
- Rediseños mayores de la definición de cohorte y de la lógica temporal, con nuevas reglas, secuencias y ventanas, o múltiples ciclos de iteración tras la validación final.

PTBG-IA-ES52d Extracción de una cohorte de pacientes a partir de dato estructurado: complejidad alta

Descripción del servicio

Extracción de cohortes complejas desde datos estructurados y notas clínicas no estructuradas, combinando múltiples dominios, reglas avanzadas y lógica temporal detallada. Se incluyen secuencias de eventos, episodización avanzada, integración de fuentes heterogéneas y procesamiento de notas clínicas mediante NLP con extracción y normalización de conceptos, gestión de negación e incertidumbre y reconocimiento del contexto temporal. Orientado a estudios multicéntricos, definiciones incidentes con ventanas estrictas y escenarios que requieren información no disponible en estructuras codificadas. Indicado para definiciones complejas con secuencias estrictas, ventanas temporales y episodización avanzada, con necesidad de integrar texto libre y múltiples sistemas, y con requisitos de documentación y trazabilidad exhaustivos, especialmente en estudios multicéntricos o regulatorios.

Alcance del servicio

- Identificación de la cohorte según criterios complejos definidos por el solicitante, con secuencias, ventanas, jerarquías y definiciones clínicas avanzadas; documentación de la especificación de la extracción y del formato de salida.
- Identificación de fuentes de datos y variables de interés: mapeo de sistemas y repositorios de múltiples orígenes; localización de tablas y campos necesarios en seis o más dominios (urgencias, hospitalización, laboratorio, farmacia,
-

- procedimientos, cuidados intensivos, atención primaria, constantes vitales, histórico clínico).
- Verificación terminológica y de codificación: uso intensivo y verificación de ICD-10-ES, SNOMED CT, ATC, DDD, LOINC, OPS, UCUM y CIE-9 para históricos cuando proceda.
- Procesamiento de notas clínicas mediante NLP: extracción de entidades clínicas (diagnósticos, síntomas, procedimientos, tratamientos); detección de negación, afirmación e incertidumbre; normalización de conceptos a terminologías de referencia; reconocimiento de temporalidad con distinción entre situación actual, antecedentes y planificación; vinculación de contexto con paciente, profesional y episodio.
- Lógica temporal y construcción avanzada: definición de eventos índice multinivel; construcción de secuencias estrictas de eventos (A dentro de X días tras B y antes de C); establecimiento de ventanas de exposición y periodos de riesgo; periodos de limpieza washout y periodos de entrada para cohortes incidentes; episodización avanzada, consolidación de estancias y modelado de trayectorias clínicas longitudinales.
- Integración y armonización avanzada: integración de fuentes heterogéneas y armonización de identificadores entre sistemas, con reconciliación a nivel de persona, episodio y visita y documentación de la estrategia de enlace y la trazabilidad.
- Caracterización y control de calidad ampliados: validación de coherencia temporal, consistencia entre dominios y control de episodización; validación clínica dirigida y validación manual de subconjuntos NLP cuando corresponda; informe completo de calidad y trazabilidad del algoritmo; diccionario de variables; repositorio de criterios y documentación reproducible.
- Generación del dataset y documentación: entrega del conjunto de datos final seudonimizado con sus tablas asociadas y del registro detallado de la extracción y de las transformaciones aplicadas.

No incluye

- Extracción de texto no estructurado ni procesamiento de lenguaje natural; véase PTBG-IA-ES07.
- Estandarización a modelos OMOP o FHIR y mapeo terminológico avanzado; véase PTBG-IA-ES64.
- Análisis estadístico, cuadros de mando o modelización; véase el bloque Análisis y Modelización.
- Armonización multicentro avanzada con mapeo interinstitucional específico; solicitar servicio dedicado.

- Rediseños mayores de la definición de cohorte y de la lógica temporal, con nuevas reglas, secuencias y ventanas, o múltiples ciclos de iteración tras la validación final.

PTBG-IA-ES34 Identificación de cohortes digitales (estandarizadas a OMOP)

Descripción del servicio

Definición y selección de poblaciones de interés a partir de datos clínicos estandarizados al modelo común de datos OMOP, con generación de cohortes digitales homogéneas y reproducibles que facilitan la investigación comparativa y el análisis de resultados en salud.

