



midas

Meaningful Integration of Data, Analytics and Services

Grant Agreement No. 727721

Proyecto MIDAS: una experiencia de explotación de datos en salud para investigación

Roberto Bilbao
BIOEF



This project is funded by
the European Union

H2020-SC1-2016-CNECT

SC1-PM-18-2016 - Big Data Supporting Public Health Policies



<http://www.midasproject.eu>



@MIDASProjectEU



Big Data in the MIDAS context

- **Big data:** data sets that are so **large** or **complex** that **traditional data processing applications are inadequate to deal** with them.
- **Challenges:** analysis, capture, data curation, search, sharing, storage, transfer, visualization, querying, updating and information privacy.

(https://en.wikipedia.org/wiki/Big_data)



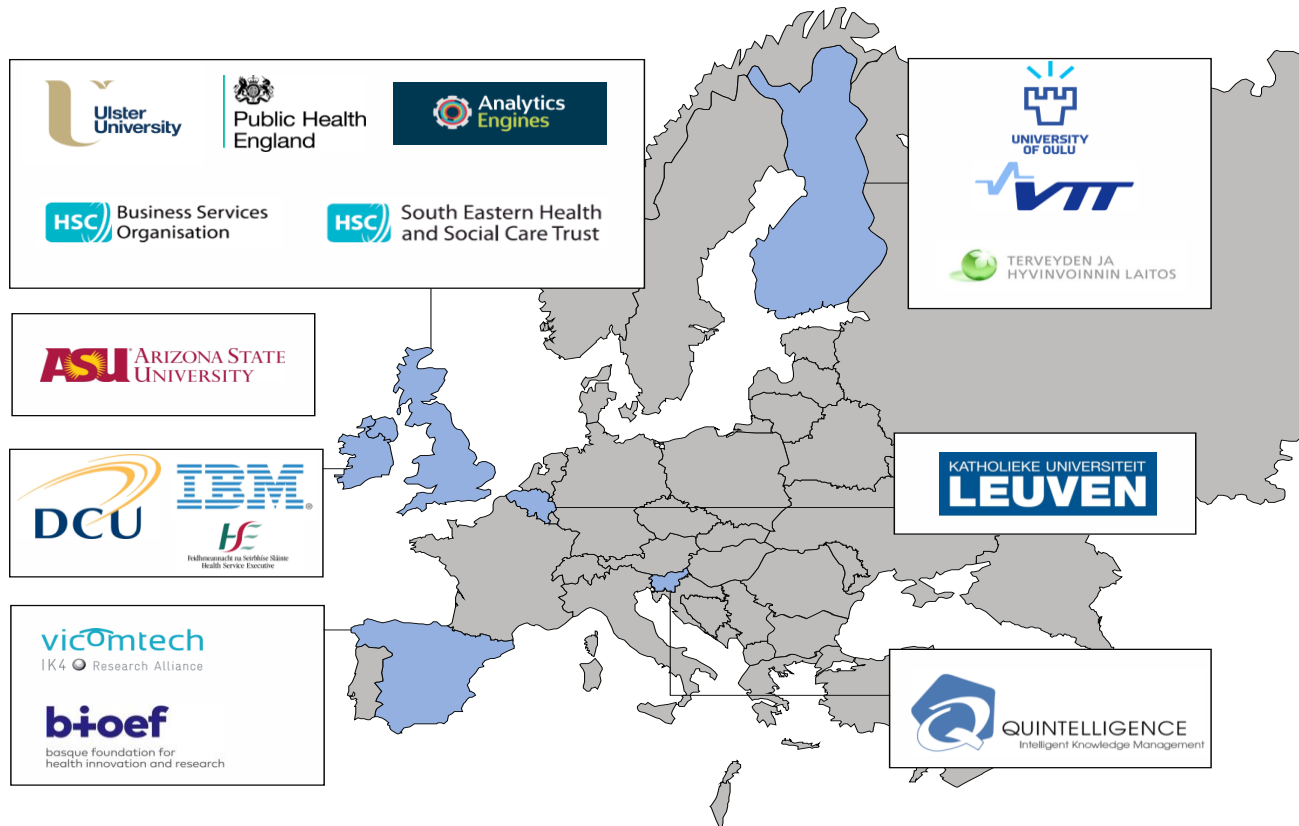
- **Funding: Horizon 2020**
- **Objetivo de la convocatoria:**
SC1-PM-18-2016: Big Data supporting Public Health policies

Scope: how to better acquire, manage, share, model, process and exploit the huge amount of data to develop integrated solutions that support public health authorities



Consorcio MIDAS

Grant Agreement No. 727721



This project is funded by
the European Union



<http://www.midasproject.eu>



@MIDASProjectEU



**Extracción de
datos**

**Depuración de
datos**

**Integración de
datos**

Análisis de datos

Visualización


midas

Meaningful Integration of Data, Analytics and Services



This project is funded by
the European Union



<http://www.midasproject.eu>



@MIDASProjectEU





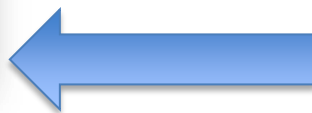
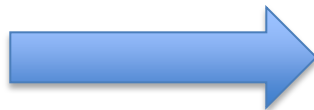
Panel customizado con:

- Panel de análisis masivo de datos
- Valoración de respuesta ante una propuesta-chatbot (Twitter)
- Buscador especializado de medline y artículos de periódico



¿Cuál es la pregunta? entender la etiología y los espacios de intervención en salud pública para la obesidad infantil en Euskadi

