



FONDO EUROPEO DE DESARROLLO REGIONAL
Una manera de hacer Europa

Jornada de Presentación de Pliegos CPI de FID 3

Medigenomics

Dr. Pablo Lapunzina Badía

Facultativo especialista de área del Servicio de Genética del Hospital Universitario La Paz. Experto científico para el proyecto Medigenomics

FONDO EUROPEO DE DESARROLLO REGIONAL

Una manera de hacer Europa

ÍNDICE

DEFINICIÓN GENERAL DEL SERVICIO

I

CONSIDERACIONES TÉCNICAS

II

DEFINICIÓN GENERAL DEL SERVICIO



FONDO EUROPEO DE DESARROLLO REGIONAL
Una manera de hacer Europa

Secuenciación genómica completa y análisis de genomas completos:

La compra del servicio de secuenciación de genomas completos y análisis de genomas dentro de la plataforma de genomas está dirigida a cumplir los objetivos específicos definidos:

- a) Utilizar los datos de los genomas para testar la solución propuesta I.
- b) Validar y analizar casos de uso real de pacientes
- c) Establecer sinergias con otros proyectos CPI

Resolución de los retos 1 a 5

Reto 1- Sistema de obtención automática de datos (voz/texto).

Reto 2- Integración de un Sistema de análisis terciario de la información genómica en la Historia Clínica Electrónica (en adelante HCE) de la Comunidad de Madrid.

Reto 3- Almacenamiento y acceso restringido/controlado a la información genómica.

Reto 4- Actualización automática de la información médica asociada a la información genómica y captura de datos proveniente de la literatura científica/médica.

Reto 5- Actualización de la información en HCE del paciente en el sistema de salud y alertas en el dispositivo personal de cada persona. Acceso de los pacientes a sus registros médicos..

DEFINICIÓN GENERAL DEL SERVICIO

FONDO EUROPEO DE DESARROLLO REGIONAL

Una manera de hacer Europa

Actualización de la información en Hist. Clínica del paciente en el sistema de salud y alertas en el dispositivo personal de cada persona. Acceso de los pacientes a sus registros médicos.

Sistema de obtención automática de datos (voz/texto)

Se trata del diseño para la incorporación en los sistemas de HCE de los sistemas de la Comunidad de Madrid (todos los disponibles) de una herramienta de extracción de información desde texto (probablemente aplicando tecnologías de NLP u otras similares) y de extracción a texto de voz (de la consulta de personal sanitario y pacientes).



Tras completar el proceso de revisión, actualización, curación y aprobación por un comité ad hoc de las evidencias científicas que vayan apareciendo en la literatura científica internacional. Estas nuevas evidencias deberán ser incorporadas en una aplicación dentro de la HCE del paciente y alertar al médico tratante y al paciente

Integración de un Sistema de análisis terciario de la información genómica en la HCE de la CM.

Consiste en una plataforma de análisis (software) que se incorpore dentro de la HCE de los sistemas de la Comunidad de Madrid (todos los disponibles) con el fin de procesar la información genómica terciaria, y mantener actualizados los datos genómicos

Almacenamiento/acceso restringido/controlado a la información genómica.

Deberá resolver el problema del almacenamiento de una enorme cantidad de información genética y genómica de cada paciente, en el formato de un servidor local de capacidad operativa, pero además la forma de almacenamiento masivo en la nube a través de un servidor seguro

Actualización automática de la información médica asociada a la información genómica. Captura de datos proveniente de la literatura médica.

El reto pretende generar una herramienta que de forma automática pueda extraer de las fuentes de información científica médica habituales (tales como PubMed y similares) y de forma periódica toda la información referente a nuevos genes, nuevas enfermedades y/o nuevas dianas o marcadores genéticos y genómicos asociados a las enfermedades

FONDO EUROPEO DE DESARROLLO REGIONAL

Una manera de hacer Europa

ÍNDICE

DEFINICIÓN GENERAL DEL SERVICIO

I

CONSIDERACIONES TÉCNICAS

II



La Adjudicataria desplegará un servicio integral que permita disponer de los recursos técnicos necesarios para satisfacer las necesidades del pliego técnico según los niveles de calidad requeridos y dentro de los plazos acordados.

