



DESARROLLO DE UN PROYECTO DE BIG DATA EN SALUD PÚBLICA: PROYECTO MIDAS

(Meaningful Integration of Data Analytics and Services)

Roberto Bilbao, Oihana Belar, Iker Ezkerra

Fundación Vasca de Innovación e Investigación Sanitarias, Biobanco Vasco

Joseba Bidaurrezaga

Departamento de Salud del Gobierno Vasco, Subdirección de Salud Pública y Adicciones de Bizkaia

Pilar Amiano

*Departamento de Salud del Gobierno Vasco, Subdirección de Salud Pública y Adicciones de Gipuzkoa, Unidad Vigilancia Epidemiológica.
Instituto de Investigación Sanitaria Biodonostia, Grupo de Epidemiología de las Enfermedades Crónicas y Transmisibles*

Nicolás González

Osakidetza, Dirección General, Subdirección de Informática y Sistemas de Información

Carlos Sola

Osakidetza, Dirección General, Subdirección asistencia sanitaria

Gorka Epelde, Mónica Arrúe

Vicomtech, eSalud y Aplicaciones Biomédicas. Biodonostia, Área de Bioingeniería, Grupo de eSalud

Joao Pita Costa, Flavio Fuart, Luka Stopar

Quintelligence

1 ¿Qué es el Big Data?

La **sociedad crea y almacena diariamente ingentes cantidades de datos** que no pueden ser analizadas por las tecnologías existentes.

Las **entidades gestoras** de esta información se enfrentan a un reto: **cómo analizarla para crear valor**, ya sea económico en las organizaciones con ánimo de lucro u ofrecer un mejor servicio en las que no lo tienen, teniendo en cuenta el volumen de los datos, su heterogeneidad y rápido crecimiento¹.

Extracción de información de valor mediante sistemas analíticos avanzados soportados por computación en paralelo

El **dominio en la gestión y análisis de esta gran cantidad de información es lo que en realidad trata Big Data** o analítica avanzada de datos. Según la **Agencia Española de Protección de Datos y la Asociación Española para el Fomento de la Seguridad de la información**², el Big Data o analítica avanzada de datos se define como el conjunto de tecnologías, algoritmos y sistemas empleados para recolectar datos a una escala y variedad no alcanzada hasta ahora y a la extracción de información de valor mediante sistemas analíticos avanzados soportados por computación en paralelo.

El término **"Big Data"** se ha abierto camino en **numerosos ámbitos**, especialmente en aquellas organizaciones que dependen en gran medida de los datos y en el intercambio de información. **El control, la gestión y el análisis de esta información ha permitido el desarrollo de empresas, Google, Facebook...** que determinan en gran medida nuestras vidas. Esta oportunidad es **multisectorial** e incluye a los **sistemas de Salud** que precisamente originan diariamente multitud de datos de diferente tipología (incluidos los datos del propio sistema sanitario, pacientes, industria y proveedores).

En este sentido, la **principal fuente de datos** es la **historia clínica de cada paciente** pero actualmente se puede ampliar a otras fuentes de información adicionales, en parte de fuera de las propias organizaciones, como por ejemplo los datos de los ciudadanos recogidos a través de **aplicaciones móviles relacionadas con la Salud (mhealth), las redes sociales o la industria sanitaria**, que se podrían utilizar para refinar los modelos para ver qué está pasando, qué va a pasar y por qué³.

En Salud, se estima que el Big Data ofrecerá nuevas posibilidades para **entender mejor la biología de sistemas en los humanos** así como para desarrollar herramientas de **medicina personalizada, previsión de enfermedades** y en definitiva, **mejorar la Salud de las personas**^{4,5,6}.

Asimismo, se espera que el Big Data suponga un gran **ahorro de los sistemas sanitarios** gracias a la mejora de la coordinación de la **atención al paciente** y la **reducción de la ineficiencias** administrativas y clínicas⁷.



Las principales fuentes de datos son la historia clínica, mhealth, las redes sociales y la industria sanitaria



2 Big Data y Salud pública

La **Salud pública** es una especialidad no clínica de la Medicina enfocada a la promoción, prevención e intervención de la Salud desde una perspectiva multidisciplinaria y colectiva⁸.

En este sentido, sus funciones son principalmente la gestión, vigilancia y mejoramiento del nivel de Salud en la población, así como la prevención, el control y la erradicación de enfermedades. Además, se encarga de desarrollar políticas públicas, garantizar el acceso y el derecho al sistema sanitario, crear programas educativos, administración de servicios e investigación⁹.

Son **muchos los factores** que se combinan para **afectar la Salud de individuos y comunidades**. La Salud de las personas viene determinada por su **circunstancia y entorno**. En gran medida, factores como el lugar donde vivimos, el estado de nuestro entorno, la genética, nuestros ingresos y nivel de educación, y nuestras relaciones con amigos y familiares, tienen un impacto considerable en la Salud, mientras que los factores más comúnmente considerados como el acceso y uso de servicios de sanitarios tiene menos impacto del que se le supone.