De acuerdo con el plan estratégico de Osakidetza 2017-2020



De acuerdo con el plan estratégico del departamento de salud 2017-2020



This project is funded by the European Union



<http://www.midasproject.eu>



@MIDASProjectEU

3. KPIs at Basque Pilot

Grant Agreement No. 727721

- What drives the childhood obesity?
 - We focused on understanding the etiology of the childhood obesity in the Basque Country and identifying areas of intervention
- Who are the customers of the MIDAS platform?
 - The customers will be pediatricians, epidemiologist and policy makers.
- What do we want to do to drive our KPI in the right direction?
 - MIDAS research core group will follow the strategic plan of the Basque Health Department to tackle child-obesity, analytical and monitoring tools needed to focus the interventions and to measure results.
 - There are some mandatory reporting requirements within the Basque Midas pilot:
 - To provide and clearly identify both crude and adjusted rates.
 - To provide information at different granularity levels, adequate for each stakeholder.
 - To cover time, place, and person analysis for epidemiological studies



This project is funded by
the European Union



<http://www.midasproject.eu>



@MIDASProjectEU

Sistema sanitario vasco en MIDAS



This project is funded by
the European Union



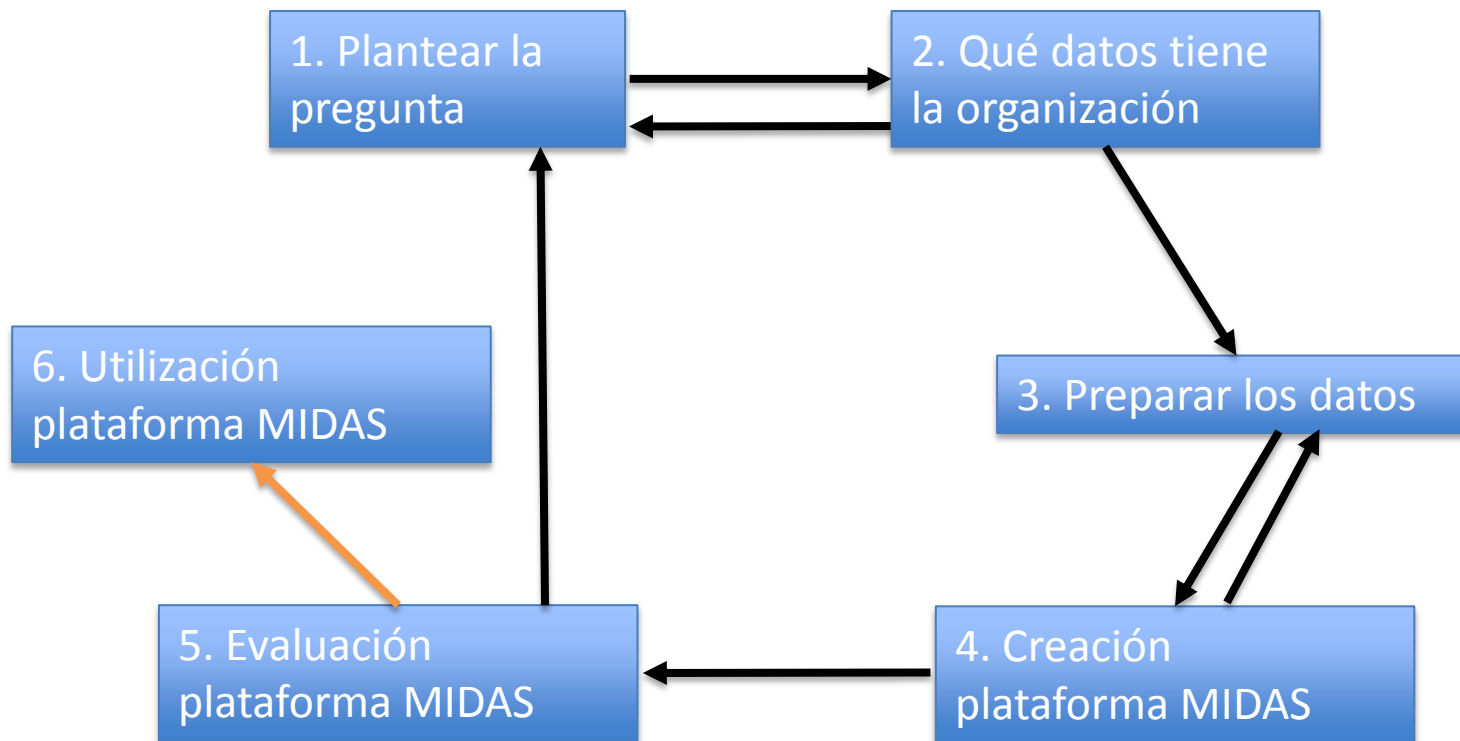
<http://www.midasproject.eu>



@MIDASProjectEU

Ciclo del proyecto MIDAS

Grant Agreement No. 727721



This project is funded by
the European Union



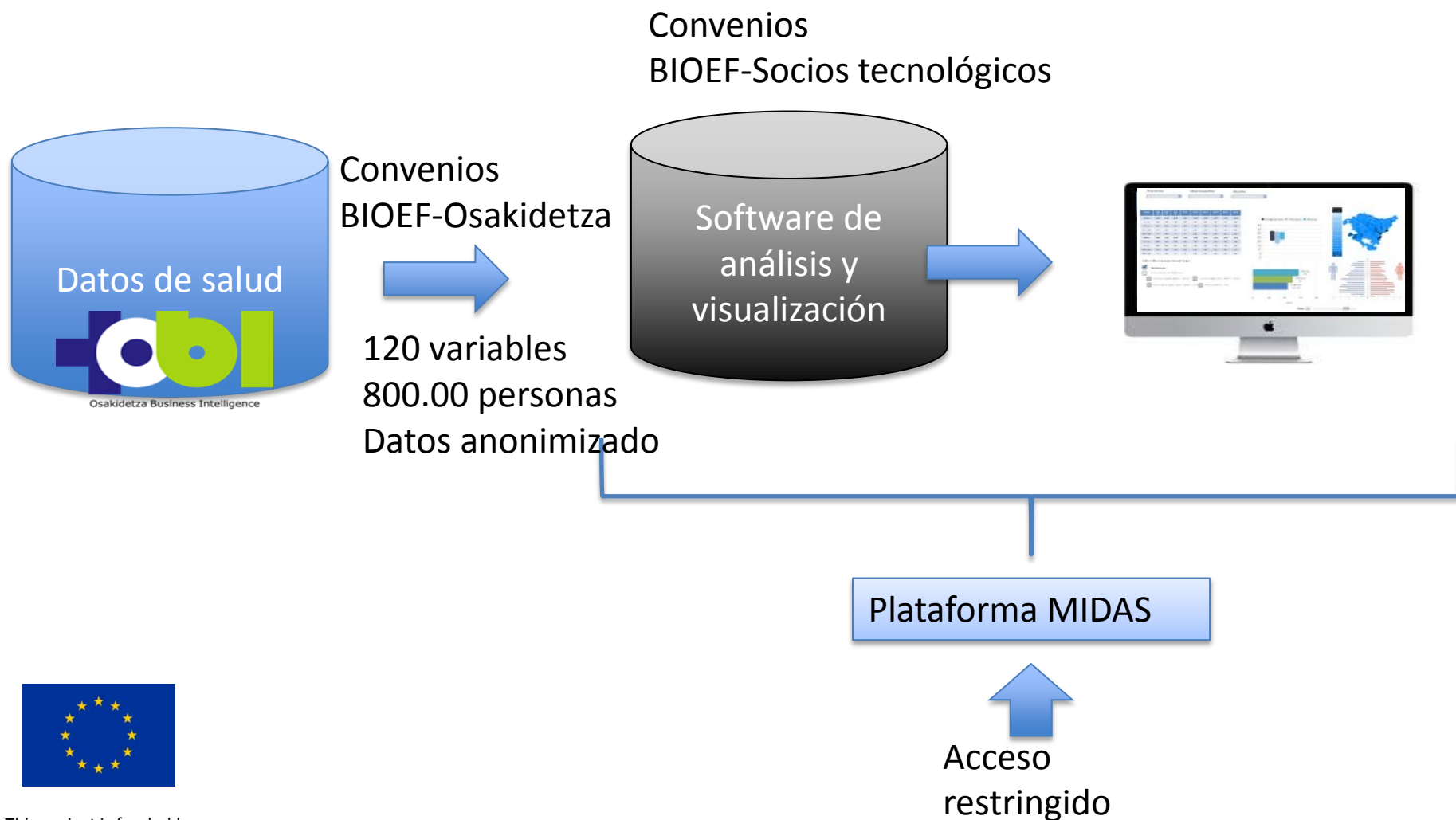
<http://www.midasproject.eu>



@MIDASProjectEU

Creación de una plataforma MIDAS con datos anonimizados de salud

Grant Agreement No. 727721



This project is funded by
the European Union



<http://www.midasproject.eu>



@MIDASProjectEU

Privacidad desde el diseño

Grant Agreement No. 727721

Presentación proyecto al Comité de etica de Euskadi.



Elaboración de un análisis de riesgos



Presentación al EPAG (comité de ética interno de MIDAS)



Firma acuerdo entre BIOEF y Osakidetza para la transferencia de datos anonimizados



