

IMPACT

DATA



E5.6 Informe final sobre el funcionamiento de la red IMPACT de beacons

v 1.0

Programa	IMPACT: Infraestructura de Medicina de Precisión asociada a la Ciencia y la Tecnología		
Nombre Proyecto	IMPACT-Data: Programa de Ciencia de Datos de IMPACT		
Expediente	IMP/00019		
Duración	Enero 2021 – Diciembre 2023		
Paquete Trabajo	WP5. Integración de Datos		
Tarea	Tarea 5.2 – Implantación de mecanismos para la identificación automática de información genómica relevante, o contenido de las imágenes médicas que den respuestas anonimizadas al estilo beacon.		
Entregable	E5.6. Informe final sobre el funcionamiento de la red IMPACT de beacons		
Versión	Número de versión del documento		
Fecha Entrega	31/03/2024	Fecha Aprobado	30/06/2024
Responsable	FPS		
Nivel	X	PU	Público
Diseminación		CO-IMP	Confidencial, sólo participantes de los pilares de IMPACT, incluyendo la comisión de evaluación de IMPACT.
		CO-DATA	Confidencial, sólo participantes de IMPACT-Data, incluyendo la comisión de evaluación de IMPACT.

Autores		
Organización	Nombre	Rol
FPS	Javier Pérez Florido	Coordinador/Autor
FPS	Joaquín Dopazo	Coordinador/Autor
CRG	Jordi Rambla	Revisor
CNIO	Fátima Al-Shahrour	Revisor
BSC	Sergi Aguiló-Castillo	Autor
BSC	Dmitry Repchevsky	Autor
CRG	Amy Curwin	Autor
CRG	Oriol Lopez-Doriga Sagales	Autor
CRG	Arnau Soler Costa	Autor

Historial de versiones			
Nro.	Fecha	Descripción	Autor
v 0.1	10/03/2024	Borrador del índice	Javier Pérez Florido (FPS)
v 0.2	13/03/2024	Índice propuesto	Javier Pérez Florido (FPS)
v 0.3	15/03/2024	Aprobación del índice	Javier Pérez Florido (FPS) Fátima Al-Shahrour (CNIO) Jordi Rambla (CRG)
v 0.4	19/04/2024	Elaboración de contenidos	Javier Pérez Florido (FPS)
v 0.5	25/04/2024	Sección 3.2.2 y 3.2.3	Sergi Aguiló-Castillo (BSC) Dmitry Repchevsky (BSC)
v 0.6	29/04/2024	Sección 3.2.1	Amy Curwin (CRG) Oriol Lopez-Doriga Sagales (CRG) Arnau Soler Costa (CRG)
v 0.7	06/05/2024	Revisión interna final	Javier Pérez Florido (FPS) Joaquín Dopazo (FPS)
v 0.7	07/05/2024	Entregable listo para su revisión	
v 0.8	23/05/2024	Sugerencias adoptadas de	Javier Pérez Florido (FPS)

		acuerdo a revisor 2	
V 0.9	18/06/2024	Sugerencias adoptadas de acuerdo a revisor 1	Javier Pérez Florido (FPS)
V 0.9	18/06/2024	Entregable finalizado	FPS
V 1.0	30/06/2024	Entregable aprobado	CD de IMPaCT-Data

Contenido

Contenido	3
Tablas	5
Figuras	6
Resumen Ejecutivo	7
Introducción	8
Audiencia	8
Ámbito	8
Relación con otros Entregables	8
Estructura Entregable	8
1. Sistemas Beacon: aspectos técnicos y casos de uso existentes	9
2. Beacon-Network	12
2.1 ¿Qué es una red Beacon?	12
2.2 Ejemplos de redes <i>Beacon</i>	13
2.3 Arquitectura de una Red <i>Beacon</i>	14
2.3.1 Proveedor de Identidades (<i>IdP</i>)	15
2.3.2 Agregador de beacons (<i>Aggr</i>)	15
2.3.3 Registro de Beacon	15
2.4 Tipos de Redes <i>Beacon</i>	16
2.5 Cómo desplegar una red <i>Beacon</i>	16
3. Red IMPaCT de beacons	17

E5.6. Informe final sobre el funcionamiento de la red IMPaCT de beacons v. 1.0

3.1	Nodos participantes y características.....	18
3.2	Arquitectura de la <i>Beacon Network</i> de <i>IMPACT-Data</i>	20
3.2.1	Beacon network: <i>front-end</i>	21
3.2.2	Beacon Network: <i>back-end</i>	26
3.2.3	Recursos técnicos y humanos de la <i>Beacon Network</i> de <i>IMPACT</i>	26
4.	Conexión con otros programas.....	27
5.	Conclusiones.....	28
	Referencias.....	29
	Acrónimos y Abreviaturas.....	30

Tablas

Tabla 1. Listado de nodos participantes en la <i>Beacon Network</i> de <i>IMPACT-Data</i>	18
Tabla 2. Descripción de los datos y su naturaleza de cada instancia <i>Beacon</i> de la red	19
Tabla 3. Entidades implementadas con datos de cada nodo <i>Beacon</i> de la red.....	19

Figuras

Figura 1. Ejemplo de instancia <i>Beacon v1</i>	9
Figura 2. <i>Beacon v2: Beacon Framework y Modelo Beacon</i>	11
Figura 3. Representación conceptual de una <i>Beacon Network</i>	13
Figura 4. Aspectos técnicos de una <i>Beacon Network</i>	14
Figura 5. Distribución geográfica de la <i>Beacon Network</i> de <i>IMPACT-Data</i>	20
Figura 6. <i>Front-end</i> de la <i>Beacon Network</i> de <i>IMPACT-Data</i>	21
Figura 7. Botón para desplegar listado de <i>filtering terms</i>	22
Figura 8. Ejemplo de selección de un <i>filtering term</i>	22
Figura 9. Ejemplo de adición de un <i>filtering term</i> correctamente seleccionado	23
Figura 10. Ejemplo de selección de un <i>filtering term</i> numérico	23
Figura 11. Ejemplo de selección de un <i>filtering term</i> alfanumérico	24
Figura 12. Ejecución de una consulta con los <i>filtering terms</i> seleccionados	24
Figura 13. Ejemplo de respuesta en la <i>Beacon Network</i> para la consulta realizada	25
Figura 14. Ejemplo de consulta para la entidad <i>Variants</i>	25
Figura 15. Ejemplo de consumo de memoria asociado a una consulta en el <i>agregador</i> de la <i>Beacon Network</i>	27

Resumen Ejecutivo

Un sistema *Beacon* es una herramienta para el descubrimiento de variantes genómicas que considera, bajo una misma entidad, numerosos conjuntos de datos genómicos. La API de *GA4GH Beacon* habilita la búsqueda de variantes genómicas e información asociada sin comprometer la privacidad del conjunto de datos de forma que cualquier institución puede hacer que sus conjuntos de datos ómicos sean descubribles.

Los sistemas *Beacon* han evolucionado a lo largo de los últimos años, partiendo de un sistema básico (*Beacon v1*) donde se pregunta acerca de la presencia de una determinada variante genómica obteniendo una respuesta afirmativa o negativa hasta nuestros días donde la versión v2 permite consultas más complejas de forma que el protocolo *Beacon* cubre diferentes entidades y detalles asociados a las mismas. Esta nueva versión viene a cubrir las demandas necesarias para la generación y compartición de datos genómicos en el ámbito de la salud, incluyendo información sobre la biomuestra y el procedimiento de análisis molecular, así como las observaciones medidas que describen el estado fenotípico de los individuos.