La Adjudicataria deberá tener en cuenta todas las tareas relativas a las consideraciones técnicas del PPT(Cláusula Tercera) así como todas aquellas tareas de gestión asociadas al servicio y todas aquellas necesarias para el correcto funcionamiento del servicio.

Consideraciones generales (parcial)

- La plataforma de Medigenomics deberá disponer de **un sistema de NLP (natural language processing) que obtenga información clínicamente relevante a partir de varios formatos (voz a texto durante la consulta, informes previos en papel, historia clínica electrónica, imagen, etc.),**
- La solución contará con una **herramienta que automatice las búsquedas y filtros de información científica publicada para la obtención de información y su actualización continua.**
- La plataforma Medigenomics dispondrá de **sistemas de análisis terciarios de genomas y sistemas de interrogación de datos.**
- Los datos manejados serán tratados de manera que se **cumpla con los requisitos de seguridad y privacidad establecidos por la General Data Protection Regulation (en adelante GDPR) 2016/679.(Ver Anexo I del PPT).**
- Medigenomics incluirá **sistemas de anonimización y/o pseudoanonimización** de los datos de cada paciente.
- La plataforma deberá **ofrecer una gestión de usuarios que diferencie las funcionalidades accesibles.**
- Se deberá garantizar el **máximo nivel de seguridad en todas las transacciones de datos.**
- Deberá existir un mecanismo de conexión bidireccional con el proyecto INFOBANCO.**

❑ Consideraciones de requerimientos mínimos informáticos

Resolución de los retos 1 a 5

- a) La Adjudicataria deberá garantizar que la plataforma **sea modular, flexible y escalable a su futura implementación en toda la Comunidad de Madrid.**
- b) La solución deberá tener **capacidades de integración para ser interconectada a los diferentes sistemas de información que conviven en la Comunidad de Madrid.**
- c) Las soluciones deben basarse principalmente en **herramientas abiertas.**
- d) **La Adjudicataria debe garantizar la seguridad de la solución en todas las áreas a desarrollar.**
- e) **La plataforma innovadora debe permitir su integración y compatibilidad con los proyectos CPI Infobanco e Integra-Cam.**
- f) **La Adjudicataria deberá garantizar la gestión e integración de diferentes socios que desarrollen actividades dentro de la propuesta.**
- g) La solución deberá seguir **un modelo de servicio híbrido entre alojamiento en la nube y local, ya que los datos sanitarios se almacenan en las dependencias de los centros sanitarios.**
- h) Se ha de tener en cuenta que es **responsabilidad de la Adjudicataria el proporcionar el servicio completo Medigenomics integrando los cinco retos que forman parte del alcance de este proyecto.**
- i) **La arquitectura técnica permitirá el proceso de actualización de historiales clínicos a partir de nuevas referencias bibliográficas.**
- j) Asimismo, dispondrá de los requisitos técnicos necesarios para el **filtrado y curado de información proveniente de fuentes de artículos científicos (principalmente PubMed) y su actualización a los pacientes.**
- k) Deberá existir un mecanismo de **conexión bidireccional con el proyecto INFOBANCO utilizando el estándar HL7 FHIR.**
- l) Deberá permitir **la capacidad de escalado suficiente para poder ser prestado a gran escala.**

❑ Consideraciones de requerimientos mínimos informáticos

• Secuenciación de genomas completo y análisis de genomas.