Firma acuerdo entre BIOEF y socios tecnológicos del proyecto MIDAS



This project is funded by
the European Union



<http://www.midasproject.eu>



@MIDASProjectEU



Datuak Babesteko Euskal Bulegoa
Agencia Vasca de **Protección de Datos**

Análisis de riesgo

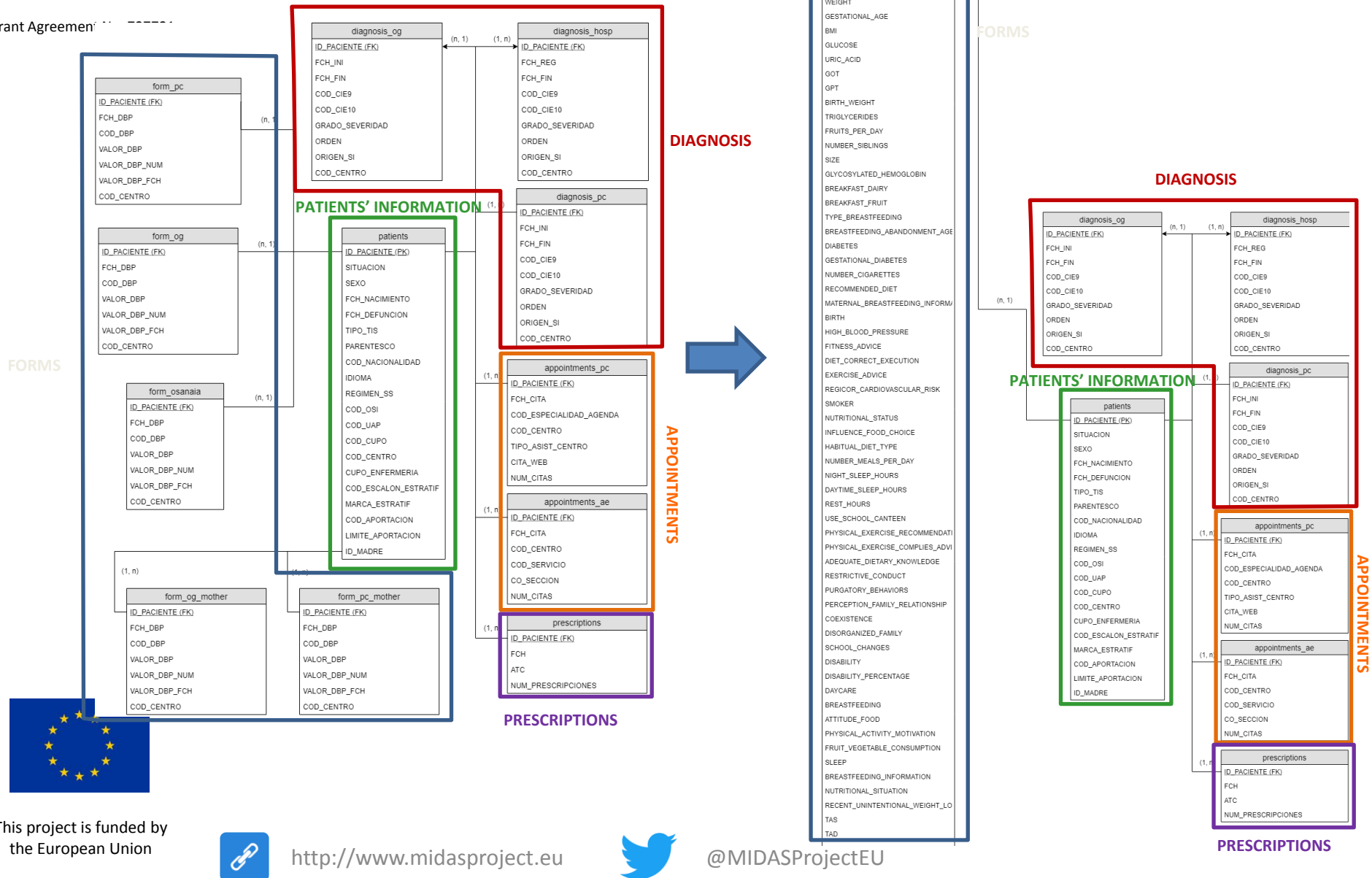
Grant Agreement No. 727721

			GRAVEDAD		
			BAJO 1	MEDIO 2	ALTO 3
PROBABILIDAD	MUY ALTA	5	5	10	15
	ALTA	4	4	8	12
	MEDIA	3	3	6	9
	BAJA	2	2	4	6
	MUY BAJA	1	1	2	3

	Riesgo muy grave. Requiere medidas preventivas urgentes. No se debe iniciar el tratamiento de datos sin la aplicación de medidas y garantías urgentes y sin acotar sólidamente el riesgo.
	Riesgo importante. Medidas preventivas obligatorias. Se deben controlar las variables de riesgo previamente o durante el tratamiento de datos.
	Riesgo apreciable. Estudiar si es posible introducir medidas preventivas para reducir el nivel de riesgo. Si no fuera posible, mantener las variables controladas.
	Riesgo marginal. Se vigilará aunque no requiere medidas preventivas de partida.



BASQUE dataset structure





LA K-ANONIMIDAD COMO MEDIDA DE LA PRIVACIDAD

I. RESUMEN

Esta nota técnica se dirige a responsables y encargados de tratamiento que aborden procesos de anonimización sobre conjuntos de datos. En una realidad en la que fuentes de datos independientes se interconectan y que, por diseño, pueden compartir atributos comunes, cabe la posibilidad de crear un rastro electrónico de los individuos, incluso cuando se hayan suprimido los datos que explícitamente les identifican, pudiendo llegar a establecerse vínculos entre dichas fuentes de información y constituir así una amenaza para la privacidad de los interesados cuyos datos están sujetos a tratamiento.

En aplicación del principio de Responsabilidad Proactiva establecido en el Reglamento (UE) 2016/679 General de Protección de Datos, el responsable debe abordar el estudio del riesgo inherente de reidentificación de los sujetos de los datos e implementar las medidas para gestionarlo. El objetivo de dicho análisis es alcanzar un balance correcto entre la necesidad de obtener unos resultados con una determinada fidelidad y el coste que el tratamiento puede tener para los derechos y libertades de los ciudadanos.

En la presente nota se desarrolla una de las posibles técnicas para gestionar el riesgo de reidentificación conocida como k-anonimización.

II. INTRODUCCIÓN

La Directiva 95/46 en su considerando 26 establecía que, para determinar si una persona era identificable, era necesario considerar el conjunto de los medios que pudieran ser razonablemente utilizados por el responsable del tratamiento, o por cualquier otro, para identificar a dicha persona. De esta forma, dejaban de ser aplicables los principios de protección de datos en aquellos casos en los que el conjunto de datos fuera hecho "anónimo" o disociados de manera tal que ya no fuera posible identificar al interesado.

En la misma línea, el considerando 26 del RGPD señala que datos personales "seudonimizados" constituyen información sobre una persona física a partir de la cual es posible realizar su identificación dentro de una probabilidad razonable teniendo en cuenta medios y factores objetivos, así como los costes, el tiempo y la tecnología necesarios para materializar su identificación.

La diferencia del término utilizado en ambas normas ha evolucionado desde una limitada "anonimización" a una materialización de ésta en el término "seudonimización" del RGPD donde viene a ponerse de manifiesto la dificultad de conseguir, en la actualidad, una anonimización perfecta o que garantice, en términos absolutos, el enmascaramiento de la identidad de las personas. No obstante, a lo largo del presente documento utilizaremos el término anonimización con independencia de que la identificación del sujeto de datos sea o no reversible en mayor o menor grado.

El tratamiento masivo de datos procedentes de los ciudadanos mediante el uso de técnicas basadas en Big Data, Inteligencia Artificial o Machine Learning obliga a la