Para desarrollar un “internet para la genómica”, es útil establecer una red de organizaciones que adopten protocolos como *Beacon* y un mecanismo eficiente para buscar a través de dicha red, algo que se conoce como *red de beacons* o *Beacon network*. Así pues, en el presente documento se describirán los aspectos técnicos relacionados con la implantación de una *Beacon network* y se darán los detalles de la red de *beacons* desplegada en IMPACT-Data.

Introducción

Audiencia

Este entregable está dirigido a todas aquellas entidades (hospitales, centros de investigación, grupos de investigación) que deseen compartir datos genómicos y datos biomédicos asociados (al nivel que decidan) sin comprometer la privacidad de los mismos mediante un sistema *Beacon* e integrar dicho *Beacon* en una red de *beacons*.

Ámbito

Este entregable supone un punto de partida para aquellas entidades que participan, en el contexto del proyecto IMPaCT-Data, en la compartición de datos genómicos e información asociada. La compartición de este tipo de información es fundamental para el desarrollo del programa de medicina personalizada en los diferentes sistemas de atención sanitarios.

Relación con otros Entregables

Este entregable está relacionado con el documento E5.4. *Requisitos Técnicos para puesta en marcha de Sistemas Beacon*, así como con el entregable E5.2 *Demostrador de Integración de Datos Biomédicos*. También está relacionado con el entregable del WP3: E3.5 “Comparativa de Procesos en Entornos Hospitalarios”

Estructura Entregable

El entregable está estructurado de la siguiente forma:

1. Sistemas *Beacon*: aspectos técnicos y casos de uso existentes.
2. *Beacon Network*: ¿qué es una red de *beacons*? Ejemplos de redes *Beacon*. Arquitectura de una red *Beacon*. Cómo desplegar una red *Beacon*. Buenas prácticas asociadas a una red *Beacon*.
3. Red IMPaCT de *beacons*. Nodos participantes y características. *Front-end* y *back-end* de la Red *Beacon*.
4. Conexión con otros programas.

1. Sistemas Beacon: aspectos técnicos y casos de uso existentes

En sus orígenes, *Beacon* es un protocolo básico de descubrimiento de datos impulsado por la iniciativa *Global Alliance for Genomics and Health (GA4GH)*. En su versión 1, permite a los usuarios obtener información de la presencia o ausencia de una variante genómica en un conjunto de datos alojado por el proveedor o nodo *Beacon* (Figura 1).

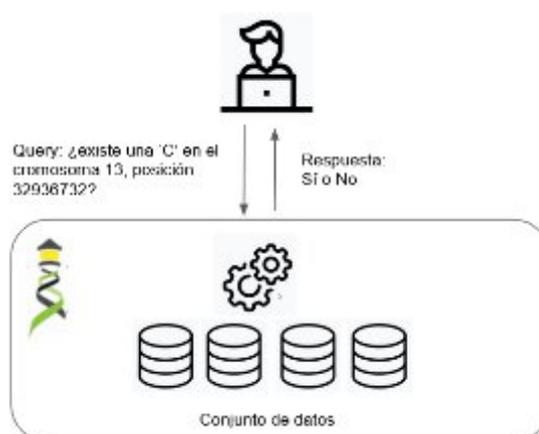


Figura 1. Ejemplo de instancia Beacon v1

Aunque el concepto *Beacon* es sencillo, la potencia del procedimiento de compartición de datos es evidente, especialmente si se implementa a través de un modelo federado, distribuyendo las consultas *Beacon* a diversos nodos y proporcionando una agregación de las respuestas individuales. Debido a esta gran aceptación, ELIXIR, la organización europea de infraestructura de bioinformática, se unió al desarrollo del proyecto con el objetivo de disponer de una especificación estándar que permitiera una usabilidad mejorada y un uso más amplio en el ámbito de la biomedicina. De hecho, en 2016, más de 35 organizaciones a lo largo y ancho del mundo adoptaron el protocolo *Beacon* en más de 90 conjuntos de datos, muchos de ellos conectados al agregador de red proporcionado por DNAtack (<https://beacon-network.org/>).

En 2018, el protocolo *Beacon v1* fue aceptado como estándar oficial de la GA4GH [1] tras un proceso de revisión formal. Aunque esta versión y sus actualizaciones introdujeron algunas mejoras respecto a las versiones anteriores, el protocolo *Beacon v1* se mantuvo fiel al concepto original de “consultas de variantes y respuesta agregada”, pero quedó claro que las

E5.6. Informe final sobre el funcionamiento de la red IMPaCT de beacons v. 1.0

mejoras adicionales solicitadas del protocolo, especialmente en el contexto de aplicaciones clínicas en enfermedades raras y genómica del cáncer, requerían una redefinición del protocolo *Beacon* para satisfacer numerosos casos de uso, lo que llevó a la iniciación del proceso de diseño de *Beacon v2*.

Muchos de estos casos de uso provienen de entornos clínicos y la necesidad de ampliar la funcionalidad del protocolo *Beacon* en dichos entornos. Por ejemplo, una consulta específica de una variante genómica debería ofrecer opciones adicionales para consultar y recuperar datos fenotípicos asociados, metadatos (por ejemplo, edad de inicio de la enfermedad, sexo genotípico), diagnósticos asociados, intervenciones terapéuticas, información de pedigrí, entre otros. Así pues, existía un fuerte argumento para expandir su usabilidad en entornos relacionados de la salud también como consecuencia del incremento en entornos hospitalarios de pruebas genéticas/genómicas como parte integral del proceso de diagnóstico o en la selección de procedimientos terapéuticos.

Por tanto, la generación y compartición de datos genómicos en el ámbito de la salud debe proporcionar contexto para la búsqueda de variantes genómicas, incluyendo información sobre la biomuestra y el procedimiento de análisis molecular, así como las observaciones y medidas que describen el estado fenotípico de los individuos. Por tanto, el modelo de datos de *Beacon v2* implementa unas entidades por defecto: “Individual” (individuo), “Biosample” (muestra biológica), “Genomic Variations” (variantes genómicas) y “Variant Information” (información de variante) y sus relaciones lógicas, así como conceptos técnicos adicionales [2]. Permite realizar preguntas combinando parámetros fenotípicos de un individuo junto con parámetros de variantes genómicas, por ejemplo “existen individuos con fenotipo X y variante Y”. Adicionalmente, el modelo *Beacon v2* soporta el uso de conceptos de agrupación como “Dataset” y “Cohort”, por ejemplo, para preguntar datos particulares a un determinado recurso en una instancia *Beacon* grande o preguntar datos particulares a un conjunto de individuos de una determinada cohorte de estudio.

Beacon v2 está organizado en dos bloques principales: el denominado *Beacon Framework* y los Modelos de *Beacon* [2]. Por un lado, el *framework* es agnóstico al conocimiento (Figura 2) y al dominio e incluye las características relacionadas con la descripción de instancias de *Beacon* (metadatos), solicitudes de consulta, respuestas de consulta y filtros, entre otros. Por otro lado, los modelos de *Beacon* describen las entidades del dominio y las relaciones entre ellas. Para obtener interoperabilidad entre instancias de *Beacon*, *Beacon v2* incluye un modelo recomendado para diagnóstico e investigación en genómica clínica, tal como se describe en el párrafo anterior. Separar el *marco* del *modelo* permite que otras disciplinas adopten el concepto de *Beacon* sin apartarse del estándar mismo y sin la obligación de implementar un modelo que no sea relevante para su dominio.

