

Jornada Ciencia Abierta

Inicio Agenda

# Agenda

**DATOS**

10:00 - 10:30 Data Management Plan  
Alex Sánchez (UB/VHIR)  
Ángel del Rey (IRICYs)

Con la colaboración de:

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN Instituto de Salud Carlos III ORIÓN open science IdiPAZ Instituto de Investigación Hospital Universitario La Paz eatris+

Y con la financiación de:

ORION has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement (GA) No 741527 and since from May 2017 to April 2021.  
EATRIS Plus has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement (GA) No 871096

Gracias al programa de mentoring del proyecto Open Life Science.

Open Life Science

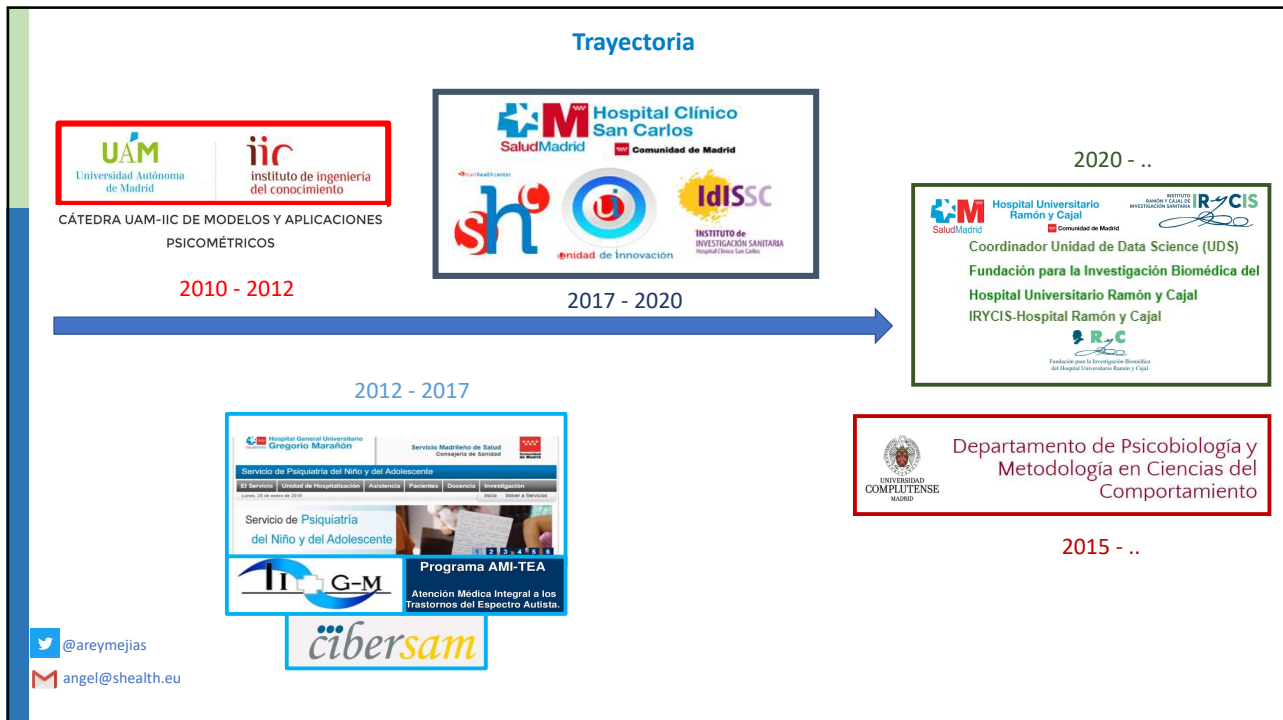
IRYCIS  
Hospital Universitario Ramón y Cajal  
Comunidad de Madrid

## Data Management Plan Caso de uso IRYCIS Ángel Luis Del Rey-Mejías

Coordinador Unidad de Data Science (UDS)  
Fundación para la Investigación Biomédica del  
Hospital Universitario Ramón y Cajal

@areymejas  
angel@shealth.eu

29/06/2021



**CONTENIDO**

- 1) Data Management Plan
- 2) Casos de mejora
  - 1) Biobanco
  - 2) Anonimización
  - 3) Tratamiento de Imágenes (ARX)
- 3) El Lago de Datos
- 4) Casos de éxito

@areymejas  
angel@shealth.eu

## CONTENIDO

- 1) **Data Management Plan**
- 2) Casos de mejora
  - 1) Biobanco
  - 2) Anonimización
  - 3) Tratamiento de Imágenes (ARX)
- 3) El Lago de Datos
- 4) Casos de éxito

@areymejas

angel@shealth.eu

### PLAN DE GESTIÓN DE DATOS DEL IRYCIS v1.0 18/06/2021

Los datos que se utilizarán para el desarrollo de este proyecto serán tratados como se especifica a continuación: i) De acuerdo con el **Reglamento 2016/679 de Protección de Datos y con la Ley Orgánica 3/2018** de Protección de datos de carácter personal; ii) De acuerdo con la Política de Privacidad y Seguridad de la Información del IRYCIS; iii) Sólo tras obtener el dictamen favorable de un comité de ética de la investigación; iv) Siguiendo los principios éticos relacionados con la investigación y el tratamiento de datos personales.

Dado que los participantes en el proyecto son pacientes de la Consejería de Sanidad de la Comunidad de Madrid, ésta ostenta de forma permanente la titularidad de los datos que serán utilizados en el (IP) el/la responsable de su custodia para garantizar su correcto tratamiento durante el desarrollo de las actividades previstas.