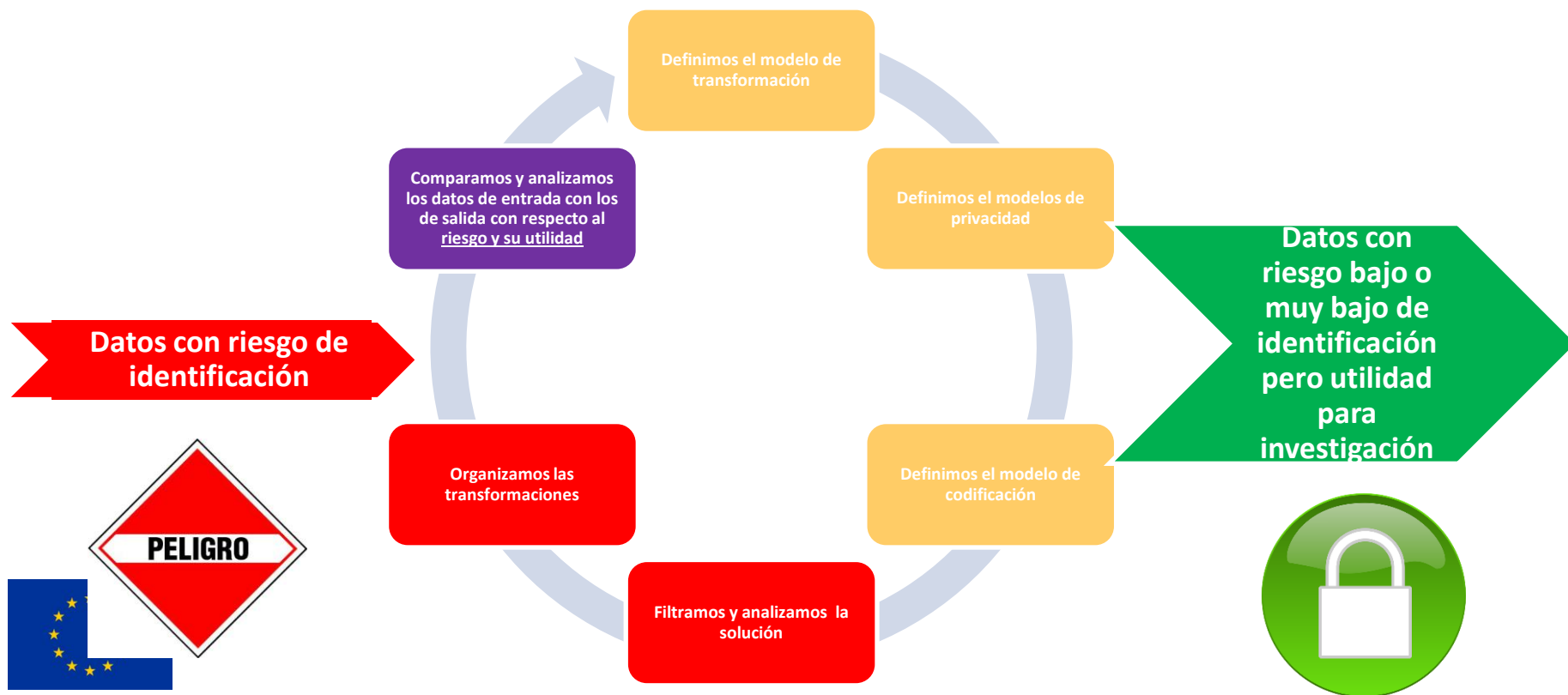
La k-anonimidad trata de resolver este problema: **“Dados unos datos estructurados con datos personales, poder asegurar con **garantía científica**, que en una nueva versión modificada de estos datos no se puedan re-identificar los individuos a los que se refieren, a la vez que los datos sigan siendo útiles en la práctica”.**

Nombre	Edad	Género	Provincia	Religión	Enfermedad
*	20 < Edad ≤ 30	Mujer	Madrid	*	Cáncer
*	20 < Edad ≤ 30	Mujer	Bizkaia	*	Infección Viral
*	20 < Edad ≤ 30	Mujer	Madrid	*	Tuberculosis
*	20 < Edad ≤ 30	Varón	Cantabria	*	Ninguna enfermedad
*	20 < Edad ≤ 30	Mujer	Bizkaia	*	Cardiológica
*	20 < Edad ≤ 30	Varón	Cantabria	*	Tuberculosis
*	Edad ≤ 20	Varón	Bizkaia	*	Cáncer
*	20 < Edad ≤ 30	Varón	Cantabria	*	Cardiológica
*	Edad ≤ 20	Varón	Bizkaia	*	Cardiológica
*	Edad ≤ 20	Varón	Bizkaia	*	Infección Viral

Vamos a transformar estos datos en datos 2-anónimos (k=2) para los atributos Edad, Género y Provincia.

LA K-ANONIMIDAD COMO MEDIDA DE LA PRIVACIDAD <https://www.aepd.es/media/notas-tecnicas/nota-tecnica-kanonimidad.pdf>

ARX es un software de código abierto que nos ayuda en la anonimización de datos personales. Admite una amplia variedad de **(1) modelos de privacidad y riesgo**, **(2) métodos para transformar datos** y **(3) métodos para analizar la utilidad de los datos de salida**.