Los **determinantes de la Salud** incluyen:

- Entorno social y económico
- Entorno físico
- Características y comportamientos individuales de la persona



Mucha de esta **información no se recoge en las historias clínicas** de los pacientes, es labor de la **Salud pública recoger esta información y analizarla** para dirigir las intervenciones adecuadas para la mejora de la Salud de la comunidad.

La revista *Lancet Public Health* ha publicado, en 2019, un editorial donde afirma que el Big Data podría tener una **aplicación en el ámbito de la Salud pública para proveer la intervención más eficiente** a la población adecuada en el periodo de tiempo más propicio¹⁰.

Actualmente, las intervenciones de Salud pública tienden a ser complejas, dependientes del contexto y potencialmente costosas¹¹. **No es fácil definir los parámetros de la intervención y evaluar su eficacia como lo es para una intervención clínica** para la cual se usa el ensayo clínico controlado aleatorizado. Además, una intervención puede no funcionar tan bien como se pretendía debido a debilidades en el diseño o porque no se implementó como se pretendía¹²⁻¹⁴. Recientemente, se postula que los datos recogidos de forma habitual en la gestión de las entidades sanitarias podrían complementar los estudios tradicionales de cohorte o de encuestas *ad hoc* para poder generar mayor información que ayude a la toma de decisiones^{10,15}.

El desarrollo de nuevos abordajes de Big Data en el que se pueda separar la información relevante de la inmensa cantidad de datos existente es un reto que puede traer un gran avance en Salud pública¹⁶



Objetivos del proyecto MIDAS: Big Data en Salud pública

En 2016, se creó el consorcio MIDAS (*Meaningful Integration of Data, Analytics and Services*) (www.midasproject.eu) para intentar dar **respuesta a retos de Salud pública** en diferentes países a través de herramientas de Big Data gracias a la **financiación de la Comisión Europea** en el marco del programa H2020 (SC1-PM-18 *Big Data Supporting Public Health Policies*).

El consorcio*, liderado por la Universidad del Ulster, está formado por representantes de **4 sistemas socio-sanitarios: Finlandia, Irlanda del Norte, República de Irlanda y Euskadi** junto con **socios tecnológicos** (empresas, PYMES y centros tecnológicos). En Euskadi, participan Osakidetza (Servicio Vasco de Salud), el Departamento de Salud del Gobierno Vasco y la Fundación Vasca de Innovación e Investigación Sanitarias-BIOEF junto con el centro tecnológico Vicomtech.

Cada **sistema socio-sanitario ha planteado problemas y preguntas** asociadas para intentar encontrar nuevos conocimientos a través del proyecto:

- **República de Irlanda**

- **Diabetes.**

Se busca analizar una foto global de la enfermedad para reservar recursos dedicados y distribuidos de forma eficiente a luchar contra esta epidemia.

- **Irlanda del Norte**

- **Niños en situación de desprotección infantil.**

Se pretende identificar intervenciones que puedan prevenir que esos niños y jóvenes pasen a una situación de desprotección infantil y tengan que estar bajo la tutela de la administración. Además, se quiere conocer la Salud y las necesidades de estos niños que entran y salen del amparo de los servicios sociales.

- **Finlandia**

- **Salud mental y abuso de estupefacientes en población joven.**

Cómo incrementar las fuentes de información de las Unidades de Salud Mental y Asuntos Sociales para promover políticas de prevención.