Todos los datos recogidos en el proyecto se registrarán y almacenarán en **REDCap (Research Electronic Data Capture, Vanderbilt University, TN, USA)**; es una aplicación web segura para gestionar bases de datos de investigación. Facilita el trabajo colaborativo y es óptima para la investigación clínica, ya que sus formularios permiten seguimiento longitudinal. REDCap incluye registros de auditoría para el seguimiento de la manipulación de datos y la actividad de los usuarios, así como herramientas de control de acceso de los usuarios, y cumple con los requisitos normativos nacionales e internacionales, incluido el Reglamento GDPR (UE) 2016/679. Para cada tipo de datos se establecerá un formulario específico de recogida de datos en colaboración con el Coordinador de la Unidad de Data Science del IRYCIS. Los datos generados se cargarán para dar acceso al IP y solo serán compartidos por éste/a con los miembros del equipo de investigación. **La seguridad del acceso a los datos será garantizada y supervisada por el Delegado de Protección de Datos (DPD)**. Todas las actividades realizadas en la plataforma estarán vinculadas al personal autorizado que accede a la plataforma, la autenticación de usuarios con datos de seguridad basados en roles permitirá al IP gestionar los diferentes privilegios entre los miembros del equipo. Los datos de pacientes serán accesibles sólo en forma codificada y anónima. El propietario de los datos tendrá la responsabilidad principal de garantizar que el análisis en el proyecto no infrinja los términos del consentimiento informado en virtud del cual se recopilarán los datos y deberá garantizar que la transferencia de datos se realice utilizando una **anonimización adecuada**. La seguridad de los datos procesados y registrados será gestionada por el IRYCIS. Esto garantizará que los datos de las bases de datos disponibles sean correctos, se mantengan constantemente actualizados y sean portables, se almacenen en un formato acordado y seguro con copias de seguridad adecuadas, que sólo se procesen los datos pertinentes conforme a la normativa y que se evite el procesamiento de datos excesivo e ilegal.

Una vez alcanzados los objetivos del proyecto, los datos utilizados serán depositados en el repositorio interno REDCap del IRYCIS, que es el Repositorio Institucional del Instituto Ramón y Cajal de Investigación Sanitaria, así como subidos a los repositorios más adecuados según el tipo de datos (por ejemplo, repositorios específicos para datos genómicos). El acceso a datos del repositorio interno para otros investigadores que puedan estar legítimamente interesados, estará sujeto a la **aprobación del Comité de Ética**. De este modo, se cumplen los requisitos de acceso abierto a los datos que deben cumplir los proyectos que cuentan con financiación pública, al tiempo que se mantienen en un entorno restringido de acuerdo con su carácter de datos personales.

@areymejas

angel@shealth.eu

Los datos pueden ser recogidos directamente de las personas que aceptan su participación en el estudio o a través de pruebas médicas y/o de laboratorio. Asimismo, en algunas circunstancias, los datos que ya existían en sus historias clínicas podrían ser recogidos con el consentimiento proporcionado por cada participante. Para la recogida de datos personales se deberán ser requerirán diferentes categorías de datos personales que deberán ser la salud, fisiológicos, sociológicos, hábitos de vida y personales, sexual u orientación sexual, características personales (antropométricas, físicas, etc.).

## Creación de un DMP

Toda la información indicada anteriormente se incluirá con mayor detalle en el **Plan de Gestión de Datos que se elaborará para el proyecto antes de la recogida de los datos personales, para hacer sus datos localizables, accesibles, interoperables y reutilizables (FAIR)**. Este documento se elaborará con el apoyo y orientación de la Unidad de Data Science del IRYCIS y, en su caso, del DPD de la Comunidad de Madrid.

### Publicaciones científicas de acceso abierto

Los resultados del proyecto se compartirán a través de revistas científicas de alto impacto revisadas por pares favoreciendo la publicación en medios de acceso abierto. Si el impacto de la revista científica es similar, se dará preferencia al acceso abierto 'gold' (acceso directo y gratuito en línea por parte del editor). El IP asignará un presupuesto para cubrir las tasas de publicación en acceso abierto, si es necesario. Las publicaciones del proyecto también se colgarán en el repositorio de la Comunidad de Madrid.