E5.6. Informe final sobre el funcionamiento de la red IMPACT de beacons v. 1.0

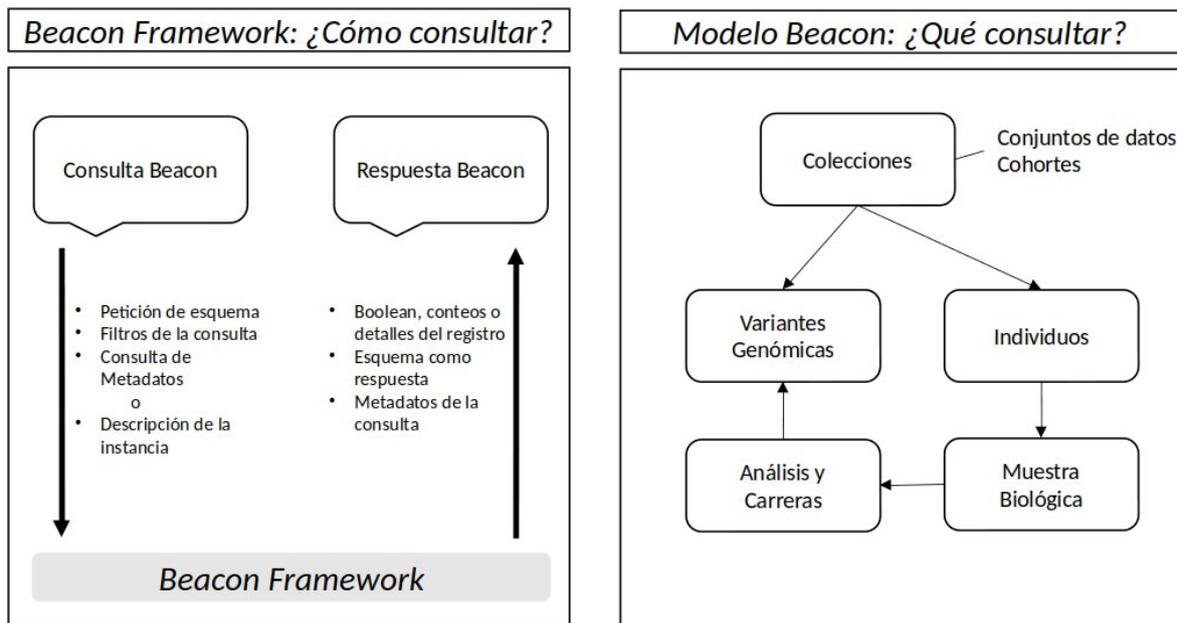


Figura 2. Beacon v2 está compuesto de dos partes: Beacon Framework y el Modelo Beacon. El primero describe el protocolo de peticiones y respuestas mientras que el segundo describe las entidades típicas en el dominio de investigación clínica, aunque otros modelos pueden ser utilizados. Figura extraída de [2]

Para simplificar el proceso de desplegar un *Beacon*, existe una implementación de referencia *open source* disponible con la especificación más actual (<https://b2ri-documentation.readthedocs.io/en/latest/>). Esta implementación permite crear un *Beacon* público a partir de un conjunto de ficheros de variantes genómicas VCF. Puede ser desplegado de forma local o en un entorno *cloud* mantenido por otros proveedores (Amazon, Google o Microsoft).

Por otro lado, *Beacon* utiliza un modelo de acceso de 3 niveles o *tiers* [2]: anónimo, registrado y controlado. Un *Beacon* que admite acceso anónimo responde a las consultas independientemente de la fuente de la consulta. Para que un *Beacon* responda a una consulta en el nivel registrado, el usuario debe identificarse ante el *Beacon*. Para ello, puede utilizar una identidad de ELIXIR controlada por la Infraestructura de Autorización y Autenticación de ELIXIR (*ELIXIR Authorization and Authentication Infrastructure, ELIXIR AAI*; <https://elixir-europe.org/services/compute/aai>). La ELIXIR AAI permite a los investigadores utilizar las credenciales de su organización de origen, comunidad o identidades comerciales (por

ejemplo, ORCID, LinkedIn) para iniciar sesión y acceder a los datos y servicios que necesitan. Finalmente, para que un *Beacon* responda a una consulta de acceso controlado, el usuario debe haber solicitado y obtenido acceso al *Beacon* (o datos derivados de uno o más individuos dentro del *Beacon*) antes de enviar la consulta. Es importante señalar que un *Beacon* puede contener conjuntos de datos (o colecciones de individuos) cuyos datos solo son accesibles en niveles específicos dentro del *Beacon*. Este modelo de acceso por niveles o *tiers* permite al propietario de un *Beacon* determinar qué respuestas se devuelven a quién dependiendo de la consulta en sí misma y del usuario que realiza la solicitud. De esta forma, se puede garantizar que la respuesta respete el consentimiento o la base legal bajo la cual se recopilaron los datos, de acuerdo a, por ejemplo, ciertos principios limitantes dentro del Reglamento General de Protección de Datos (General Data Protection Regulation, GDPR).

2. Beacon-Network

Beacon representa un protocolo simple que, al igual que los protocolos de Internet como HTTP, describe un método para descubrir e intercambiar datos entre sistemas distribuidos y colaborativos. Para desarrollar un 'internet para la genómica', es útil establecer una red de organizaciones que adopten protocolos como *Beacon* y un mecanismo eficiente para buscar a través de dicha red [1].

2.1 ¿Qué es una red Beacon?

La red *Beacon* o *Beacon Network* es un directorio y motor de búsqueda para *beacons*. Aunque los *beacons* individuales responden a la pregunta "¿Has observado esta variante genómica?", la Red Beacon responde a la pregunta "¿Quién tiene esta variante genómica?". Por tanto, la red *Beacon* sirve como un canal de distribución de datos genómicos potente y en tiempo real a través del cual los usuarios pueden descubrir la existencia de variantes genómicas de interés y ser dirigidos a las organizaciones anfitrionas que los tienen. Un esquema de la Red *Beacon* como una red federada global para el descubrimiento de información genómica se muestra en la Figura 3.

E5.6. Informe final sobre el funcionamiento de la red IMPACT de beacons v. 1.0

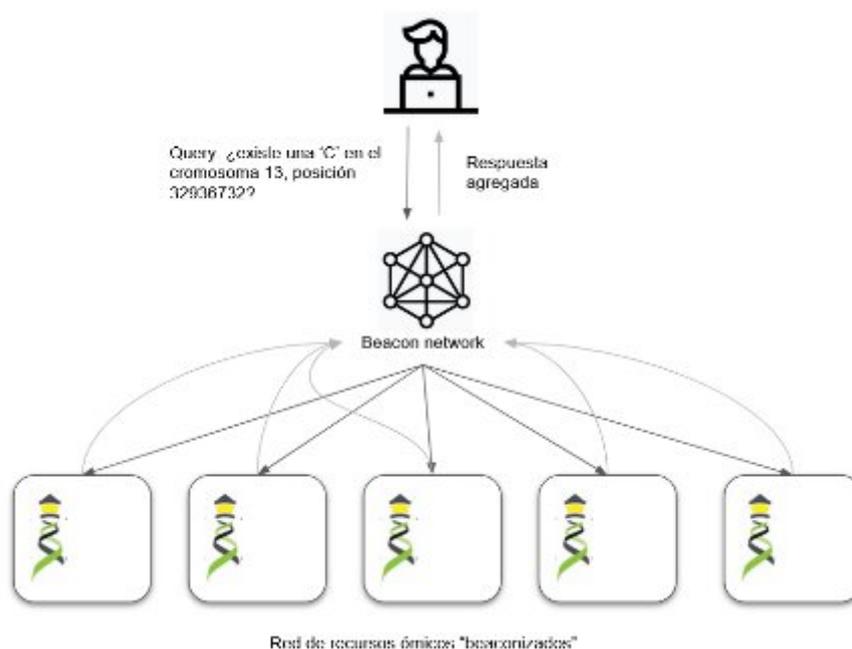


Figura 3. Representación conceptual de una Beacon Network. El usuario interactúa con la red Beacon por medio de una pregunta relacionada con la existencia de una variante genética determinada. La red realiza esta pregunta a lo largo de todas las instancias Beacon integradas en dicha red. Las respuestas obtenidas por cada entidad son recolectadas y presentadas de forma agregada al usuario.