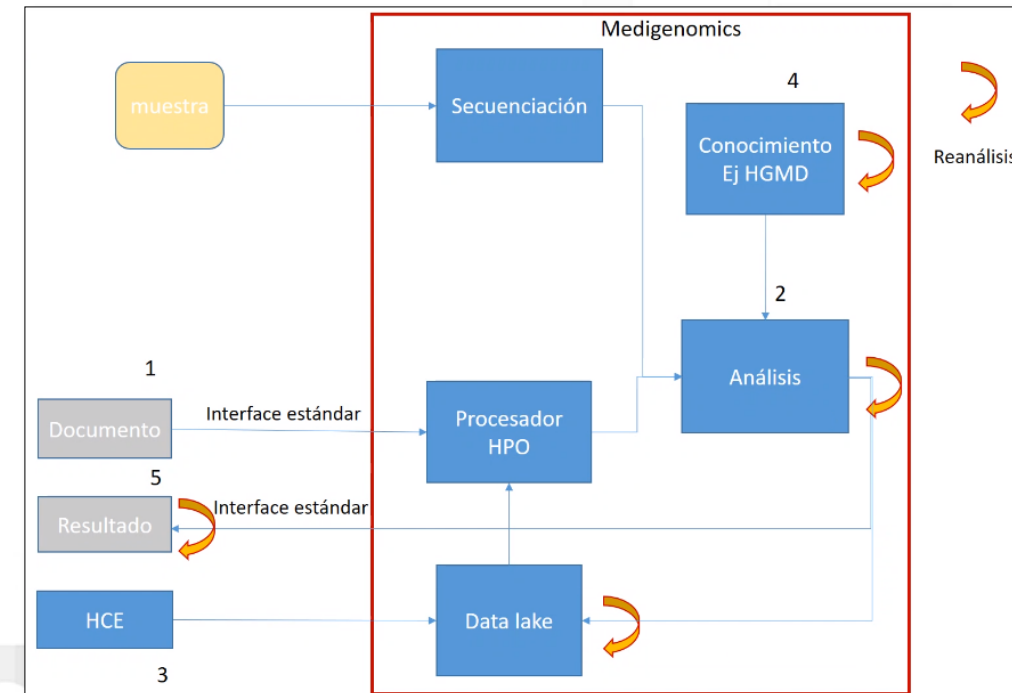
La Adjudicataria deberá tener en cuenta los siguientes requerimientos de los genomas:

- La **secuenciación de los 380 genomas** se debe realizar utilizando la tecnología más actual.
- En cuanto a los **parámetros de calidad**, debe ser 150bp Paired-End (PE), con una profundidad media de 30x, Q30 (valor de calidad Phred) mayor del 80%, 90Gb de archivo BAM de lecturas, y obtener los archivos .fastq, .bam, .bai, .vcf anotados en la versión del genoma humano hg38.
- Deben tener un **informe de calidad de los resultados obtenidos del alineamiento, anotación y calidad de las lecturas obtenidas.**

• Pruebas

Consistirán en:

- Probar que el **sistema es capaz de leer los archivos .fastq (o bcl) y procesarlos para realizar el análisis de calidad, alineamiento, anotación y generación** de archivos .bam, .bai, .vcf.
- Probar que el **sistema es capaz de automatizar el proceso de priorización de variantes siguiendo los criterios establecidos previamente y es capaz de generar un informe final de resultados automatizado**, que exponga los **resultados de aquellas variantes más relevantes con utilidad clínica.**
- Probar que el **sistema es capaz de almacenar los datos de una manera eficiente**, siendo capaces de explotar los datos desde la fuente donde se encuentren almacenados, de una manera rápida, sin necesidad de copiar los datos a otras fuentes y teniendo en cuenta que los datos deben tener un respaldo para evitar la pérdida.



Procesos Medigenomics.