<p style="text-align: center;"><b>PLAN DE GESTIÓN DE DATOS DEL IRYCIS V1.0 18/02/2021</b></p> <p>Los datos que se utilizarán para el desarrollo de este proyecto serán tratados como se especifica a continuación: (i) De acuerdo con el Reglamento 2016/679 de Protección de Datos y con la Ley Orgánica 3/2018 de Protección de datos de carácter personal; (j) De acuerdo con la Política de Privacidad y Seguridad de la Información del IRYCIS; (k) Sólo tras obtener el dictamen favorable de un comité de ética de la investigación; (l) Siguiendo los principios éticos relacionados con la investigación y el tratamiento de datos personales.</p> <p>Dado que los participantes en el proyecto son pacientes de la Consejería de Sanidad de la Comunidad de Madrid, ésta ostenta de forma permanente la titularidad de los datos que serán utilizados en este proyecto, siendo e/la investigador/a principal (IP) e/la responsable de su custodia para garantizar su correcto tratamiento, así como el cumplimiento ético y normativo durante el desarrollo de las actividades previstas.</p> <p>Todos los datos recogidos en el proyecto se registrarán y almacenarán en la herramienta de captura electrónica de datos REDCap (Research Electronic Data Capture, Vanderbilt University, TN, USA); es una aplicación web segura para gestionar bases de datos de investigación. Facilita el trabajo colaborativo y es óptima para la investigación clínica, ya que sus formularios permiten de datos y la actividad normativa nacional un formulario específico. Los datos generados en la investigación, la seguridad del acceso a los datos será garantizada y supervisada por el Delegado de Protección de Datos (DPD). Todas las actividades realizadas en la plataforma estarán vinculadas al personal autorizado que acceda a la plataforma, la autenticación de usuarios con datos de seguridad basados en roles permitirá al IP gestionar los diferentes privilegios entre los miembros del equipo. Los datos de pacientes serán accesibles sólo en forma codificada y anónima. El propietario de los datos tendrá la responsabilidad principal de garantizar que el análisis en el proyecto no infrinja los términos del consentimiento informado en virtud del cual se recopilarán los datos y deberá garantizar que la transferencia de datos se realice utilizando una anonimización adecuada. La seguridad de los datos procesados y registrados será garantizada por el IRYCIS. Esto garantizará que los datos de las bases de datos disponibles sean correctos, se mantengan constantemente actualizados y sean portables, se almacenen en un formato acordado y seguro con copias de seguridad adecuadas, que sólo se procesen los datos pertinentes conforme a la normativa y que se evite el procesamiento de datos excesivo e ilegal.</p> <p>Una vez alcanzados los objetivos del proyecto, los datos utilizados serán depositados en el repositorio interno REDCap del IRYCIS, que es el Repositorio Institucional del Instituto Ramón y Cajal de Investigación Sanitaria, así como subidos a los repositorios más adecuados según el tipo de datos (por ejemplo, repositorios específicos para datos genómicos). El acceso a datos del repositorio interno para otros investigadores que puedan estar legítimamente interesados, estará sujeto a la aprobación del Comité de Ética. De este modo, se cumplen los requisitos de acceso abierto a los datos que deben cumplir los proyectos que cuentan con financiación pública, al tiempo que se mantienen en un entorno restringido de acuerdo con su carácter de datos personales.</p> <p>Los datos pueden ser recogidos directamente de las personas que aceptan su participación en el estudio o a través de pruebas médicas y/o de laboratorio. Asimismo, en algunas circunstancias, los datos que ya existían en sus historias clínicas podrían ser utilizados y accedidos por el personal sanitario con el consentimiento proporcionado por cada participante. Para la ejecución de las actividades del proyecto, se requerirán diferentes categorías de datos personales que deberán ser enumerados: identificativos, relacionados con la salud, fisiológicos, sociológicos, hábitos de vida y personales, comportamiento, demográficos, genéticos, vida sexual u orientación sexual, características personales (antropométricas, profesionales, geográficas), etc. Toda la información indicada anteriormente se incluirá con mayor detalle en el Plan de Gestión de Datos que se elaborará para el proyecto antes de la recogida de los datos personales, para hacer sus datos localizables, accesibles, interoperables y reutilizables (FAIR). Este documento se elaborará con el apoyo y orientación de la Unidad de Data Science del IRYCIS y, en su caso, del DPD de la Comunidad de Madrid.</p> <p>Publicaciones científicas de acceso abierto</p> <p>Los resultados del proyecto se compartirán a través de revistas científicas de alto impacto revisadas por pares favoreciendo la publicación en medios de acceso abierto. Si el impacto de la revista científica es similar, se dará preferencia al acceso abierto (gold) (acceso directo y gratuito en línea por parte del editor). El IP asignará un presupuesto para cubrir las tasas de publicación en acceso abierto, si es necesario. Las publicaciones del proyecto también se cargarán en el repositorio de la Comunidad de Madrid.</p>	<p style="text-align: center;"><b>IRYCIS DATA MANAGEMENT OF RESEARCH DATA PLAN v1.0 18/06/2021</b></p> <p>The data that will be used for the development of this project will be treated as specified below:</p> <ul style="list-style-type: none"> <li>- In accordance with the Regulation 2016/679 on Data Protection and with Organic Law 3/2018 on the Protection of personal data.</li> <li>- In accordance with IRYCIS's Privacy and Information Security Policy.</li> <li>- Only after obtaining a favorable opinion from a research ethics committee.</li> <li>- Following the ethical principles related to research and processing of personal data.</li> </ul> <p>Since the project participants are patients of the "Comunidad de Madrid" of the region of Madrid, it permanently holds the ownership of the data that will be used in this project, and its principal investigator (PI) is responsible for their custody to ensure its proper treatment, as well as ethical and regulatory compliance during the development of planned activities.</p> <p>All data collected in the project will be recorded and stored in an electronic data capture tool developed for this project under REDCap (Research Electronic Data Capture, Vanderbilt University, TN, USA). REDCap is a secure web application for managing online research databases. It facilitates collaborative work and is an optimal tool for clinical research, as its forms allow easy set up. REDCap include audit trails for tracking data manipulation and user activity, as well as tools for data export, and complies with national and international regulatory and legal requirements, including GDPR 2016/679.</p> <p>Once a specific data collection form will be established in collaboration with the REDCap manager (Coordinator of the Data Science Unit) at IRYCIS. The data generated will be updated to provide access to the PI. Data will be shared by the PI only with the research team members. Access data security would be granted and supervised by the Data Protection Officer (DPO). All activities realized in the platform will be linked to the authorized personal who access to the platform, the user authentication with a role-based security data allow the PI to manage the different privileges between team members. All patient level data will be accessed only in coded and anonymized form. The data owner will have the primary responsibility of ensuring that analysis in the context of the project will not infringe the terms of the informed consent under which the data will be collected and must ensure that data transfer will be conducted using appropriate anonymization. The safety of data processed and recorded will be managed by IRYCIS. This will ensure that the data in the available databases will be correct, constantly kept up-to-date and portable, stored in an agreed and secure format with adequate backups, that only relevant data will be processed adherent to the ethics standard, avoiding excessive and unlawful data processing.</p> <p>Once the objectives of the project have been achieved, the data used will be deposited in the internal REDCap repository of IRYCIS, which is the Institutional Repository of the Ramón y Cajal Health Research Institute, as well as uploaded on most appropriate repositories according to the type of data (e.g., specific repositories for genomic data). Access to the internal repository data for other researchers who may be legitimately interested in them will be subject to the approval of an ethics commission. In this way, the Open Access data requirements are fulfilled, which must be met by projects that have public funding, while they are kept in a restricted environment in accordance with their character of personal data.</p> <p>Data can be collected directly from people who accept their participation in the study directly or through medical and / or laboratory tests. Likewise, under some circumstances, data that already existed in their medical records could be used and accessed by healthcare personnel with the consent provided by each participant.</p> <p>For the execution of project activities, different categories of personal data will be required and should be listed: Identifiers, related to health, physiological, sociological, life and personal habits, behavior, demographics, genetics, sex life or sexual orientation, personal characteristics (anthropometric, professional, geographic), etc.</p> <p>All the information indicated above will be included in greater detail in the Data Management Plan that will be prepared for the project prior to the personal data being collected, for making their data findable, accessible, interoperable, and reusable (FAIR). This document will be prepared with the support and guidance of the IRYCIS Data Science Unit and, if necessary, of the Data Protection Delegation of the region of Madrid.</p> <p><u>Open access scientific publications</u></p> <p>The project results will be shared via peer-reviewed, high impact scientific journals, adopting the 'gold' or 'green' model, favoring publishing in open access media. If the impact of the scientific journal is similar, preference will be given to 'gold' open access (direct free on line access by the publisher). The PI will allocate budget to cover open-access publication fees, if necessary. The publications of the project will also be uploaded on the repository of the region of Madrid.</p>
<p><b>Versión inglés</b></p>	