Alcance del servicio

- Identificación de la cohorte según los criterios proporcionados; construcción de eventos índice y delimitación de periodos de observación; documentación de la especificación de la definición y del formato de salida.
- Localización de dominios y tablas OMOP: identificación de dominios y tablas relevantes (Condition Occurrence, Procedure Occurrence, Drug Exposure, Visit Occurrence, Measurement, Observation) y verificación de vocabularios estándar utilizados.
- Verificación terminológica y de codificación: comprobación de coherencia y trazabilidad de conceptos y códigos estándar (SNOMED CT, RxNorm, LOINC e ICD-10-ES cuando corresponda).
- Diseño y ejecución de la definición y la extracción: filtros por fechas y contexto asistencial; depuración de duplicados; reconciliación de identificadores de persona, episodio y visita; generación de agregaciones elementales.
- Generación de la cohorte digital: listado seudonimizado de pacientes y, cuando aplique, tabla de episodios asociada en CSV, Parquet o SQL; diccionario de variables y registro detallado de la extracción.
- Control de calidad básico: comprobaciones de completitud, consistencia e integridad referencial; informe conciso con conteos y distribución por sexo y edad.

No incluye

- Extracción de texto no estructurado ni procesamiento de lenguaje natural; véase PTBG-IA-ES07.
- Análisis estadístico, cuadros de mando o modelización; véase el bloque Análisis y Modelización.
- Estandarización a modelos OMOP o FHIR y mapeo terminológico avanzado; véase PTBG-IA-ES64.
- Integración en sistemas de producción, acceso directo a HCE o iteraciones múltiples.

- Comparativas entre cohortes o uso multicéntrico.
- Rediseños mayores de la definición de cohorte y de la lógica temporal, con nuevas reglas, secuencias y ventanas, o múltiples ciclos de iteración tras la validación final.

PTBG-IA-ES35 Caracterización de cohortes digitales (formato OMOP)

Descripción del servicio

Caracterización detallada de una cohorte previamente identificada en formato OMOP, con visión estructurada de su perfil clínico y demográfico para aportar información clave que apoye la investigación biomédica y el análisis de resultados en salud.

Alcance del servicio

- Descripción demográfica: distribución por edad, sexo y otras variables básicas.
- Perfil clínico: prevalencia de comorbilidades, diagnósticos principales y procedimientos asociados.
- Uso de recursos sanitarios: hospitalizaciones, visitas ambulatorias, pruebas diagnósticas y consumo de fármacos.
- Tratamientos y evolución: análisis de patrones terapéuticos y seguimiento básico.
- Informe estructurado: resultados en formato tabular y visual (CSV o Parquet y gráficos descriptivos), con resumen interpretativo, diccionario de variables y metodología de cálculo utilizada.

No incluye

- Análisis estadístico, cuadros de mando o modelización; véase el bloque Análisis y Modelización.
- Extracción de texto no estructurado ni procesamiento de lenguaje natural; véase PTBG-IA-ES07.
- Estandarización a modelos OMOP o FHIR y mapeo terminológico avanzado; véase PTBG-IA-ES64.
- Comparativas entre cohortes o análisis multicéntrico.
- Rediseños mayores de la definición de cohorte y de la lógica temporal, con nuevas reglas, secuencias y ventanas, o múltiples ciclos de iteración tras la validación final.

PTBG-IA-ES08 Preparación de datos

Descripción del servicio

Preparación, depuración, integración y estructuración de datos para transformar información bruta clínica, administrativa o experimental en conjuntos de datos coherentes y trazables, listos para análisis. Trabajo sobre datos ya extraídos conforme a PTBG-IA-ES06 y PTBG-IA-ES52, o sobre datos aportados por el solicitante.

Alcance del servicio

- Especificación técnica del trabajo: documentación de variables objetivo, fuentes, periodo y formato de salida definidos por el solicitante; verificación de viabilidad respecto de los datos disponibles.
- Perfilado inicial del dato: conteos y tipos; determinación de rangos y distribuciones; identificación de duplicados, valores atípicos y valores faltantes.
- Limpieza y depuración: corrección de formatos; normalización de unidades y fechas; eliminación de duplicados; armonización de claves internas.
- Integración de fuentes heterogéneas: resolución de identificadores y enlace de dominios (diagnósticos, procedimientos, farmacia, laboratorio, episodios), garantizando coherencia entre sistemas.
- Derivación básica de variables: generación de recuentos, indicadores simples y agregaciones por periodo o por episodio.
- Control de calidad y trazabilidad: comprobaciones de completitud, consistencia e integridad referencial; registro de la trazabilidad mediante registros detallados.
- Generación del dataset y documentación: acceso al conjunto de datos en CSV, Parquet o SQL, acompañado del diccionario de variables y de un informe conciso de calidad con conteos e incidencias.