- **Euskadi**

- **Obesidad infantil.**

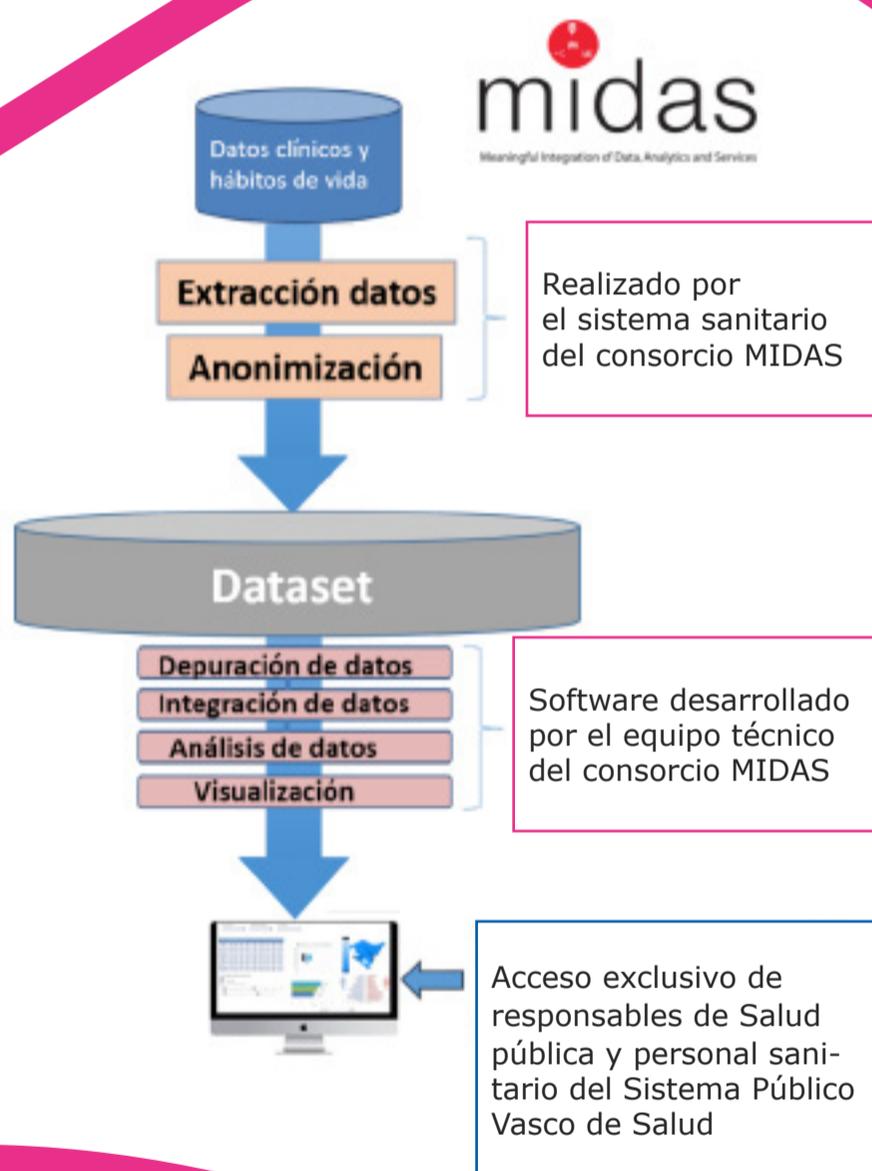
Se pretende conocer la etiología de la enfermedad e identificar áreas de intervención.

* Ver composición del consorcio en www.midasproject.eu/partners/

Los **socios tecnológicos del consorcio desarrollan una plataforma tecnológica de Big Data** que posteriormente se adapta a cada sistema sanitario de forma individualizada por lo que no se cruzan bases de datos entre sistemas. La plataforma MIDAS permite mapear, adquirir, gestionar, modelar, procesar y explotar los datos de Salud existentes junto con otras fuentes de datos abiertos (*open data*) para crear información procesable basada en la evidencia. Los usuarios finales de estas plataformas son los profesionales, gestores y responsables sanitarios.

La **ejecución, desarrollo e implementación** de la plataforma acabará en **febrero de 2020**.

La plataforma MIDAS permite mapear, adquirir, gestionar, modelar, procesar y explotar los datos de Salud



4 Creación de la Plataforma MIDAS en Euskadi

En Euskadi, el Departamento de Salud del Gobierno Vasco identificó la **obesidad infantil**¹⁷ como un **problema de Salud pública en el Plan de Salud 2013-2020**, que es el instrumento superior de planificación y programación del sistema¹⁸.

Por ello, los socios del consorcio MIDAS en Euskadi decidieron tomar la obesidad infantil como el objetivo principal a analizar con las herramientas a desarrollar con una visión de que el problema tiene una **influencia multifactorial**¹⁹.



En Euskadi, el Departamento de Salud del Gobierno Vasco identificó la obesidad infantil como un problema de Salud pública en el Plan de Salud 2013-2020

Así, el diseño de la plataforma MIDAS se centró en la creación de una herramienta de Big Data que permitiera **visualizar las posibles correlaciones en el ámbito de la obesidad infantil** y así encontrar información oculta, patrones recurrentes, nuevas correlaciones, etc. El **reto** ha sido doble: **generar la base de datos con variables asociadas a la obesidad infantil** bajo el marco legal vigente de protección de datos, a partir de información existente en la historia clínica y, por otro lado, **crear el software** que permitiera capturar, curar, analizar, almacenar, visualizar, etc., los datos para obtener conocimiento. Además, se ha planteado **producir herramientas para incluirlas en la plataforma**, como un chatbot capaz de mantener conversaciones con personas para evaluar la percepción en las redes sociales sobre una política pública concreta y un sistema inteligente de búsquedas de noticias, referencias científicas y aplicaciones que extraen inteligencia de las noticias mundiales y de los artículos publicados en MEDLINE.