@areymejas

angel@shealth.eu

## CONTENIDO

- 1) Data Management Plan
- 2) Casos de mejora
  - 1) Biobanco
  - 2) Anonimización
  - 3) Tratamiento de Imágenes (ARX)
- 3) El Lago de Datos
- 4) Casos de éxito

@areymejas

angel@shealth.eu

## CONTENIDO

- 1) Data Management Plan
- 2) Casos de mejora
  - 1) **Biobanco**
  - 2) Anonimización
  - 3) Tratamiento de Imágenes (ARX)
- 3) El Lago de Datos

@areymejas

angel@shealth.eu

La gestión preanalítica de las muestras y los datos asociados se va a realizar en colaboración con el Biobanco. La preparación preanalítica de las muestras será realizada por el **Biobanco conforme a los procedimientos normalizados de trabajo y el Sistema de Gestión de la Calidad implantado (Certificado ISO 9001SGS nº ES13/14407).**

**La información difundida y obtenida por la puesta en marcha del presente proyecto es considerada confidencial y deberá ser tratada en todo momento como tal.** Los sujetos del estudio se identificarán con un código alfanumérico. Sólo aquellos datos de la historia clínica que estén relacionados con el estudio serán objeto de comprobación. Esta comprobación se realizará en presencia del Investigador Principal/Investigadores Colaboradores, responsables de garantizar la confidencialidad de todos los datos de las historias clínicas pertenecientes a los sujetos participantes en el estudio. Los investigadores firman a tal efecto el compromiso de confidencialidad **CEIm**.

Los datos serán por tanto **tratados con medidas de seguridad de nivel alto** y conforme a lo establecido en el **Reglamento General de Protección de Datos** (Reglamento [UE] 2016/679) así como la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales. La aplicación permite una gestión integral de los datos asociados.

Si el participante consiente los **excedentes de las muestras** y los datos asociados podrán ser conservados en el Biobanco en vez de destruirlos, de forma que una vez finalizado el proyecto inicial, cuyo investigador principal/colaboradores sí conocen la identidad, a cualquier nuevo investigador se le entregarán las muestras codificadas junto con los datos clínicos asociados, sabiendo sólo el personal del Biobanco, con acceso autorizado a la base de datos, a quien corresponde el código. Se garantiza por tanto el mantenimiento de la confidencialidad.

Cumpliendo con la **LEY 14/2007, de 3 de julio, de Investigación biomédica y el RD 1716/2011**, las muestras y los datos asociados se utilizarán únicamente en proyectos científica y éticamente aprobados, a nivel nacional e internacional, incluyendo la posibilidad de estudios genéticos. Además, el Biobanco está

@areymejas

angel@shealth.eu

**Tratamiento muestras, excedentes, genética, reglamentación extra, etc.**

**Describir la tipología y formato de los datos a recoger / generar en el marco del proyecto, procedimiento previsto para acceso a los mismos (quién, cómo y cuándo podrá acceder a ellos), titularidad de los datos, repositorio en que se prevé realizar su depósito, y procedimiento previsto para garantizar los requisitos éticos o legales específicos de aplicación.**

(Añadir la tipología y formato de los datos de que se trata...)

La gestión preanalítica de las muestras y los datos asociados se va a realizar en colaboración con el Biobanco. La preparación preanalítica de las muestras será realizada por el Biobanco conforme a los procedimientos normalizados de trabajo y el Sistema de Gestión de la Calidad implantado (Certificado ISO 9001SGS nº ES13/14407).