La red Beacon es accesible tanto a través de su sitio web como *programáticamente* mediante una API, y permite la búsqueda rápida y simultánea de cientos de conjuntos de datos de cientos de miles de individuos a través de *beacons* en todo el mundo. Los *beacons* pueden registrarse libremente en la red Beacon y pueden ser buscados de forma independiente o en conjunto con otros *beacons* conectados.

2.2 Ejemplos de redes Beacon

Existen varios ejemplos de redes de Beacon. Por ejemplo, la Beacon Network [1] de DNASTack (<https://beacon-network.org>) en el que participan 35 instituciones de todo el mundo y entre ellas se encuentran tres españolas: Fundación Progreso y Salud donde el Beacon está desplegado sobre la base de datos de variabilidad española [3] y otras dos instituciones, como el Centro Nacional de Análisis Genómico (CNAG) y el Centro de Regulación Genómica (CRG).

Otro ejemplo de red *Beacon* es el prototipo de la red de ELIXIR (<https://beacon-network.elixir-europe.org/>). La futura red *Beacon* de ELIXIR agregará *beacons* de conjuntos de datos de nodos ELIXIR y de la Infraestructura Europea de Datos Genómicos (*European Genomic Data Infrastructure, GDI*). Está subvencionada por ELIXIR y estará disponible desde instituciones designadas en nodos ELIXIR.

2.3 Arquitectura de una Red *Beacon*

Una instancia de una red *Beacon* está compuesta por una serie de componentes software (Figura 4) que proporcionan determinadas funcionalidades [4]:

1. Una interfaz web (*User Interface, UI*) que permite el usuario hacer *login*, explorar los *beacons* que son miembros de la red, realizar preguntas, obtener resultados y acceder a enlaces que permiten obtener detalles sobre los resultados obtenidos.
2. Un proveedor de identidad (*Identity Provider, IdP*) local o remoto que permite la autenticación de los usuarios.
3. Un registro que aloja una lista de los *beacons* que son parte de la red
4. Un *agregador* que recopila las acciones de la interfaz de usuario, las envía a cada *beacon* y recopila todas las respuestas obtenidas hacia la interfaz de usuario. Hay que tener en cuenta que este *agregador* tiene disponible una API tipo REST que puede ser consultada directamente sin requerir una interfaz de usuario.

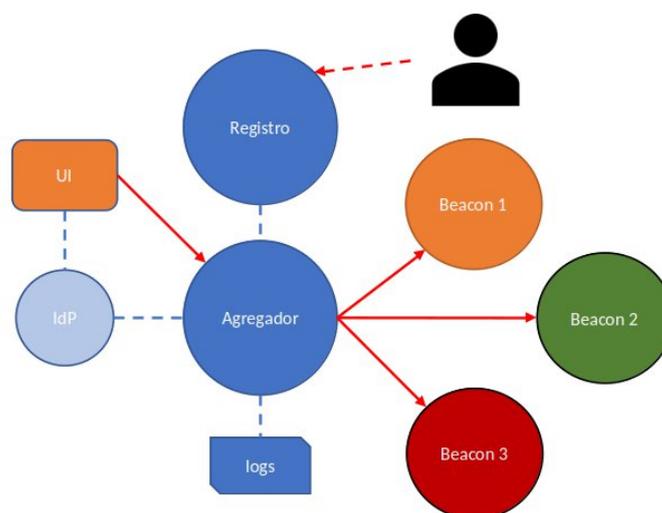


Figura 4. Aspectos técnicos de una Beacon Network. Componentes de una red Beacon. UI: User Interface; IdP: Identity Provider; Aggr: Beacon Aggregator; B1-B3: instancias beacon. Extraído de [4]

A continuación, describiremos algunas consideraciones importantes de los componentes descritos en la Figura 4.

2.3.1 Proveedor de Identidades (IdP)

De acuerdo a las recomendaciones tanto de ELIXIR como de GA4GH [4], se recomienda *OpenID Connect* (<https://openid.net/developers/how-connect-works/>) como el protocolo para el proceso de autenticación. Una vez que el usuario está autenticado, el IdP genera un *token* que es devuelto a la Interfaz de Usuario (UI). De esta manera, en cualquier consulta o *query* realizada por el usuario, la interfaz pasa el *token* al agregador, el cual lo envía a cada *beacon* de la red. Cada *beacon*, debe validar el token contra el IdP original y tomar decisiones relacionadas con el acceso del usuario y los permisos particulares que éste tiene sobre el *beacon* en concreto.

2.3.2 Agregador de beacons (Aggr)

Como ya se ha comentado, el agregador es la entidad que realiza las preguntas a los *beacons* de la red y recopila los resultados. También agrega y *parsea* sus metadatos. Por este motivo, es conocido también como *Meta-Beacon*. Algunas cuestiones a tener en cuenta del agregador en la red son las siguientes:

- Si el *beacon* a consultar requiere autenticación, el agregador debe proporcionar una cabecera con la autenticación del usuario a dicho *beacon*. Si la autenticación no es necesaria, se realiza el siguiente paso.
- El *agregador* *parsea* el *endpoint /map* y comprueba si el *beacon* tiene el esquema del modelo necesario (individuals, biosamples, etc).
- El *agregador* comprueba si los términos de filtrado existen en la consulta.
- En la red, puede haber *beacons* con diferentes esquemas, como *Phenopackets* o diferentes versiones de *Beacon*. Si la red no es capaz de leer estos esquemas, debe eliminar la consulta para el *beacon* en concreto.

2.3.3 Registro de Beacon

Respecto al registro de *Beacon*, su papel principal es comprobar los *beacons* disponibles en la red y recopilar su información básica. La caché del registro almacena la siguiente información sobre los *beacons* de la red:

- Tipos de entradas disponibles para consultar

E5.6. Informe final sobre el funcionamiento de la red IMPaCT de beacons v. 1.0

- Tipo de esquemas para la representación de datos y metadatos
- Términos de filtrado que el *beacon* soporta
- Mecanismo de autenticación/autorización necesarios para garantizar que solamente los usuarios autorizados acceden a los datos.

Esta información debe ser actualizada cada vez que existe un cambio en uno de los *beacons*. En su configuración actual, el Registro de beacons usa el modelo *pull* y actualiza los datos sobre cada *beacon* periódicamente.

2.4 Tipos de Redes Beacon

Podemos distinguir dos tipos principales de redes *Beacon*:

- Redes opacas. Este tipo de red permite al usuario consultar los datos sin conocer su origen. El usuario sabe que los datos están disponibles en la red, pero desconoce de qué *Beacon* procede. Este tipo de red es particularmente útil para casos en los que identificar la fuente de los datos no es necesario o se prefiere un control centralizado del acceso a los *beacons* de la red.
- Redes transparentes. En este tipo de red, la fuente de los datos es conocida. Así, cuando se realizan consultas, la red proporcionará información sobre el *beacon* que contiene los datos. De esta forma, el usuario puede rastrear la fuente de los datos y solicitar acceso al conjunto de datos completo al propietario de esos datos, permitiendo la compartición de los mismos.