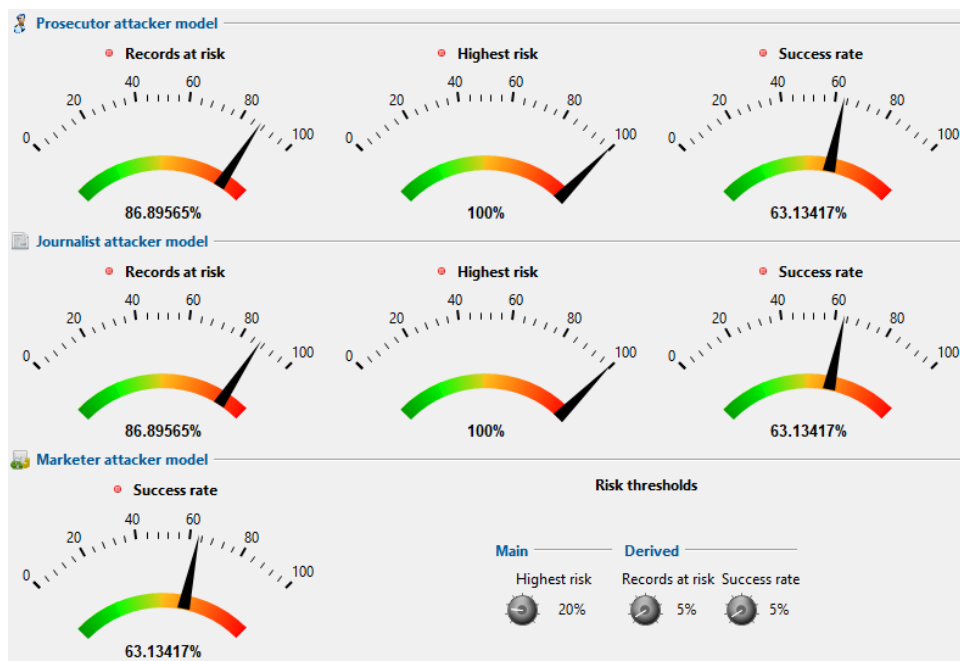


Resultados para k = 10

Grant Agreement No. 727721

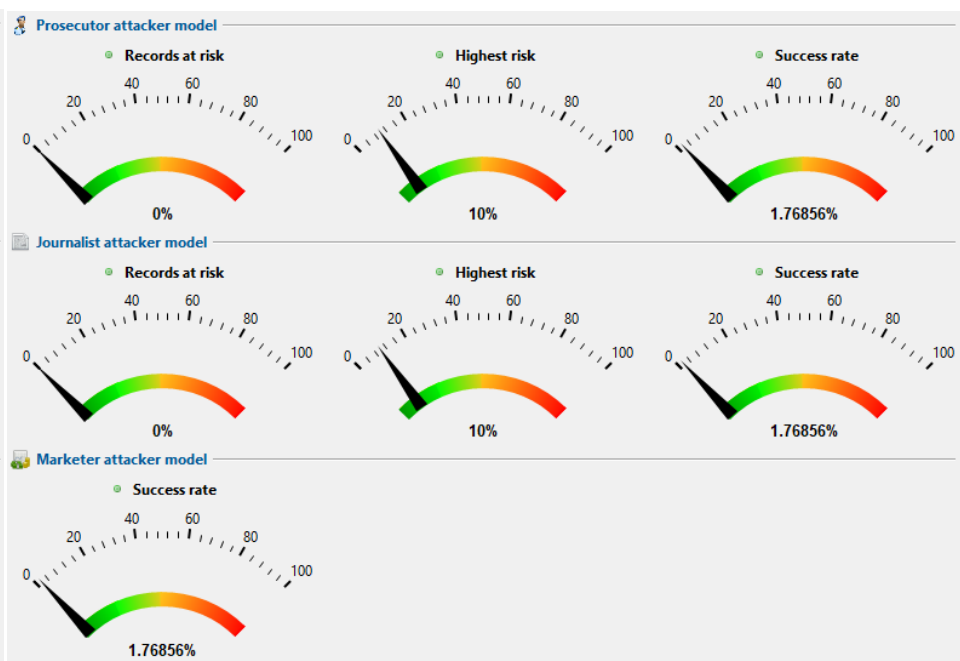
Después de varias iteraciones conseguimos reducir el riesgo de re-identificación de forma muy significativa