La información difundida y obtenida por la puesta en marcha del presente proyecto es considerada confidencial y deberá ser tratada en todo momento como tal. Los sujetos del estudio se identificarán con un código alfanumérico. Sólo aquellos datos de la historia clínica que estén relacionados con el estudio serán objeto de comprobación. Esta comprobación se realizará en presencia del Investigador Principal/Investigadores Colaboradores, responsables de garantizar la confidencialidad de todos los datos de las historias clínicas pertenecientes a los sujetos participantes en el estudio. Los investigadores firman a tal efecto el compromiso de confidencialidad que hay anejar al proyecto para su evaluación en el **CEIm**.

Los datos serán por tanto tratados con medidas de seguridad de nivel alto y conforme a lo establecido en el Reglamento General de Protección de Datos europeo (Reglamento [UE] 2016/679) así como la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales. La aplicación permite una gestión integral de los procesos, trazabilidad de muestras y datos asociados.

El acceso está controlado por clave y facilita el registro vía web en el cuestionario de los datos asociados requeridos para el proyecto. Solo el investigador principal/colaboradores podrán relacionar dichos datos con el paciente y con su historia clínica.

Todos los datos necesarios para el estudio serán tratados fuera de la aplicación informática de forma codificada, sin incluir la identidad del participante para preservar la confidencialidad.

Si el participante consiente los excedentes de las muestras y los datos asociados podrán ser conservados en el Biobanco en vez de destruirlos, de forma que una vez finalizado el proyecto inicial, cuyo investigador principal/colaboradores sí conocen la identidad, a cualquier nuevo investigador se le entregarán las muestras codificadas junto con los datos clínicos asociados, sabiendo sólo el personal del Biobanco, con acceso autorizado a la base de datos, a quien corresponde el código. Se garantiza por tanto el mantenimiento de la confidencialidad.

Cumpliendo con la **LEY 14/2007, de 3 de julio, de Investigación biomédica y el RD 1716/2011**, las muestras y los datos asociados se utilizarán únicamente en proyectos científica y éticamente aprobados, a nivel nacional e internacional, incluyendo la posibilidad de estudios genéticos. Además, el Biobanco está adscrito a un comité de ética y a un comité científico que evaluará cada una de las solicitudes de muestras que reciba, siendo necesaria su aprobación para cederlas. Este comité de ética también evaluará el acceso a la Historia Clínica por personal cualificado, en caso de ser necesario para la obtención de datos imprescindibles para la realización de los proyectos para los que se solicitan sus muestras.

Los participantes podrán ejercer el derecho de supresión, oposición, portabilidad, limitación del tratamiento, acceso y rectificación de sus datos, así como, el derecho de solicitar la destrucción o anonimización de las muestras y/o los datos que aún no hayan sido utilizados. Para ello deberá ponerse en contacto con nosotros en el teléfono 913369075 o en la dirección de correo electrónico: [biobanco@salud.madrid.org](mailto:biobanco@salud.madrid.org).

Así mismo tienen derecho a dirigirse a la Agencia de Protección de Datos si no quedaran satisfechos ([www.agpd.es](http://www.agpd.es)).

## CONTENIDO

- 1) Data Management Plan
- 2) Casos de mejora
  - 1) Biobanco
  - 2) **Anonimización**
  - 3) Tratamiento de Imágenes (ARX)
- 3) El Lago de Datos
- 4) Casos de éxito

@areymejas

angel@shealth.eu

El tema es la pseudoanonimización que nos pide la OSSI, para lo cual debemos encontrar un sistema sencillo para aplicar a una base de datos (CMDB) de unos 1,5 millones de registros, un HASH SHA512 con una clave HMAC

### Anonimización concreta

### ANÁLISIS PRELIMINAR DE VIABILIDAD DE NUEVOS APLICATIVOS

[Estudio observacional sobre el efecto de la pandemia por COVID-19 en la incidencia de eventos de patologías tiempo-dependientes atendidas en el SUMMA 112]

@areymejas

angel@shealth.eu

```

## Proceso
1. Cargamos los paquetes.
> library(readr)

# Prueba 1: Importar csv de 150mil registros
wgm_2018_publiccsv <- read_csv("Pruebas/wgm_2018_publiccsv.csv")
# Creamos la diferencia de tiempos

DNIS = as.character(as.vector(wgm_2018_publiccsv$PROJWTT))
for (i in 1:length(DNIS)){
  a[i] = sha256(DNIS[i], key = password)
}
cbind(a,wgm_2018_publiccsv)

### [? setup, include=FALSE]
knitr::opts_chunk$set(echo = TRUE)
library("digest")
library("openssl")
...

Prueba 1: Importar csv 150.000 registros de prueba
--- [? importar csv 150mil]
library(readr)
wgm_2018_publiccsv <- read_csv("Pruebas/wgm_2018_publiccsv.csv")
view(wgm_2018_publiccsv)

Prueba 2: Hashear y contar tiempo
--- [? prueba csv de 150mil registros]
start=Sys.time()
password="cba"
a=0
DNIS = as.character(as.vector(wgm_2018_publiccsv$PROJWTT))
for (i in 1:length(DNIS)){
  a[i]=sha256(DNIS[i], key = password)
}
HASH_Pruebas_csv = cbind(a,wgm_2018_publiccsv)
stop=Sys.time()
stop-start
HASH_Pruebas_csv$PROJWTT<-NULL
write_csv(HASH_Pruebas_csv,"pruebas_Hasheado.csv")
  