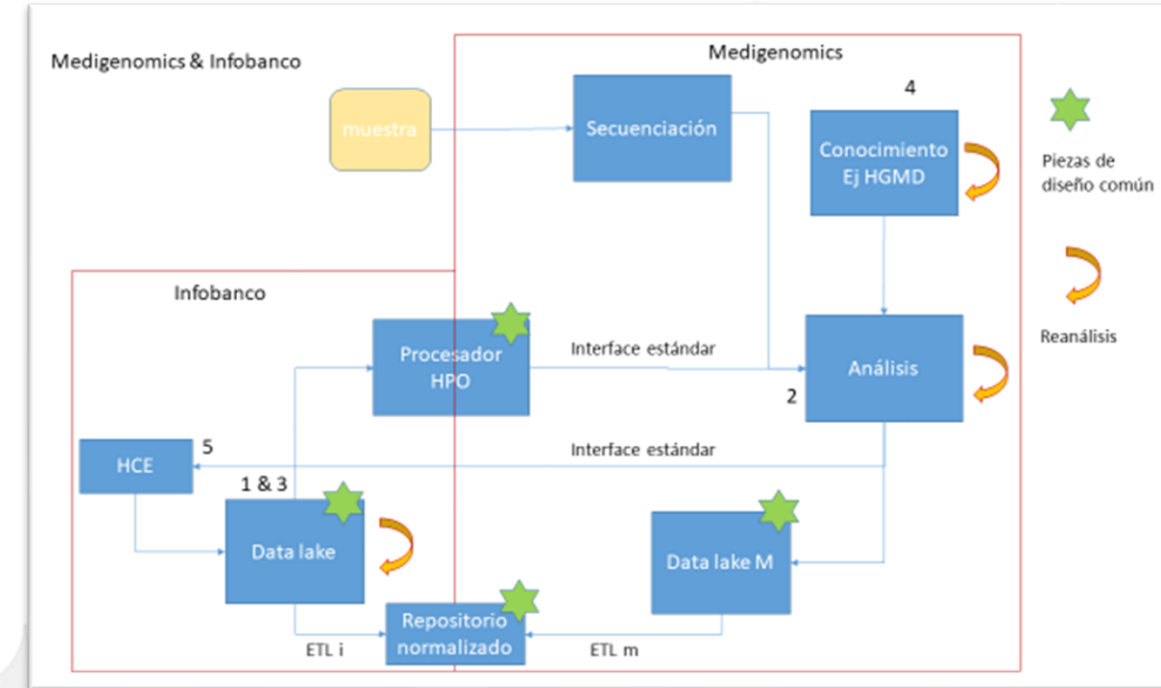
❑ Consideraciones de integración y gobernanza

A continuación, se especifican **los requerimientos técnicos de Medigenomics para la integración con Infobanco**:

- ❑ La integración de **Medigenomics con Infobanco se hará a través de entradas y salidas de datos acordadas y/o estándares**, así como haciendo uso de componentes compartidos.

Para ello es necesario:

- a) **Capacidad de incorporar al flujo de trabajo los servidores de modelos, de terminologías y de ETL desarrollados en la plataforma Infobanco**
- b) **Capacidad de desarrollar e incorporar el componente 'Procesador fenotipo HPO' de manera coordinada con Infobanco**. La Adjudicataria de Infobanco realizará la inferencia de los conceptos HPO a partir de datos estructurado normalizado, mientras que **La Adjudicataria de Medigenomics realizará el procesamiento de lenguaje natural para extraer los conceptos HPO**.
- c) **Interfaz de entrada y salida conforme al estándar HL7 FHIR**.
- d) **Pasarela de datos en bruto desde el lago de datos del sistema a través de ficheros (XML, JSON, CSV, TXT, etc.)**.
- e) **Acceso a tablas y/o vistas del lago de datos del sistema (JDBC, ODBC, etc.)**.



Integración Medigenomics con el proyecto Infobanco.



FONDO EUROPEO DE DESARROLLO REGIONAL
Una manera de hacer Europa

Jornada de Presentación de Pliegos CPI de FID 3

Medigenomics

Muchas Gracias