MIDAS se centró en la creación de una herramienta de Big Data que permitiera visualizar las posibles correlaciones en el ámbito de la obesidad infantil

El **desarrollo de la plataforma** en Euskadi se ha realizado llevando a cabo los siguientes **pasos**:

1 Identificación de las variables asociadas a la obesidad infantil

La creación de la **base de datos con información almacenada en la historia clínica ha sido una tarea multidisciplinar** en la que han participado diferentes tipos de profesiones: pediatras de atención primaria y especializada de Araba, Bizkaia y Gipuzkoa, así como representantes de Salud pública del Departamento de Salud del Gobierno Vasco, expertos en Big Data, gestores informáticos y responsables de calidad de la historia clínica. Como resultado, se ha generado una base de datos de **120 variables** relacionadas con la obesidad infantil de personas con **historia clínica de edad igual o menor a 18 años** en el año 2000. Se ha calculado que el tamaño muestral es de **890.000** personas.



2 Garantía ético-legal del proyecto

En este proyecto, se estableció el diseño la **privacidad** a través de la adopción de medidas técnicas y organizativas apropiadas, con el fin de garantizar el cumplimiento de los requisitos del **Reglamento General de Protección de Datos (RGPD)** que regula la protección de los derechos y libertades de las personas físicas con respecto al tratamiento de datos personales.

En este sentido, el **proyecto MIDAS** fue presentado y aprobado por el **Comité ético de Investigación Clínica de Euskadi**²⁰. Por otro lado, el consorcio MIDAS cuenta con un **EPAG** (*Ethical and Privacy Advisory Group*) que es un comité interno que evalúa la idoneidad del uso de las bases de datos que se utilizan dentro de un proyecto y que también aprobó la ejecución del proyecto en Euskadi. Además, se ha contado con la colaboración de la **Agencia Vasca de Protección de Datos** como criterio experto.



Además MIDAS cuenta con un EPAG y la colaboración de la Agencia Vasca de Protección de Datos

3 Extracción datos de historia clínica y anonimización

Personal de Osakidetza realizó una carga de información a partir de **herramientas de extracción, y la correspondiente anonimización** de los datos siguiendo las instrucciones derivadas del análisis de riesgos según la metodología descrita por la **Agencia Española de Protección de Datos**²¹. Así, los datos fueron completamente anonimizados y, por lo tanto, no son datos de carácter personal.

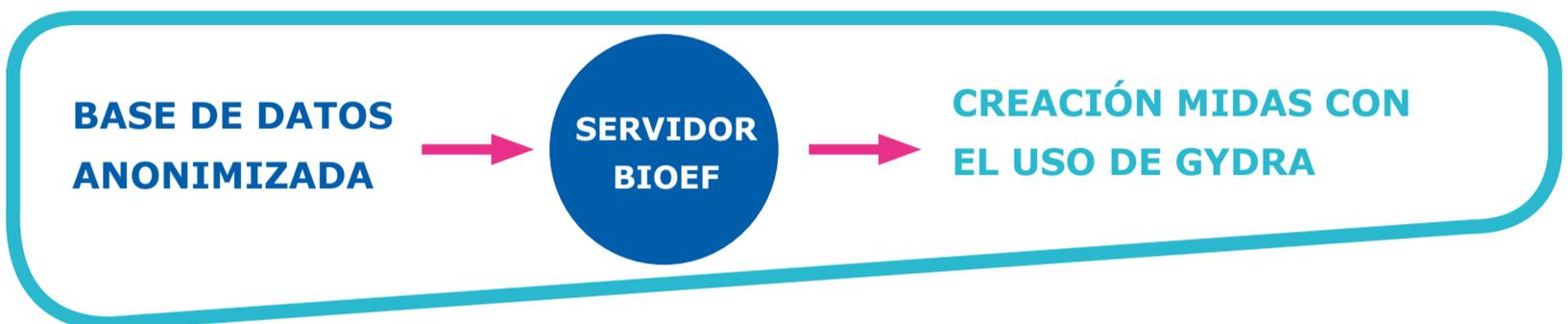
En este sentido, la anonimización conlleva que **no es posible identificar al donante de la información**, teniendo en cuenta todos los medios que puedan ser razonablemente utilizados.

MIDAS cuenta con herramientas de extracción y la correspondiente anonimización de los datos



4 Creación de la plataforma MIDAS: software-base de datos

La **base de datos anonimizada** fue transferida al servidor de **BIOEF** para la creación de la plataforma MIDAS junto con el software desarrollado por el consorcio. Para esta tarea se ha desarrollado **GYDRA** que es la herramienta implementada para el tratamiento, curación y carga de los datos en el proyecto MIDAS. GYDRA trabaja sobre el sistema de archivos distribuidos **HDFS** (común en entorno Big Data).



Las **funcionalidades** que nos permite esta herramienta son:

1. Estadística general del set de datos

que hemos subido, para obtener un resumen con el tamaño del set de datos, número de registros que contiene, número de variables, características, tipos, etc.