2.5 Cómo desplegar una red Beacon

Al margen de los componentes descritos para una *Beacon Network* y debido al éxito de la red de *beacons* de ELIXIR mencionada anteriormente, existe la necesidad de una nueva implementación de redes tipo *Beacon* con nuevas funcionalidades. De esta manera y como consecuencia de la aprobación de *Beacon v2* como un estándar oficial de GA4GH, la nueva versión de *Beacon Network* permite realizar preguntas sobre datos genómicos y fenoclinicos de pacientes a lo largo de las instancias *Beacon* pertenecientes a la red.

Para poder desplegar una red *Beacon v2*, podemos seguir los pasos indicados en <https://beacon-network-v2-documentation.readthedocs.io/en/latest/light-beacon-network/>. En líneas generales:

E5.6. Informe final sobre el funcionamiento de la red IMPaCT de beacons v. 1.0

1. Instalar el *Backend* y el *Frontend* por medio de *Docker* disponible en el repositorio github de Elixir-Europa: <https://github.com/elixir-europe/beacon-network-docker.git> . Alternativamente, se pueden instalar el *Backend* y el *Frontend* de forma separada descargándolos de los siguientes enlaces: <https://github.com/elixir-europe/beacon-network-backend.git> y <https://github.com/elixir-europe/beacon-network-ui.git>—. En cualquiera de las dos alternativas, tendremos creados un *FrontEnd* y un *Backend* para la red *Beacon*.
2. Añadir una instancia *Beacon* a la red. En este caso, tenemos dos opciones dependiendo del rol:
 - a. Usuario: cuando éste es el propietario o el mantenedor de un *beacon* y se desea añadirlo a la red. En este caso, el usuario debe realizar los siguientes pasos:
 - i. Añadir un identificador único del *beacon* a la red.
 - ii. Validar la instancia *beacon*, por medio del validador, disponible en <https://github.com/elixir-europe/neat-beacon-v2-validator/tree/master>
 - iii. Una vez validada la instancia, enviar la instancia *Beacon* a la red, junto con la salida de la validación al administrador de la red *Beacon*
 - b. Mantenedor de la *Red Beacon*: cuando éste es el propietario o el mantenedor de la red *Beacon* y un *beacon* ha solicitado incluir la URL de su API a la red. En este caso, el mantenedor es el responsable de solicitar el fichero de validación del *beacon* para confirmar que la instancia cumple con todos los criterios necesarios para poder añadirlo a la red. Una vez validado, el mantenedor de la red incorporará la URL de la instancia en la red.

3. Red IMPaCT de beacons

Tal y como está descrito en el Plan de Trabajo de IMPaCT-Data (<https://b2drop.bsc.es/index.php/apps/files/?dir=/IMPACT-Data/Documentos%20Generales/Plan%20Estrategico&openfile=556494>), una de las tareas específicas (T5.2) a desarrollar está relacionada con “la implantación de mecanismos para la identificación automática de información genómica relevante o contenido de las imágenes médicas que den respuestas anonimizadas al estilo *Beacon*”. Por este motivo, se ha creado una red ex profeso para IMPaCT-Data de *beacons*. En este apartado, describiremos los nodos incorporados a esa red de *beacons* (genómicos), sus características principales así como los medios técnicos y humanos necesarios para llevarla a cabo.

3.1 Nodos participantes y características

En la tabla 1, listamos los 12 nodos participantes en la *Beacon Network* de IMPaCT

Nodo
Centro Nacional de Análisis Genómico (CNAG)
Navarrabiomed & Nasertic
Institut Germans Trias i Pujol (IGTP)
IIS La Fe
Fundación Progreso y Salud (FPS)
Sant Joan de Déu Research Institute (IRSJD)
Institut d'Investigació en Atenció Primària Jordi Gol (IDIAPJGol)
Centro Nacional de Investigaciones Cardiovasculares (CNIC)
Biobizkaia
Fundación Pública Galega de Medicina Xenómica (FPGMX)
Centro Nacional de Investigaciones Oncológicas (CNIO)
Hospital del Mar and Hospital del Mar Research Institute (IMIM)

Tabla 1. Listado de nodos participantes en la Beacon-Network de IMPaCT-Data

Cada entidad de la tabla 1 hace uso de conjuntos de datos sintéticos o reales, dependiendo de la disponibilidad de los datos o de los consentimientos informados / aprobación de sus comités éticos correspondientes. En la tabla 2 se muestra qué datos sirve cada nodo *Beacon* y su naturaleza (sintéticos o reales).

Nodo	Descripción de los datos	Categoría
CNAG	Datos sintéticos de GPAP (RD-Connect Genome-Phenome Analysis Platform)	S
Navarrabiomed & Nasertic	Dos beacons implementados <ul style="list-style-type: none"> – Datos sintéticos que contienen 2504 muestras del proyecto 1000 genomas y 76 atributos de individuo y datos fenotípicos de UKBiobank – Conjunto de datos de NAGENPediatrics (758 WGS samples), con la intención de incorporar todos los datos generados dentro del programa NAGEN 	S y R

E5.6. Informe final sobre el funcionamiento de la red IMPaCT de beacons v. 1.0

IGTP	Datos de la cohorte GCAT de aproximadamente 20.000 individuos con consentimiento informado, permitiendo el uso de los datos clínicos y genómicos. Se utiliza un conjunto de 808 genomas completos.	R
IIS La Fe	No especificado	S
FPS	Datos de 2105 muestras no relacionadas del servidor de variabilidad española CSVS (http://csvs.clinbioinfospa.es/). Próximamente se incluirán muestras del programa de Enfermedades No Diagnosticadas (ENoD) del CIBERER y de otros proyectos locales como el relacionado con el estudio de variantes genéticas asociadas con riesgo reproductivo en una cohorte de mujeres.	R
IRSJD	Datos sintéticos no especificados	S
IDIAPJGol	Datos sintéticos no especificados	S
CNIC	Variantes cuidadosamente seleccionadas identificadas como <i>drivers</i> en la hematopoyesis clonal de potencial indeterminado	R
Biobizkaia	Datos sintéticos no especificados	S
FPGMX	Pertencia a la red <i>Beacon</i> a través de GPAP en los datos de secuenciación del proyecto IMPaCT-Genómica.	R
CNIO	Datos de una cohorte de datos de metástasis cerebral recopilados en el contexto del proyecto RENACER. La información clínica y de variantes está disponible.	R
IMIM	Datos sintéticos no especificados	S

Tabla 2. Descripción de los datos y su naturaleza de cada instancia Beacon de la red

De los nodos con URL disponibles en la actualidad, las entidades del modelo *Beacon* (sección 1) implementadas con datos por cada uno de los nodos está reflejadas en la tabla 3

Nodo	Entidades con datos
CNAG	Biosamples, Individuals, Genomic Variants
Navarrabiomed & Nasertic	Beacon de datos sintéticos: analysis, biosamples, cohorts, datasets, genomic variants, individuals, runs Beacon NAGEN Pediatrics: analysis, Biosamples, cohorts, datasets, genomic variants, individuals, runs
IIS La Fe	Analysis, Biosamples, Cohorts, Datasets, Genomic Variants, Individuals, Runs

FPS	Genomic Variants, Biosamples
IRSJD	–
IDIAPJGoI	–
CNIC	Genomic Variants
Biobizkaia	Analysis, Biosamples, Cohorts, Datasets, Genomic Variants, Individuals, Runs
FPGMX	–
CNIO	–
IMIM	–

Tabla 3. Entidades implementadas con datos de cada nodo Beacon de la red

3.2 Arquitectura de la Beacon Network de IMPaCT-Data

Tanto el *front-end* como el *back-end* (sección 2.5) están instalados en el *Centre de Regulació Genòmica (CRG)* desde donde se recogen las peticiones de consulta de los usuarios y se recopilan los resultados de cada uno de los nodos *Beacon* y se muestran a dichos usuarios (Figura 5).