Análisis de riesgo dataset original



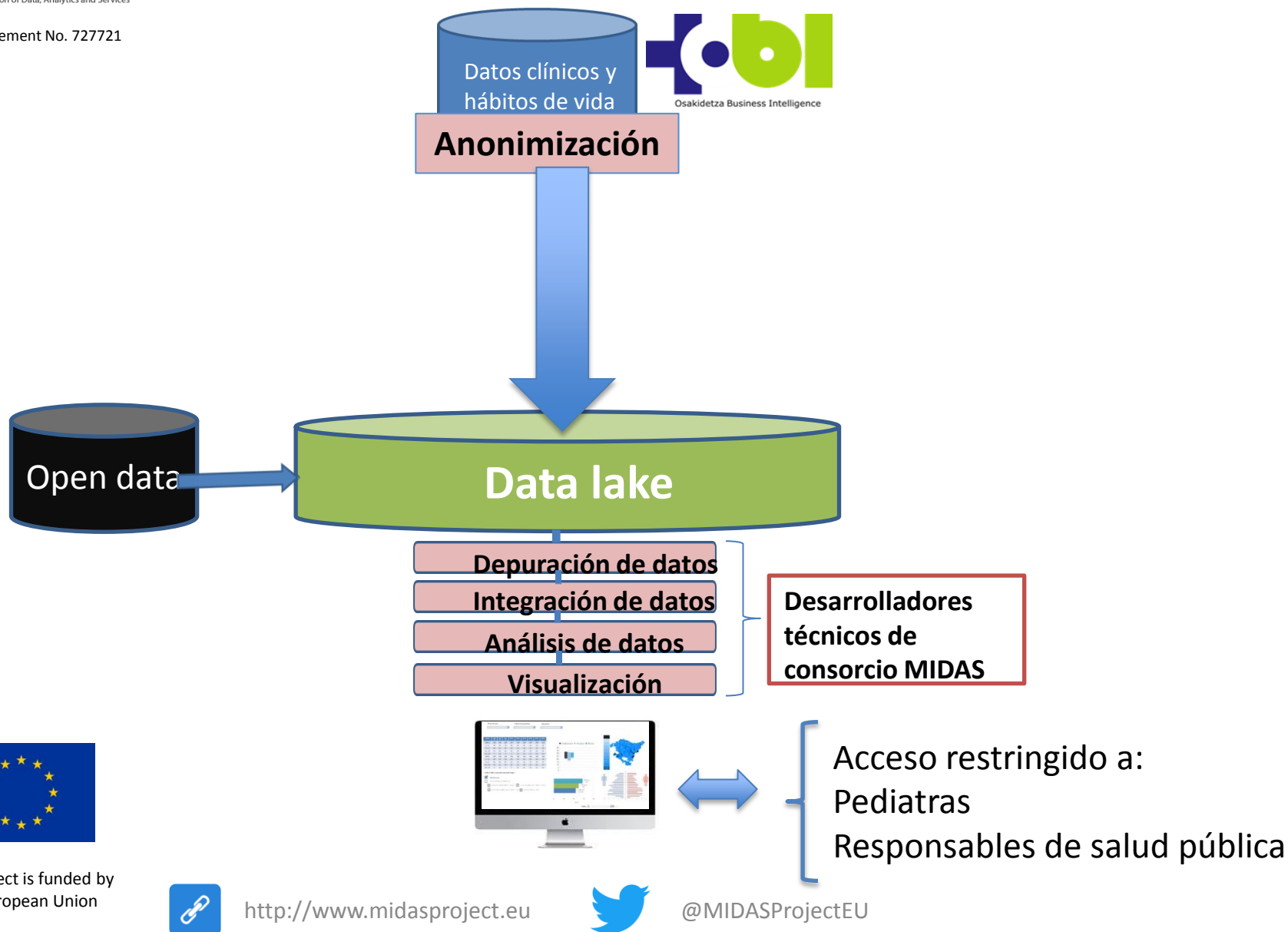
86% de registros en riesgo

Análisis de riesgo dataset final

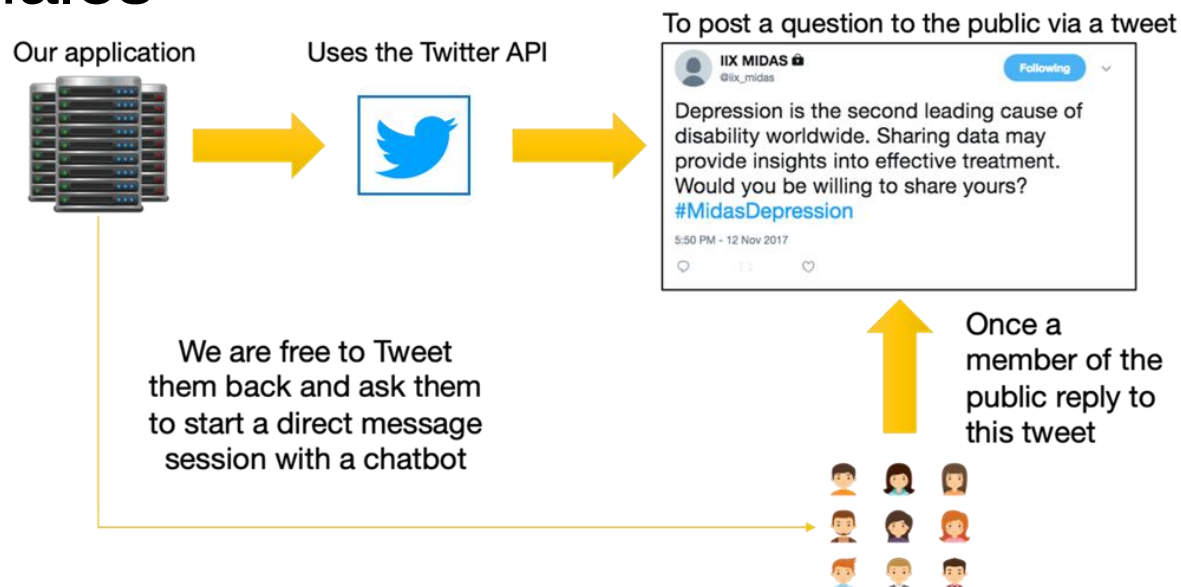


0% de registros en riesgo

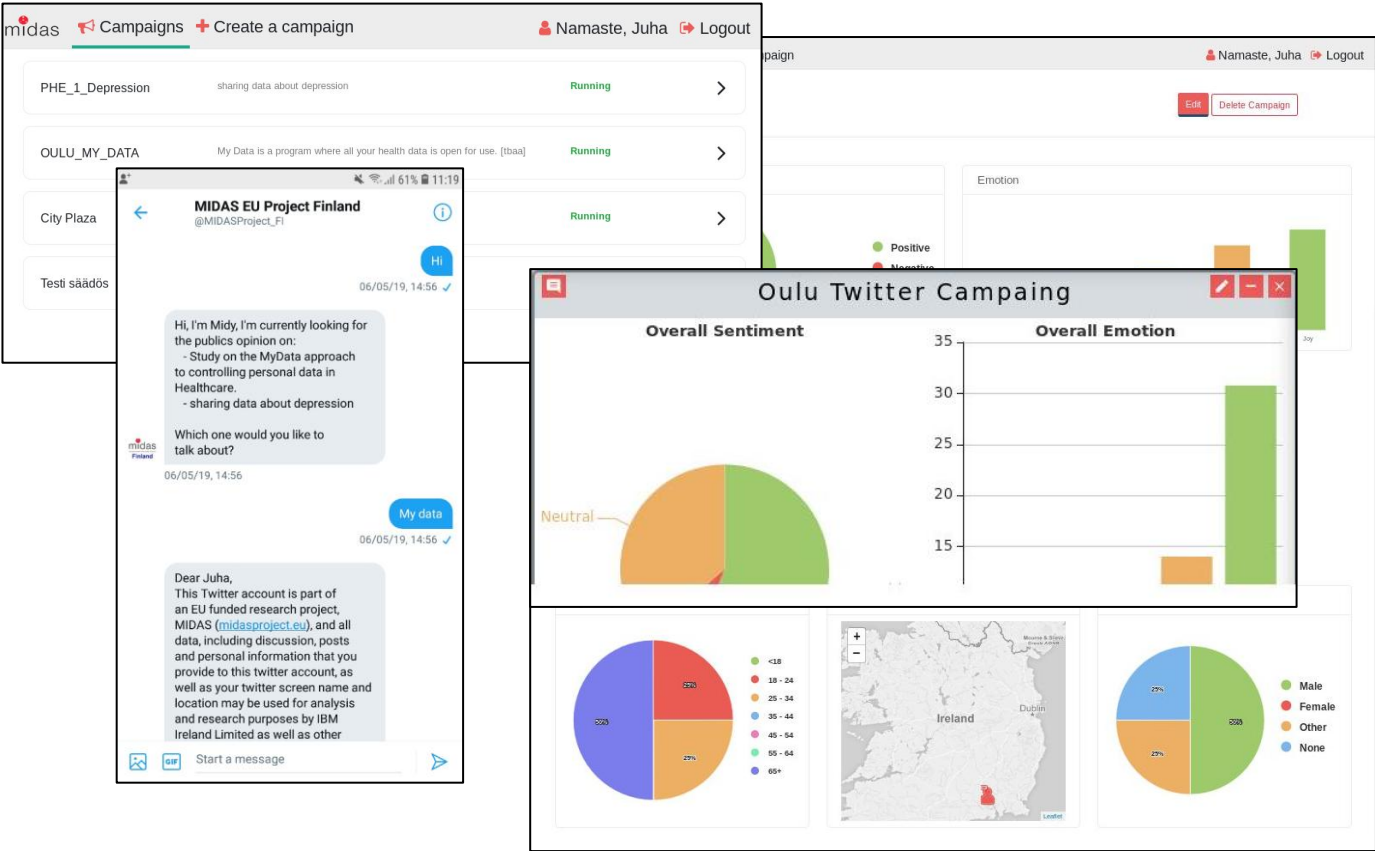




- Conseguir la percepción de la sociedad ante una pregunta mediante las redes sociales



Social Media Engagement Platform



This project is funded by the European Union



<http://www.midasproject.eu>

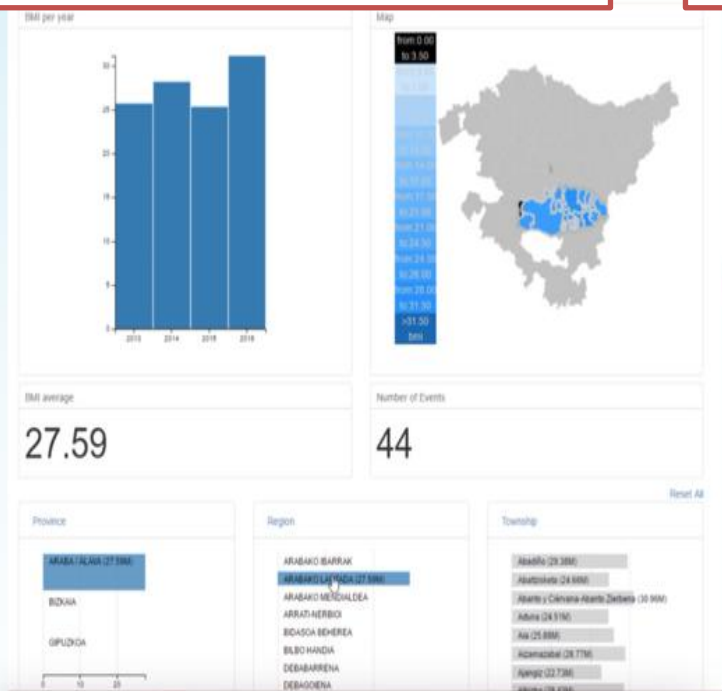


@MIDASProjectEU

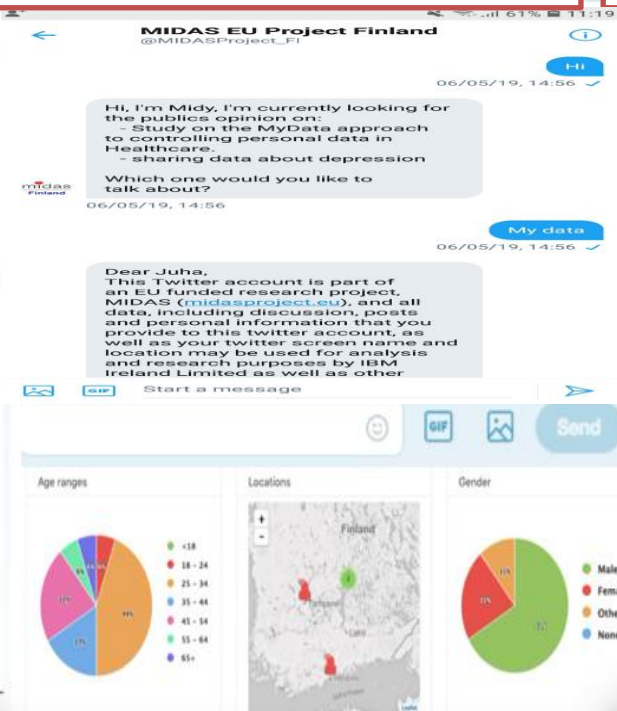
Panel actual de MIDAS: piloto Euskadi

Grant Agreement No. 727721

Bases de datos de salud



Chatbot twitter



Medline y Mesh headlines



This project is funded by the European Union



<http://www.midasproject.eu>



@MIDASProjectEU

Investigadores del grupo MIDAS

Grant Agreement No. 727721

- OSAKIDETZA
 - Dirección asistencia sanitaria
 - Carlos Sola
 - Mikel Ogueta
 - Servicios informáticos-servicios horizontales
 - Nicolás González
 - Grupo pediatría
 - M^a Pilar Aizpurua -OSI Donostialdea;
 - Francisco Javier Nuñez -OSI Bilbao-Basurto;
 - Jesus Arranz - OSI Araba;
 - Iratxe Salcedo OSI Arabako Errioxa - Rioja Alavesa;
 - Ainhoa Zabaleta OSI Donostialdea
- Departamento de Salud, Subdirección de Salud pública
 - Joseba Bidaurrázaga
 - Pilar Amiano
- BIOEF
 - Roberto Bilbao
 - Oihana Belar



This project is funded by
the European Union



<http://www.midasproject.eu>



@MIDASProjectEU

Eskerrik asko
Moltes gràcies
Muchas gracias

Roberto Bilbao
bilbao@bioef.org



This project is funded by
the European Union



<http://www.midasproject.eu>



@MIDASProjectEU