2. Información sobre las variables del set de datos:

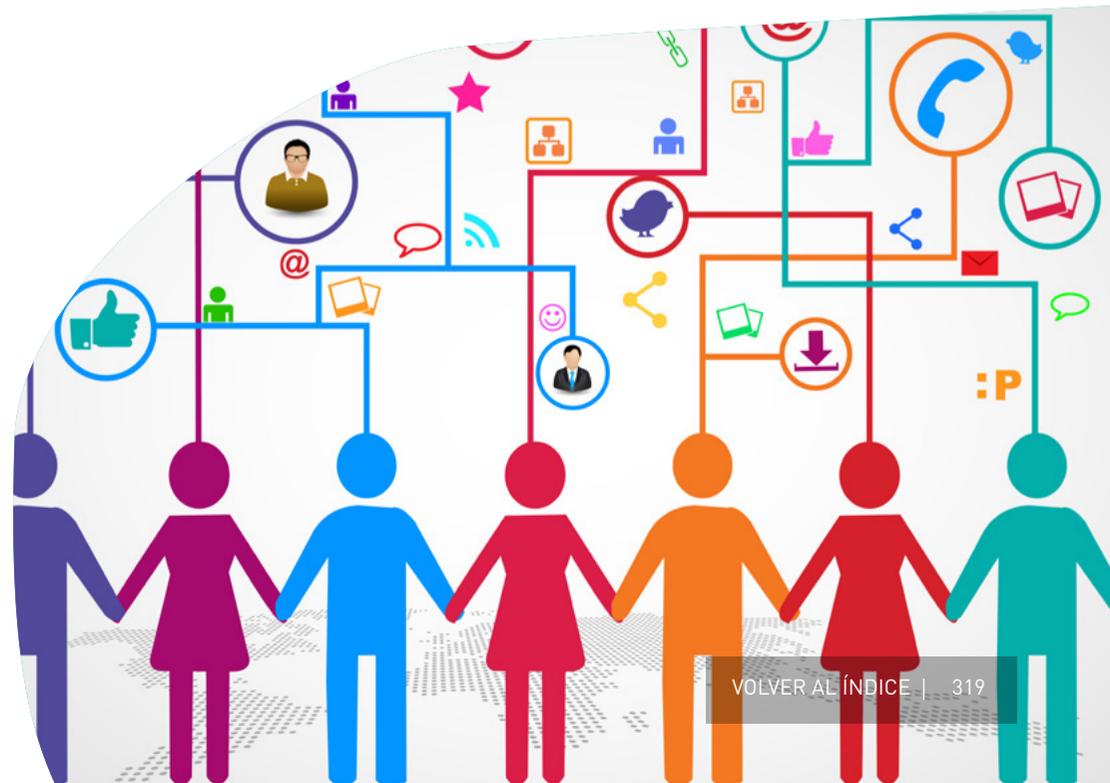
valor máximo, valor mínimo o los valores diferentes que se recogen en el set de datos; valores nulos, se puede ver si alguna variable recogida tiene todos sus valores nulos y un histograma por variable.

3. Correlaciones para los campos numéricos

para ver el coeficiente de correlación de Spearman que existe entre ellos, esto permite ver el grado de correlación que existe entre las variables del set de datos.

4. Identificación de

outliers, para las variables numéricas sin valores nulos el sistema indica qué valores pueden ser valores atípicos sobre los que se debe poner atención.



5. Mejora de la calidad de los datos. Además de obtener la información descrita anteriormente se puede transformar los datos con el fin de mejorar la calidad de los mismos antes de subirlos a la plataforma MIDAS, las transformaciones que podemos hacer son:

- a. Eliminar registros específicos.
- b. Eliminar variables específicas.
- c. Renombrar variables.
- d. Modificar un valor concreto de una variable.
- e. Reemplazar caracteres de una variable.
- f. Realizar operaciones aritméticas con variables numéricas.
- g. Reemplazar valores nulos por la media de esa variable.
- h. Reemplazar valores nulos por la mediana de esa variable.
- i. Reemplazar valores nulos por una constante.
- j. Reemplazar valores *outlier*.
- k. Establecer una variable como numérica.

Una vez **realizadas las transformaciones se generó la base de datos** para cargarlos en la plataforma MIDAS, listo para ser **analizados**.

En paralelo a la fase de curación de los datos, éstos fueron descritos siguiendo una metodología de metadatos establecida para el proyecto. La herramienta **GYDRA permite la alineación de los datos con sus metadatos** durante la fase de curación de datos. Una vez finalizada la curación de datos, la herramienta GYDRA **facilita la carga de los datos, junto con sus metadatos a la plataforma MIDAS**.

La **plataforma MIDAS** se compone de **3 capas**:



Para la **capa de gestión de datos** la plataforma MIDAS hace uso de **Apache Hive**. Apache Hive es una infraestructura de almacenamiento de datos construida sobre HDFS que permite el análisis de grandes conjuntos de datos almacenados mediante un lenguaje de consultas basado en SQL llamado **HiveQL**. Una vez los datos han sido curados se procede a cargarlos junto a sus metadatos en Hive.