Figura 5. Distribución geográfica de la Beacon Network de IMPaCT-Data

3.2.1 Beacon network: *front-end*

La página web de la *Beacon network* creada para IMPaCT-Data, está disponible en el siguiente enlace: <https://impact-beacon-network-demo.ega-archive.org/> (Figura 6)

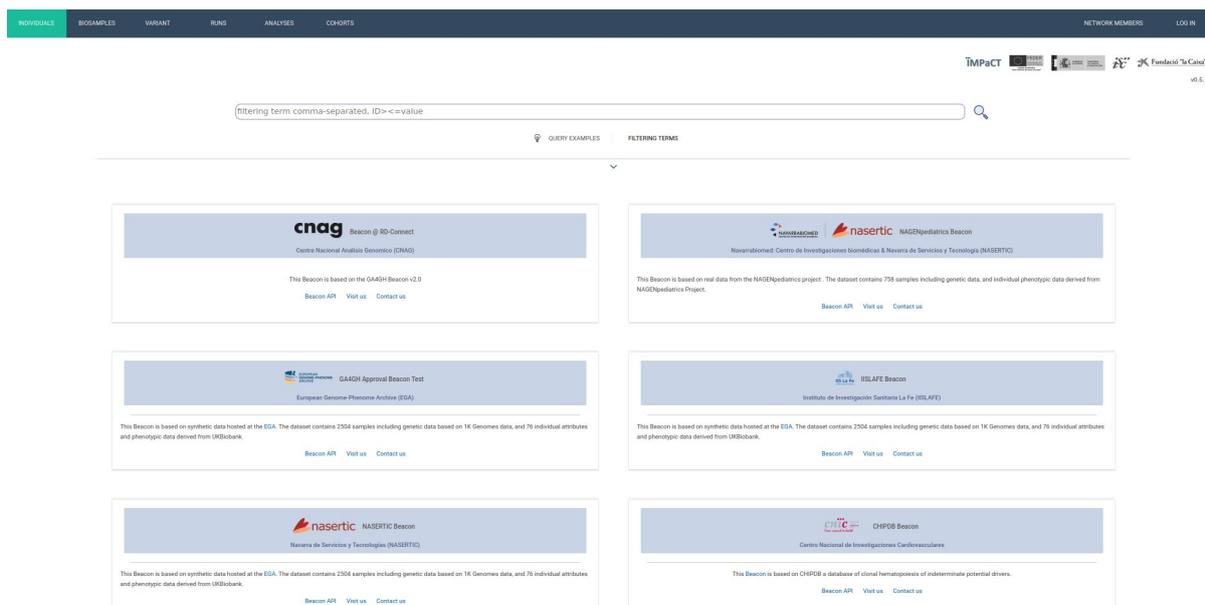


Figura 6. Front-End de la Beacon Network de IMPaCT-Data

Como se puede ver en la figura 6, se pueden realizar consultas a la red *Beacon* teniendo en cuenta las diferentes entidades del modelo: *Individuals*, *Biosamples*, *Variant*, *Runs*, *Analysis* y *Cohorts*. Los resultados obtenidos dependerán de la implementación realizada por cada uno de los nodos (ver Tabla 3).

Si seleccionamos una entidad que no sea *Variants* y *Cohorts*, por ejemplo *Individuals*, hay que definir un criterio de búsqueda para la consulta que queremos lanzar a la *Beacon Network*, es decir, elegir mínimo un concepto para aplicar un filtro de búsqueda. Los conceptos que se pueden aplicar para filtrar nuestra búsqueda se encuentran dentro del apartado *Filtering Terms*, que se puede desplegar pulsando en el botón que lleva este mismo nombre inmediatamente debajo de la barra de búsqueda de la interfaz web (Figura 7).

E5.6. Informe final sobre el funcionamiento de la red IMPaCT de beacons v. 1.0



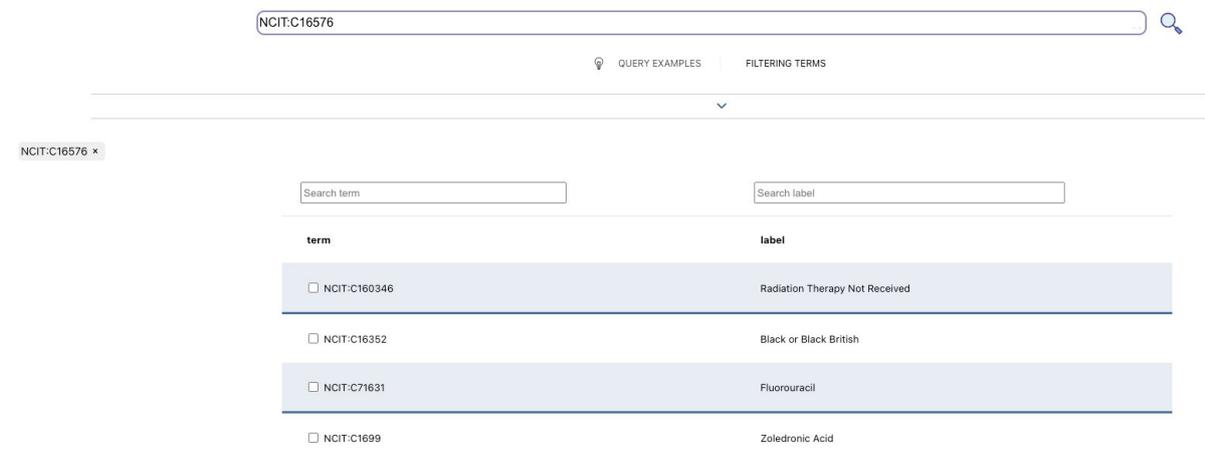
Figura 7. Botón para desplegar listado de filtering terms.

Esta acción desplegará un listado de *filtering terms*, del que podremos seleccionar los términos que sean elegidos para definir nuestra búsqueda, marcando la casilla del *filtering term* correspondiente (Figura 8).

term	label
<input checked="" type="checkbox"/> NCIT:C16576	female
<input type="checkbox"/> NCIT:C160346	Radiation Therapy Not Received
<input type="checkbox"/> NCIT:C16352	Black or Black British
<input type="checkbox"/> NCIT:C71631	Fluorouracil
<input type="checkbox"/> NCIT:C1699	Zoledronic Acid

Figura 8. Ejemplo de selección de un filtering term.

Una vez elegido el *filtering term*, aparecerá etiquetado inmediatamente encima del listado de *filtering terms* y en la barra de búsqueda (Figura 9).



NCIT:C16576

QUERY EXAMPLES | FILTERING TERMS

NCIT:C16576 x

Search term	Search label
term	label
<input type="checkbox"/> NCIT:C160346	Radiation Therapy Not Received
<input type="checkbox"/> NCIT:C16352	Black or Black British
<input type="checkbox"/> NCIT:C71631	Fluorouracil
<input type="checkbox"/> NCIT:C1699	Zoledronic Acid

Figura 9. Ejemplo de adición de un filtering term correctamente seleccionado.