```

# CONTENIDO

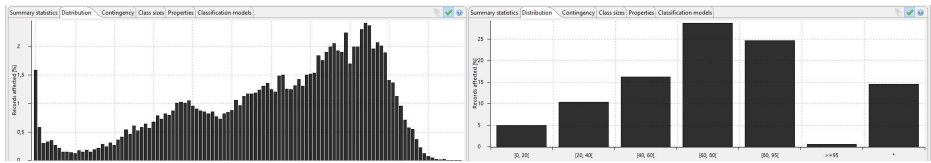
- 1) Data Management Plan
- 2) Casos de mejora
  - 1) Biobanco
  - 2) Anonimización
  - 3) Tratamiento de Imágenes (ARX)**
- 3) El Lago de Datos
- 4) Casos de éxito

@areymejas  
angel@sheath.eu



## Comprobación a posteriori de proceso anonimización

## 2. Explorar resultados




Quasi-identifier	Distinction	Separation	Quasi-identifier	Distinction	Separation
SEXO	0.00638%	99.85037%	NHC	0.00314%	0%
Edad	0.34231%	98.81185%	Edad	0.00833%	14.07921%
NHC	75.71958%	99.99754%	SEXO	0.00638%	49.85037%
SEXO, Edad	0.66302%	99.28893%	NHC, Edad	0.00833%	14.07921%
NHC, SEXO	75.71958%	99.99754%	NHC, SEXO	0.00638%	49.85037%
NHC, Edad	99.85923%	99.99923%	SEXO, Edad	0.01237%	56.9608%
NHC, SEXO, Edad	99.85923%	99.99923%	NHC, SEXO, Edad	0.01237%	56.9608%



@areymejas  
angel@sheath.eu

## 4. Análisis de riesgos

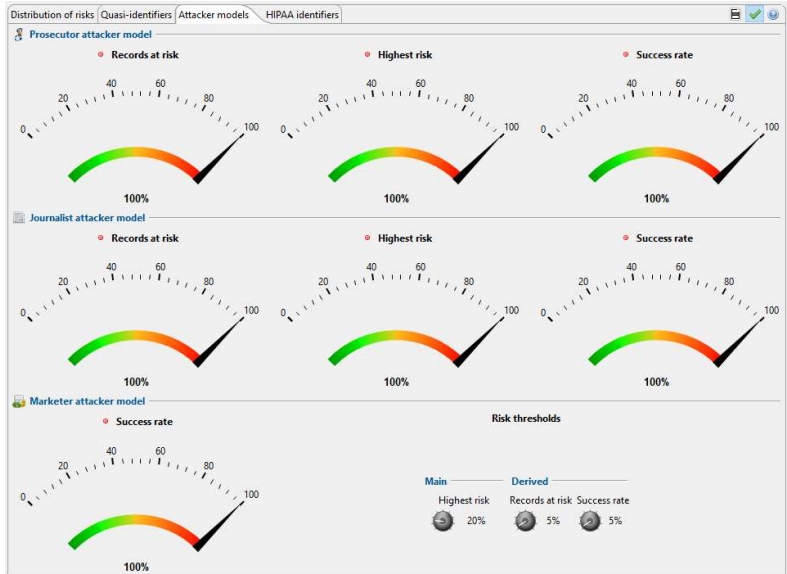


**Análisis de riesgos**  
**3. Modelos de ataque**


El programa calcula los riesgos que tiene la base de datos de ser hackeada desde 3 perfiles profesionales de hacker distintos

- Prosecutor
- Journalist
- Marketer


Pudiendo personalizarse los umbrales de alto riesgo.