Para la **capa de analítica de datos**, se planteó un servicio web que, la capa de visualización, pregunta qué analíticas hay disponibles. La capa de analítica de datos provee **dos tipos de analíticas: genéricas y específicas**. Las **analíticas genéricas** plantean las analíticas descriptivas y visualizaciones posibles en base a los metadatos que se hayan subido a la plataforma. Las **analíticas específicas** son aquellas que se han desarrollado para un piloto específico, en el caso de Euskadi, para el contexto de la obesidad infantil. Las analíticas de datos han sido programadas en Python, tanto usando librerías de análisis de datos no-distribuidos tradicionales, así como utilizando Apache Spark y sus librerías de cómputo distribuido para escalar las soluciones de analítica a grandes cantidades de datos.



Se planteó un servicio web que permite a la capa de visualización preguntar qué analíticas hay disponibles

OPTION 1

OPTION 2

OPTION 3

La solución de visualización presenta en un asistente las opciones de visualización y análisis que la capa de analítica de datos le haya informado como posibles

Por último, para la **capa de visualización** se ha planteado una solución de generación de **dashboards web**, que permite al usuario definir y personalizar según sus necesidades un cuadro de mando. Esta solución de visualización se ha planteado para que sea lo más automática posible. Para ello, presenta en un asistente las opciones de visualización y análisis que informados como “posibles” por la capa de analítica de datos. El usuario es capaz de configurar un cuadro de mando compuesto por widgets de analíticas personalizables y configurables en tamaño y posición.

Además de las analíticas de datos introducidas, el sistema de visualización permite **añadir widgets** de soluciones complementarias desarrolladas en MIDAS para la definición y monitorización de políticas de Salud pública como monitorización de noticias en medios de comunicación, monitorización de publicaciones científicas o resultados de campañas de contraste en redes sociales.

5 Otros elementos inteligentes de la plataforma

Los crecientes **desafíos y oportunidades** de las instituciones de Salud pública en la era de **Big Data revelan necesidades y problemas** que hay que abordar para **mejorar su eficiencia y aprovechar el conocimiento disponible** (a menudo en tiempo real y a menudo de acceso libre). En ese sentido, los datos abiertos tienen un papel destacado debido a su fácil acceso y potencial para complementar otras fuentes de datos propietarias. MIDAS permite la integración de estas fuentes de datos heterogéneas, agregando valor a los datos propietarios de los Servicios de Salud.

Este conjunto de herramientas propuestas por MIDAS se centra en **dos grandes conjuntos de datos abiertos: "noticias mundiales" y MEDLINE**, proporcionando soluciones integradas que se pueden personalizar fácilmente según las necesidades de los usuarios.

MIDAS permite la integración de estas fuentes de datos heterogéneas, agregando valor a los datos propietarios de los Servicios de Salud



El conjunto de herramientas propuestas ponen a disposición:

- **Panel de control** con funcionalidad general y varios módulos de visualización
- **Versión pública** de la misma que se puede compartir fácilmente mediante una **URL** e integrarlo a través de, por ejemplo, *"iframe"*
- **Widget personalizado** integrado en el panel principal de MIDAS

Además, permite a los **usuarios**:

- **Aumentar el poder informativo** de sus propios datos propietarios
- **Explorar otros aspectos** de los mismos en un contexto global/local en función de las nuevas capas de información a las que se puede acceder
- **Tener una toma de decisiones** basada en la evidencia más potente obtenida gracias a bases de conocimiento confiables

Esta tecnología utiliza **herramientas web semánticas** que extraen información de dos grandes conjuntos de datos abiertos: las **noticias mundiales**, en más de **60 idiomas**, y la base de conocimientos de investigación biomédica **MEDLINE**, donde se dispone de información comentada de más de **26 millones de artículos**.



**NOTICIAS
MUNDIALES EN
> 60 IDIOMAS**

**MEDLINE CON
> 26 MILLONES
ARTÍCULOS**

También permite la **integración del conocimiento de estos dos conjuntos de datos para habilitar un sistema de monitoreo inteligente** que se complementa con los datos propietarios del Instituto de Salud Pública.

Las tecnologías descritas se basan en herramientas ya establecidas, algunas de ellas ya en el mercado. Una selección de estas herramientas también se puede desplegar en un contexto de datos propietarios, se ofrecen a través de **APIs** que permiten a los propios sistemas informáticos del usuario integrarlas sin problemas.



La **manipulación y visualización** de una fuente de datos tan completa como MIDAS, plantea **desafíos**, especialmente en la búsqueda, revisión y presentación eficientes de los conocimientos científicos apropiados. Una de las necesidades identificadas que motivan este trabajo es asegurar la **disponibilidad de un cuadro de mando dinámico** que permita al usuario explorar los módulos de visualización de datos, representando las consultas al conjunto de datos de MEDLINE a través de gráficos circulares, gráficos de barras, etc. El cuadro de mando de MEDLINE se alimenta de ese conjunto de datos a través de la herramienta **elasticSearch**.