Por otro lado, en este listado de *filtering terms*, se pueden distinguir tres tipos diferentes: ontología, alfanuméricos y *custom* o personalizados. Los primeros son aquellos que sólo muestran una ontología, cuyo significado se encuentra en la columna derecha “*label*”. Los segundos, de tipo alfanumérico, buscan si existe el término introducido en la base de datos y que además, deben tener algún valor asociado a introducir como criterio de búsqueda adicional. Si este *filtering term* está asociado a un valor numérico, aparecerá en “*label*” las unidades por las que se medirá este valor y se podrá aplicar los comparadores de mayor (>) y menor (<), además de igual (=). Posteriormente se debe hacer *click* sobre el símbolo + (Figura 10) para añadirlo al criterio de búsqueda.



Search term

Search label

term	label
<input type="checkbox"/> measures.assayCode:Weight	
<input checked="" type="checkbox"/> Weight	Weight in Kilograms

ID Weight > Value 100 +

Figura 10. Ejemplo de selección de un filtering term numérico.

Si este *filtering term* está asociado a un valor alfanumérico, se tendrá que comparar el término a un texto que tendremos que introducir y sólo podremos aplicar los comparadores de igual (=) o de regex (%) y hacer *click* en el botón *include* (Figura 11).

E5.6. Informe final sobre el funcionamiento de la red IMPaCT de beacons v. 1.0

Search term	Search label
term	label
<input type="checkbox"/> ethnicity:Asian or Asian British	
<input type="checkbox"/> ethnicity:Irish	
<input checked="" type="checkbox"/> ethnicity	
ID ethnicity	= Value African INCLUDE
<input type="checkbox"/> phenotypicFeatures.featureType:M1A	
<input type="checkbox"/> phenotypicFeatures.featureType:M1B	

Figura 11. Ejemplo de selección de un filtering term alfanumérico.

Por último, existen los *filtering terms* del tipo personalizado o *custom*, que se utilizan del mismo modo que los de ontología, sólo hay que seleccionarlos y se añadirán a nuestra barra de búsqueda.

Una vez tengamos los *filtering terms* deseados dentro de nuestra barra de búsqueda, habrá que hacer *clic* en la lupa disponible para ejecutar la consulta (Figura 12) y la *beacon network* devolverá los resultados desglosados por beacon y por conjunto de datos.

Weight>100,NCIT:C16576 

[QUERY EXAMPLES](#) | [FILTERING TERMS](#)

Figura 12. Ejecución de una consulta con los filtering terms seleccionados.

En los resultados (figura 13), aparecerán tres tipos de granularidad para visualizar las respuestas, dependiendo del nivel de permiso que hay en un *dataset* o *beacon* para un usuario en particular. Si seleccionamos la granularidad boolean, la *Beacon Network* desglosará los resultados según se hayan encontrado (*Found*) o no (*Not Found*). Si seleccionamos *count*, se mostrarán resultados por el conteo numérico encontrado y si seleccionamos *Full Response*, podremos visualizar los datos directamente tal y como se encuentran en la base de datos tabulados o en formato JSON.

E5.6. Informe final sobre el funcionamiento de la red IMPaCT de beacons v. 1.0

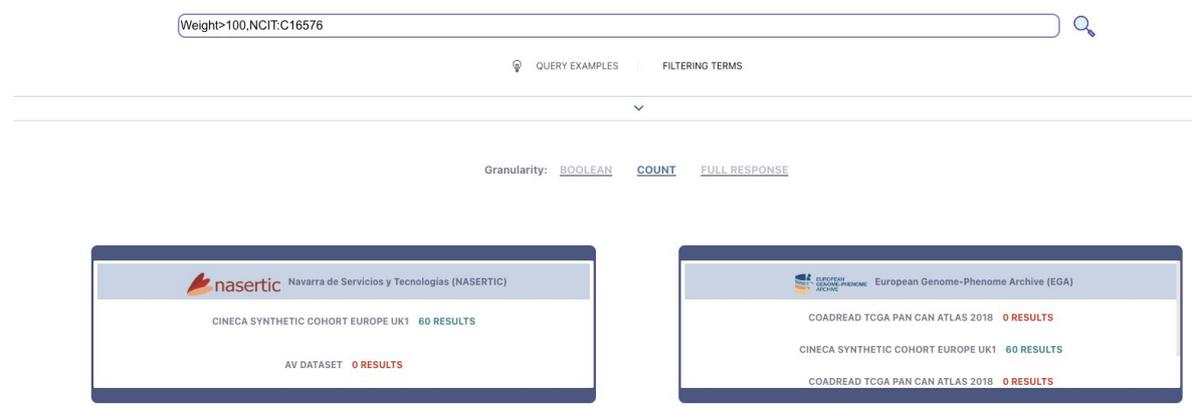


Figura 13. Ejemplo de respuesta en la Beacon Network para la consulta realizada.

Si queremos lanzar una consulta a la Entidad *Variants*, debemos utilizar uno de los tres formularios posibles: *Sequence Queries*, *Range Queries* o *Gene ID Queries*, seleccionando la pestaña correspondiente (Figura 14).

Una vez seleccionado el formulario, podemos introducir cualquiera de los valores que aparecen en el mismo y hacer *clic* en el botón *Search* para ejecutar la nueva consulta en la *Beacon Network*.

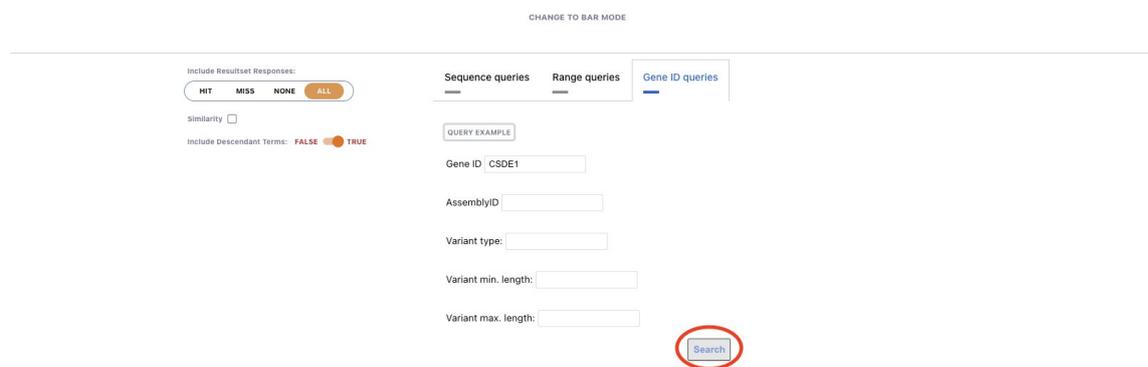


Figura 14. Ejemplo de consulta para la entidad *Variants* introduciendo el nombre del gen que desamos buscar.

3.2.2 Beacon Network: *back-end*

El *backend* de la Beacon Network es el componente principal que se encarga de distribuir las consultas a los diferentes *beacons* de la red y agregar los resultados. Se compone de dos partes: el agregador (2.3.2) y el registro (2.3.3). Además, registra todas las operaciones realizadas en la Beacon Network mediante *logs*.

En esta aplicación, el usuario envía consultas a la API de la *Beacon Network*, ya sea a través del *Frontend* u otros medios, en forma de solicitudes GET o POST. Cuando el *backend* recibe las respuestas de todos los *beacons* consultados, estos se presentan en formato JSON. Si el usuario ha realizado la consulta mediante interfaz de usuario, la respuesta se formateará de manera más visual como se enseña en el apartado 3.2.1.

Actualmente, el *backend* está implementado y desplegado en un servidor [WildFly 30.0.0](#). Está basado en Jakarta RESTful Web Services 3.0 API ([JAX-RS 3.0](#)) y se encuentra en un [repositorio](#) publico propiedad de ELIXIR.