No incluye

- Extracción desde sistemas fuente; véanse PTBG-IA-ES06 y PTBG-IA-ES52 a–d.
- Estandarización a modelos OMOP o FHIR y mapeo terminológico avanzado; véase PTBG-IA-ES64.
- Análisis estadístico, cuadros de mando o modelización; véase el bloque Análisis y Modelización.
- Automatización ETL en producción, integración en historia clínica electrónica o validación prospectiva o regulatoria.
- Rediseños mayores del procedimiento de extracción y preparación, con nuevas reglas de filtrado, mapeo y vinculación entre tablas, o múltiples ciclos de iteración tras la validación final.

PTBG-IA-ES55 Uso autónomo de la plataforma BIG DATA por personal con vinculación laboral con el H.U. i P. La Fe o el IIS La Fe

Descripción del servicio

Provisión de un entorno de trabajo en la nube para la explotación autónoma de conjuntos de datos previamente preparados por la Plataforma. Servicio limitado a personal con vinculación laboral vigente con el Hospital Universitario y Politécnico La Fe o con el IIS La Fe, con garantías de seguridad, confidencialidad, trazabilidad y cumplimiento normativo. Adecuado cuando se requiere un espacio seguro para trabajar de forma autónoma con datos ya preparados por la Plataforma, con herramientas estándar y controles de salida de información.

Alcance del servicio

- Herramientas estándar de software libre: JupyterLab, RStudio Server, clientes SQL y librerías habituales de R y Python (paquetes estadísticos y de visualización).
- Carga y organización del dataset preparado con registros de trazabilidad.
- DLP básico y controles de salida para evitar transferencia de datos personales; exportación autorizada de resultados agregados (tablas y figuras).
- Soporte básico: guía de uso, resolución de incidencias iniciales y mantenimiento menor de herramientas.
- Ajustes específicos del entorno como opción: véase PTBG-IA-ES53 Preparación de entorno de trabajo.

No incluye

- Descarga de conjuntos de datos crudos ni transferencia de datos personales fuera del entorno; no se habilita conexión directa con la HCE.
- Desarrollo a medida de infraestructura, instalación de software propietario sujeto a licencia específica o provisión de recursos especiales fuera del catálogo (por ejemplo, GPU de alta capacidad).
- Análisis estadístico, cuadros de mando o modelización; véase el bloque Análisis y Modelización.
- Soporte continuo 24×7, administración avanzada tipo SRE o almacenamiento indefinido; el entorno dispone de vigencia acotada y de políticas de retención y eliminación.
- Rediseños mayores del procedimiento de extracción y preparación, con nuevas reglas de filtrado, mapeo y vinculación entre tablas, o múltiples ciclos de iteración tras la validación final.

PTBG-IA-ES49 Feasability

Descripción del servicio

Evaluación rápida de la viabilidad de un estudio o ensayo clínico a partir de datos estructurados ya disponibles en el nodo local o en OMOP, con estimación de conteos de pacientes que cumplen criterios de inclusión y exclusión, ventanas temporales y estratificaciones clave. Objetivo: estimar el potencial de reclutamiento y ajustar el diseño antes de ejecutar extracciones o solicitar autorizaciones.

Alcance del servicio

- Consulta de viabilidad con datos disponibles: cálculo de conteos por criterio, coocurrencias y tendencias temporales con acumulación por meses o por trimestres, con reporte de estratificaciones por variables como edad o sexo cuando corresponda.

No incluye

- Extracción desde sistemas fuente; véase PTBG-IA-ES06.
- Estandarización a modelos OMOP o FHIR y mapeo terminológico avanzado; véase PTBG-IA-ES64.
- Identificación nominativa de pacientes, contacto o reclutamiento; véase PTBG-IA-ES50.
- Análisis estadístico, cuadros de mando o modelización; véase el bloque Análisis y Modelización.
- Diseño detallado del protocolo; véanse PTBG-IA-ES02 y PTBG-IA-ES03.
- Validación regulatoria o prospectiva ni integración en HCE; resultados orientados a predecisión y planificación.

PTBG-IA-ES53 Preparación entorno de trabajo

Descripción del servicio

Configuración de un entorno de trabajo personalizado en la nube, con instalación y validación de herramientas preferentemente de software libre, dimensionamiento de cómputo y almacenamiento y gestión de permisos de acceso, dejando el espacio listo para la explotación autónoma de datos en condiciones de seguridad, trazabilidad y escalabilidad.