Una de las necesidades identificadas que motivan este trabajo es asegurar la disponibilidad de un cuadro de mando dinámico

Éste se compone de varios módulos de visualización interactiva que utiliza el cursor del ratón para interactuar y proporcionar información a través de mensajes *“mouse-over”* sobre varios aspectos de los datos basados en consultas de particular interés (p. ej. un gráfico circular que representa las citas de “Salud pública” que hablan de “obesidad infantil” durante un período de tiempo seleccionado).

Este panel **MEDLINE**, tal y como fue conceptualizado e implementado en la plataforma MIDAS, **sirve al usuario menos técnico para explorar los datos disponibles** (sobre un subconjunto de los datos generados por un tema de interés).

Existen otras opciones que permiten un **mayor control de los datos por parte de los científicos** de datos a un nivel más operativo:

- **Panel de gestión**, donde el usuario técnico puede realizar el sub-muestreo adecuado basado en los temas de interés y las opciones avanzadas opcionales sobre las características de los datos disponibles
- **Creador de módulos visuales** que permite al usuario menos técnico crear fácilmente nuevos módulos interactivos de visualización
- **Cuadro de mando**, que se puede configurar a través de *“iframe”* como una ventana en tiempo real para las personas que tienen que tomar decisiones que le permite controlar el estado de las cuestiones planteadas y que son representados en cada módulo de visualización

En la plataforma MIDAS se ha añadido un chatbot para "oír" la opinión que las redes sociales pueden tener sobre un tema sin inferir opinión y que pueda ayudar a conocer la aceptación social de una propuesta

Se ha creado un formulario para crear **preguntas de opción múltiple que se presenta a un conjunto de personas** que han dado su consentimiento para ser encuestadas. Cada opción se completa automáticamente utilizando **Dbpedia**. Los metadatos que se obtienen nos permiten entrenar al bot para detectar la opción en texto de forma libre. Esto significa que se puede hacer **preguntas de forma libre a los encuestados** en lugar de pedirles que elijan "a", "b" o "c". Esto da a los respondedores la oportunidad de **añadir explicaciones adicionales** que se puedan analizar, o que el bot haga preguntas posteriores para obtener más detalles. Los chatbots pueden identificar las respuestas de opción múltiple que se espera ver y detectar cuando una persona que ha sido preguntada menciona algo inesperado. Así, se permite la oportunidad de hacer **preguntas de seguimiento** y obtener información potencialmente valiosa que se habría perdido en una encuesta tradicional.



Se ha creado un formulario para crear preguntas de opción múltiple que se presenta a un conjunto de personas que han dado su consentimiento para ser encuestadas

Da a los respondedores la oportunidad de añadir explicaciones adicionales que se pueden analizar



5 Conclusiones

- El concepto de **inteligencia artificial** en el análisis de datos ha llegado para **transformar la medicina**^{4,21}, pero Salud pública no ha estado en primera fila en la aceptación de su potencial¹⁰.
- Ahora es un **momento crítico para reflexionar sobre esta revolución digital** y lo que puede significar para la gestión e implantación de nuevas políticas de Salud pública.
- Consciente de este reto, la **Comisión Europea marcó en su programa marco de investigación H2020 (SC1-PM-18 *Big Data Supporting Public Health Policies*)**, la necesidad de desarrollar **nuevas estrategias** para vincular fuentes de datos heterogéneas, estudiar los **mecanismos causales** y **mejorar las previsiones** de desarrollo espacial y temporal de la enfermedad, poder elaborar **informes innovadores** para poder mejorar las metodologías actuales de estratificación de riesgo, poder **evaluar la eficiencia y la eficacia de las estrategias** aplicadas y **retroalimentar los resultados**, y **analizar la eficiencia de la gestión del paciente** tanto a nivel de la Atención Primaria (prevención y detección temprana) como en el hospital.
- En definitiva, se busca **alinear los datos, macrodatos y los métodos avanzados de simulación** para proporcionar a los gestores de Salud pública herramientas de análisis de políticas de alto impacto en el marco de una serie de **desafíos epidemiológicos**.
- El proyecto **MIDAS en Euskadi** se ha postulado desde su origen como una nueva e innovadora aproximación para afrontar los mencionados retos, proporcionando una herramienta para **evaluar y tomar acciones locales**, basadas en evidencia objetiva.
- **MIDAS no pretende redefinir, desde la base, las estrategias políticas existentes relacionadas con la obesidad y dirigidas a la prevalencia de la obesidad infantil sino apoyar a la epidemiología tradicional** en la manera en que son diseñadas y desplegadas en la sociedad.