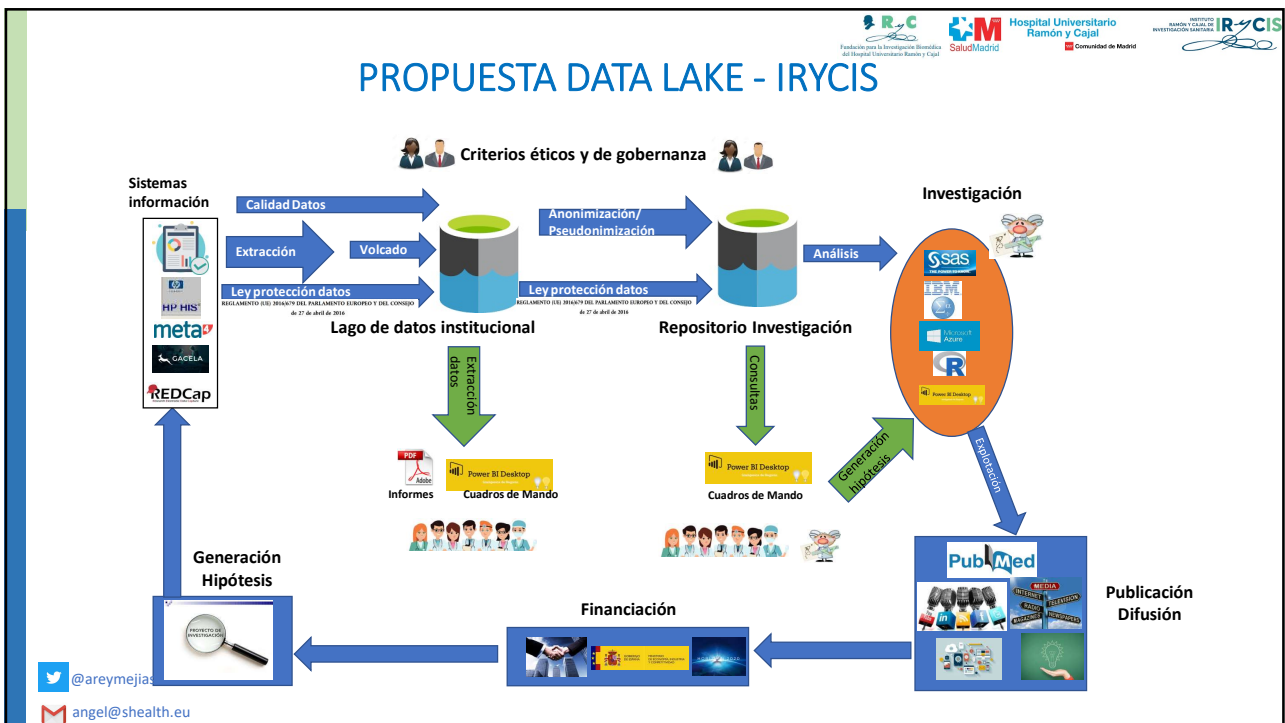
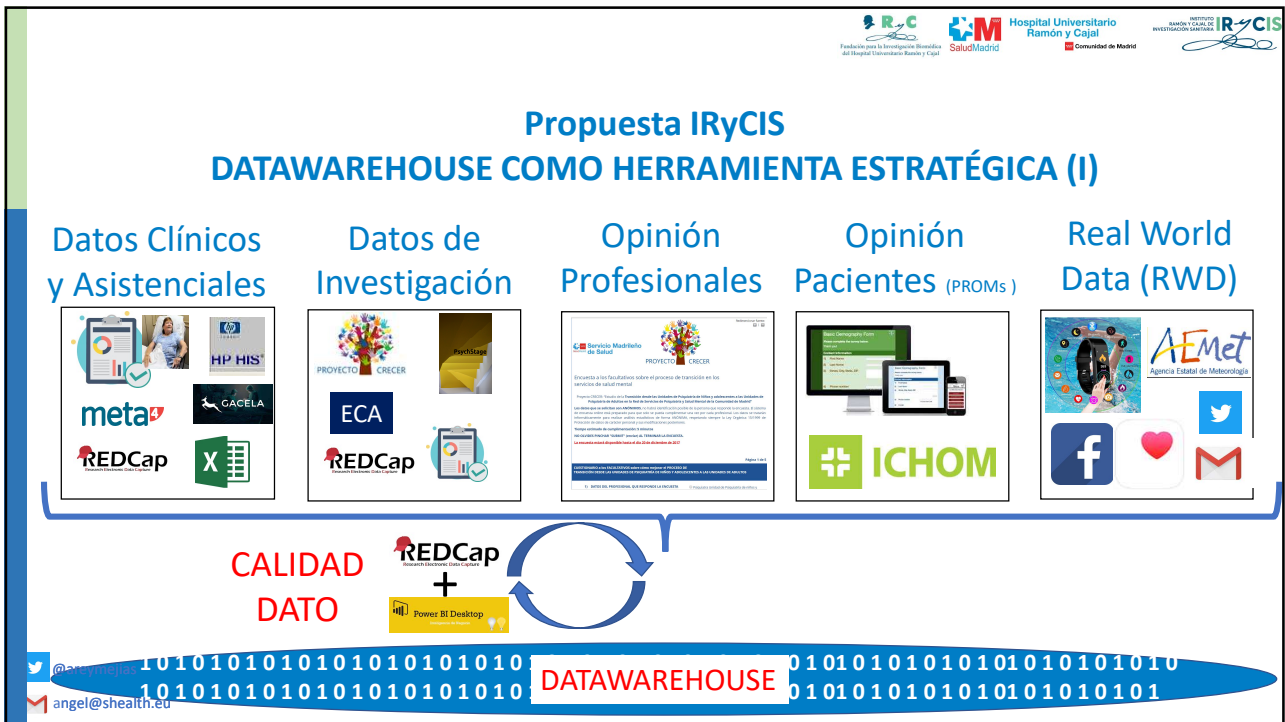
## CONTENIDO