Alcance del servicio

- Toma de requisitos: herramientas y versiones, librerías, kernels y entornos; necesidades de CPU, GPU y memoria; almacenamiento estimado; usuarios, roles y formato de salida.
- Provisión del entorno de trabajo seguro: aislamiento por proyecto, cifrado en reposo y en tránsito, estructura de carpetas, cuotas y copia de seguridad básica.
- Instalación y puesta a punto de herramientas estándar: JupyterLab, RStudio Server, clientes SQL y librerías habituales para análisis, aprendizaje automático y visualización; configuración de kernels y entornos.
- Identidades y accesos: alta de usuarios, perfiles y roles, autenticación multifactor y registro de auditoría.
- Controles DLP básicos: bloqueo de salida de datos personales y habilitación de exportación de resultados agregados en tablas y figuras.
- Pruebas de validación: lectura y escritura, rendimiento básico de CPU y GPU cuando aplique, carga de librerías y acceso a carpetas compartidas.
- Documentación: guía breve de uso, lista de paquetes y versiones instaladas y política de retención del entorno.
- No incluye
- Extracción desde sistemas fuente; véase PTBG-IA-ES06.
- Estandarización a modelos OMOP o FHIR y mapeo terminológico avanzado; véase PTBG-IA-ES64.
- Análisis estadístico, cuadros de mando o modelización; véase el bloque Análisis y Modelización.
- Soporte continuado 24×7, administración avanzada tipo SRE o escalado automático complejo.
- Instalación de software propietario sin licencias válidas; apertura de puertos externos fuera de la política de seguridad.
- Acceso o descarga de datos personales fuera del entorno; integración en HCE o producción, o validación prospectiva o regulatoria.
- Infraestructura especial fuera del catálogo sin presupuesto específico, como GPUs adicionales o almacenamiento no estándar.

PTBG-IA-ES64 Estandarización de datos

Descripción del servicio

Garantía de interoperabilidad y de calidad de los datos sanitarios aportados por el solicitante mediante su mapeo y transformación al estándar clínico requerido, por ejemplo OMOP CDM y HL7 FHIR, con validaciones básicas de conformidad y consistencia.

Alcance del servicio

- Revisión del dataset y definición del alcance: selección del estándar de destino y de los dominios y tablas a cubrir; identificación de terminologías necesarias.
- Mapeo de estructuras: correspondencia de campos y tablas; reglas de transformación de tipos, formatos, fechas y unidades; derivación sencilla de variables cuando proceda.
- Mapeo de vocabularios: alineación con LOINC, SNOMED CT, ATC y DDD, ICD-10-ES y otros terminológicos aplicables; registro de códigos y equivalencias.
- Validación y control de calidad: comprobaciones de completitud, unicidad, integridad referencial y conformidad con el estándar elegido; informe de incidencias y recomendaciones.
- Puesta a disposición del dataset estandarizado: entrega listo para análisis descriptivos y consultas en el estándar de destino, con documentación técnica básica compuesta por diccionario de datos y registro de transformaciones.

No incluye

- Extracción desde sistemas fuente ni desarrollo de conectores ETL complejos.

PTBG-IA-ES65 · Creación de catálogo de datos

Descripción del servicio

Creación y mantenimiento de fichas de metadatos de los datasets, distribuciones y servicios de la Plataforma conforme al perfil HealthDCAT-AP, para mejorar la visibilidad, la interoperabilidad y la gobernanza.

Alcance del servicio

- Captura y normalización de metadatos mínimos de HealthDCAT-AP: Dataset, Distribution, CatalogRecord y DataService; título, descripción, palabras clave, temática, cobertura temporal y espacial, responsable, punto de contacto, licencia, derechos de acceso y versión.

- Identificadores y enlaces: asignación de URIs y PIDs cuando existan; relación con proyectos y estudios; referencia a estándares en conformsTo como OMOP, HL7 FHIR y MIABIS.
- Descripción de distribuciones y servicios: formatos como CSV, Parquet u OMOP; endpoints como API, S3 o SQL; esquemas y requisitos de acceso y autenticación; tamaño y frecuencia de actualización.
- Calidad y trazabilidad: fechas de publicación y actualización, provenance, método de generación, notas de uso y validación sintáctica del perfil HealthDCAT-AP.
- Publicación y entrega: alta en el catálogo interno y exportables en JSON-LD y RDF, y ficha en PDF o HTML para revisión y reutilización; soporte básico de actualización de metadatos.

No incluye

- Curado o transformación del contenido del dataset.
- Integración con recolectores externos o portales nacionales o europeos.
- Automatización ETL del catálogo; se presupuesta aparte