3.2.3 Recursos técnicos y humanos de la *Beacon Network* de IMPaCT

Para desplegar la *Beacon Network*, se requiere al menos una persona con conocimientos técnicos en los siguientes aspectos:

- Sistema de contenedores: la implementación del *Beacon Network*, junto su *backend* y *frontend*, se realiza con un comando de *Docker*.
- Configuración de red: es necesario establecer la dirección IP pública del *frontend* para que sea accesible a las usuarios de la red *Beacon*.
- En caso de ser necesario un servicio de ayuda (*helpdesk*) para asistir a los *beacons* dentro de la red (ya sea para agregar nuevos *beacons* o para solucionar problemas técnicos) se requerirá un correo de referencia o una persona responsable.

En cuanto a los recursos técnicos necesarios, se necesita una máquina virtual, cuyo tamaño dependerá del volumen y tamaño de las consultas que los usuarios realicen, así como el número de usuarios que utilicen la red. El servidor se reserva el 75% de la memoria de la maquina virtual y una consulta realizada a la red *Beacon* no debería exceder unos decenas de mega-bytes de memoria por cliente (Figura 15)

E5.6. Informe final sobre el funcionamiento de la red IMPaCT de beacons v. 1.0

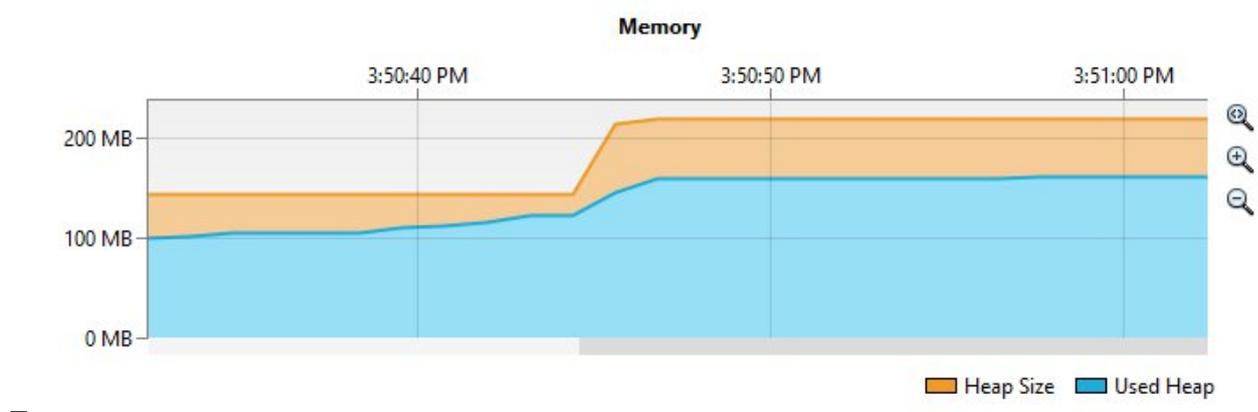


Figura 15. Ejemplo de consumo de memoria asociado una consulta en el agregador de la Beacon Network.

Para la máquina virtual, se recomienda un sistema operativo Unix con Ubuntu 18.04 o superior, con una capacidad mínima de 4GB de RAM y 3GB de memoria libre. La capacidad recomendada son 16GB de RAM y 5GB de memoria libre.

4. Conexión con otros programas

La red IMPaCT de *beacons* descrita en la sección anterior, supone un buen punto de partida para la implementación de redes de *beacons* en otros programas o iniciativas que tienen interés en compartir datos genómicos.

Es el caso del CIBERER, el Centro de Investigación Biomédica en Red de Enfermedades Raras del Instituto de Salud Carlos III (<https://www.ciberer.es/>). El CIBERER es una estructura cooperativa en red, con 62 grupos de investigación ubicados en los principales centros de investigación de toda España. Estos grupos se estructuran en 7 programas científicos:

- Programa de Medicina Genómica Traslacional
- Programa de Medicina Mitocondrial y Metabólica Hereditaria
- Programa de Enfermedad Neurológica
- Programa de Medicina Pediátrica y del Desarrollo
- Programa de Patología Neurosensorial
- Programa de Medicina Endocrina
- Programa de Cáncer Hereditario

E5.6. Informe final sobre el funcionamiento de la red IMPaCT de beacons v. 1.0

Numerosos grupos de investigación, dentro de sus líneas de trabajo, realizan la secuenciación masiva de exomas o genomas y muchos de estos datos genómicos son depositados en CSVS, la base de datos de variabilidad genómica española (<http://csvs.clinbioinfospa.es/>). Ocurre por ejemplo con las muestras de secuenciación dentro del programa ENoD (Enfermedades no Diagnosticadas) del CIBERER, donde se depositan en CSVS. Tal y como se indicó en la Tabla 2, CSVS es descubrible a través de la *Beacon Network* (<https://beacon-network.org/>) mediante una instancia de *Beacon*.

Sin embargo, es demandada en la comunidad del CIBERER la necesidad de compartir datos genómicos de forma adicional a través de instancias *Beacon* desplegadas en varios nodos e integrarlos a través de una red *Beacon* específica del CIBERER.

Por ello, dentro del Grupo de Trabajo de Bioinformática del CIBERER del año 2024, existe un paquete de trabajo específico para “Establecer sistemas colaborativos de explotación y consulta de variantes”, donde se pondrá en valor la experiencia adquirida en la *Beacon Network* del programa IMPaCT de forma que se pueda establecer una línea de trabajo dentro del CIBERER para la implantación de una *red Beacon* específica en esta organización.

5. Conclusiones

En conclusión, este documento repasa qué es un sistema *Beacon* y describe de forma más profunda qué es una red de *beacons* y sus especificaciones técnicas para su implantación. Se describe también la red IMPaCT de beacons: nodos participantes, datos que sirven, características técnicas del *back-end* y el *fron-end*, así como conexión de esta red de *beacons* con otros programas de investigación.

Referencias

1. Fiume M, Cupak M, Keenan S, et al. Federated discovery and sharing of genomic data using beacons [published correction appears in *Nat Biotechnol.* 2019 Mar 20;:]. *Nat Biotechnol.* 2019;37(3):220-224. doi:10.1038/s41587-019-0046-x
2. Rambla J, Baudis M, Ariosa R, et al. Beacon v2 and Beacon networks: A "lingua franca" for federated data discovery in biomedical genomics, and beyond. *Hum Mutat.* 2022;43(6):791-799. doi:10.1002/humu.24369
3. Peña-Chilet M, Roldán G, Perez-Florido J, et al. CSVS, a crowdsourcing database of the Spanish population genetic variability. *Nucleic Acids Res.* 2021;49(D1):D1130-D1137. doi:10.1093/nar/gkaa794
4. Rambla J, Fromont L, López-Doriga O, Repchevsky D, Aguiló-Castillo S, Michael Baudis M . *Service architecture deliverable, ELIXIR Beacon Infrastructure.* 2024. <https://docs.google.com/document/d/1BLa1H5rsLILz2MXMHBXnCFS8HI2dp04Sbj7Ov9wkpDY/edit> (disponible bajo demanda)

Acrónimos y Abreviaturas

VCF	Variant Call Format
API	Application Programming Interfaces
CSVS	Collaborative Spanish Variation Server
ENoD	Enfermedades no Diagnosticadas
CIBERER	Centro de Investigación en Red de Enfermedades Raras
JSON	JavaScript Object Notation
GA4GH	Global Alliance for Genomics and Health
IMPACT	Infraestructura de Medicina de Precisión asociada a la Ciencia y la Tecnología
IMPACT-Data	Programa de ciencia de datos de IMPACT