- Actualmente, el proyecto MIDAS ha superado el ecuador de su ejecución a través de la **cadena de valor del Big Data**, habiendo desarrollado un sistema de gobernanza del dato para asegurar la **transparencia ante la sociedad y de gestión de la calidad del dato**, así como el software para correlacionar datos reales de la práctica asistencial con fuentes de datos abiertos (*open data*).
- Durante los **próximos años comenzaremos la fase de evaluación** de la utilización y aplicabilidad de la plataforma a las necesidades de los profesionales de Salud pública.

Los autores de este trabajo quieren agradecer la aportación al proyecto MIDAS de los siguientes profesionales de Osakidetza: Mikel Ogueta, M^a Pilar Aizpurua Galdeano, Francisco Javier Núñez; Jesus Arranz, Iratxe Salcedo, Ainhoa Zabaleta, Inmaculada Moro, Miren Garrastatxu Landaluce, Jose Luis Bilbao Madariaga.

Bibliografía

1. Institute, The McKinsey Global. Big data: The next frontier for innovation, competition, and productivity. 2011.
2. Agencia Española de Protección de Datos y Asociación Española para el Fomento de la Seguridad de la información. Código de buenas prácticas en protección de datos para proyectos Big data. 2017.
3. Big Data Application in Biomedical Research and Health Care: A Literature Review. Luo J, Wu M, Gopukumar D, Zhao Y. 2016, Biomed Inform Insights, Vol. 8, págs. 1-10.
4. From Big Data to Precision Medicine. Hulsén T, Jamuar S, Moody A, Karnes J, Varga O, Hedensted S, Spreafico R, Hafler DA, McKinney EF. 2019, Front Med, Vol. 6, págs. 34-40.
5. Using predictive analytics and Big data to optimize pharmaceutical outcomes. Belle A, Thiagarajan R, Soroushmehr SM, Navidi F, Beard DA, Najarian K. 2017, Am J Health Syst Pharm, Vol. 74, págs. 1494-1500.
6. Big Data and medicine: a Big deal? Mayer-Schönberger V, Ingelsson E. 2018, J Intern Med, Vol. 283, págs. 418-429.
7. Big Data Management in US Hospitals: Benefits and Barriers. Schaeffer C, Booton L, Halleck J, Studeny J, Coustasse A. 2017, Health Care Manag (Frederick), Vol. 36, págs. 87-95.
8. WHO/Europe | Public health services. [En línea] [Citado el: 1 de 5 de 2019.] <http://www.euro.who.int/en/health-topics/Health-systems/public-health-services>
9. Ten great public health achievements--United States, 2001-2010. (CDC), Centers for Disease Control and Prevention. MMWR Morb Mortal Wkly Rep, Vol. 60, págs. 619-23.
10. Next generation public health: towards precision and fairness. Heath, The Lancet Public. 2019, Vol. 4, pág. 209.
11. Criteria for evaluating evidence on public health interventions. Rychetnik, L., Frommer, M., Hawe, P., & Shiell, A. 2002, Vol. 56, págs. 119-127.

12. Evaluating the Health Impact of Large-Scale Public Policy Changes: Classical and Novel Approaches. Basu S, Meghani A, Siddiqi A. 2017, *Annu Rev Public Health*, Vol. 38, págs. 351-370.
13. Assessing the Public's Health. Younger DS, Moon-Howard J. 2016, *Neurol Clin*. 2016 Nov;34(4):1057-1070, Vol. 34, págs. 1057-1070.
14. A conceptual framework for implementation fidelity. Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. 2007, *Implementation science*, Vol. 2, págs. 40-49.
15. How has Big data contributed to obesity research? A review of the literature. Kate A. Timmins, Mark A. Green, Duncan Radley, Michelle A. Morris, Jamie Pearce. 2018, *Epidemiology and Population Health*, Vol. 42, págs. 1951-1962.
16. Big data meets public health. Muin J. Khoury, John P. A. Ioannidis. 2014, *Science*, Vol. 346, págs. 1054-1055.
17. Dirección de Salud Pública y Adicciones del Gobierno Vasco. Estrategia de prevención de la obesidad infantil en Euskadi. 2019.
18. Departamento de Salud del Gobierno Vasco. Políticas de Salud para Euskadi 2013-2020. 2013.
19. Prevention of diabetes in overweight/obese children through a family based intervention program including supervised exercise (PREDIKID project): study protocol for a randomized controlled trial. Arenaza L, Medrano M, Amasene M, Rodríguez-Vigil B, Díez I, Graña M, Tobalina I, Maiz E, Arteche E, Larrarte E, Huybrechts I, Davis CL, Ruiz JR, Ortega FB, Margareto J, Labayen I. 2017, *Trials*, Vol. 18, págs. 372-384.
20. Comité ético de investigación clínica de Euskadi. [En línea] 1 de 5 de 2019. https://www.osakidetza.euskadi.eus/r85-pkfarm03/es/contenidos/informacion/ceic_ensayos_clinicos/es_ceic/ensayos_clinicos.html
21. Agencia Española de Protección de Datos. Orientaciones y garantías en los procedimientos de anonimización de datos personales. 2016.