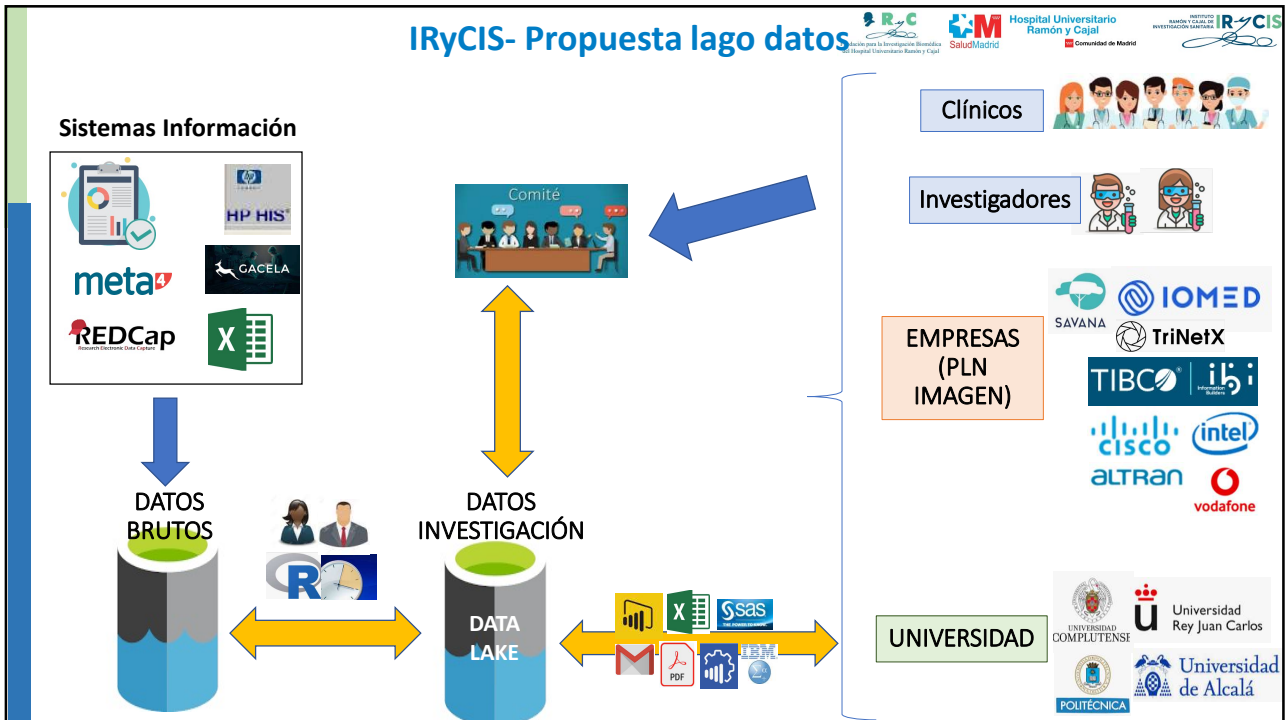
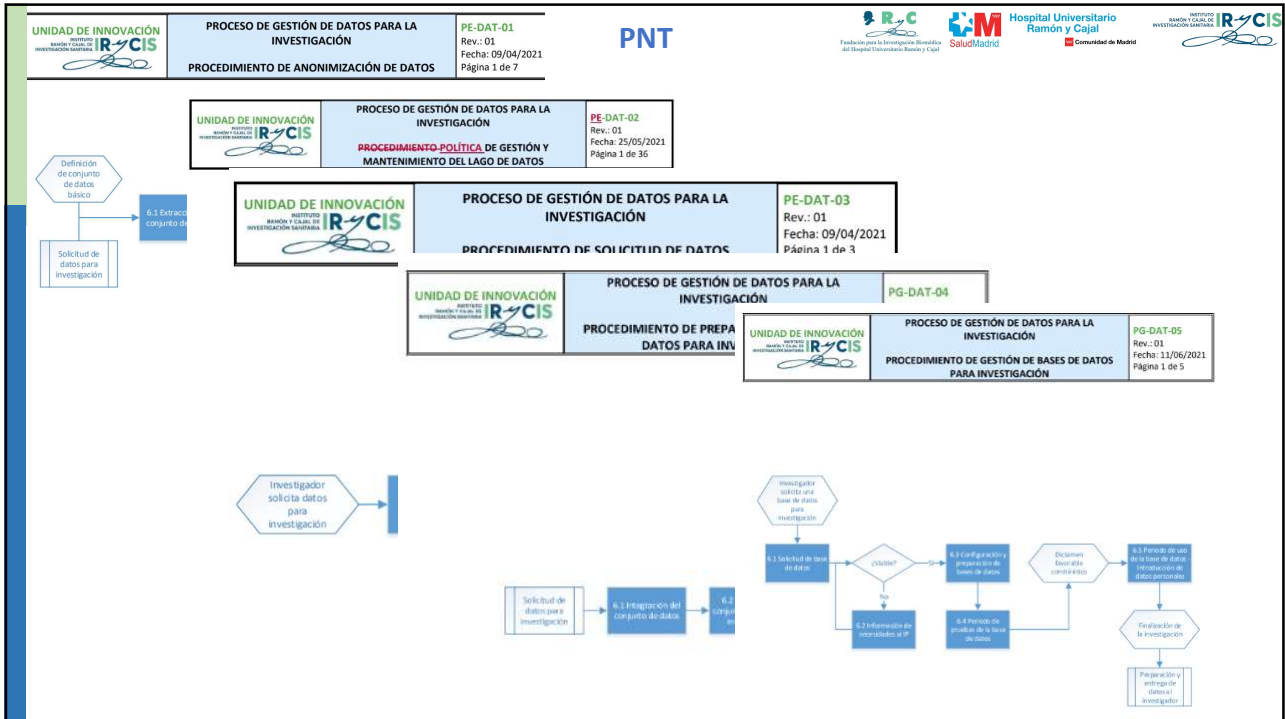


- 1) Data Management Plan
- 2) Casos de mejora
  - 1) Biobanco
  - 2) Anonimización
  - 3) Tratamiento de Imágenes (ARX)
- 3) El Lago de Datos**
- 4) Casos de éxito

 @areymejas

 angel@shealth.eu









## Caso éxito





### Sharing ICU Patient Data Responsibly Under the Society of Critical Care Medicine/European Society of Intensive Care Medicine Joint Data Science Collaboration: The Amsterdam University Medical Centers Database (AmsterdamUMCdb) Example

**OBJECTIVES:** Critical care medicine is a natural environment for machine learning approaches to improve outcomes for critically ill patients as admissions to ICUs generate vast amounts of data. However, technical, legal, ethical, and privacy concerns have so far limited the critical care medicine community from making these data readily available. The Society of Critical Care Medicine and the European Society of Intensive Care Medicine have identified ICU patient data sharing as one of the priorities under their Joint Data Science Collaboration. To encourage ICUs worldwide to share their patient data responsibly, we now describe the development and release of Amsterdam University Medical Centers Database (AmsterdamUMCdb), **the first freely available critical care database in full compliance with privacy laws from both the United States and Europe**, as an example of the feasibility of sharing complex critical care data.

Patrick J. Thoral, MD<sup>1</sup>  
 Jan M. Peppink<sup>1</sup>  
 Ronald H. Driessen<sup>1</sup>  
 Eric J. G. Silbrands<sup>1</sup>, MD, PhD<sup>2</sup>  
 Erwin J. O. Kompanje, PhD<sup>3</sup>  
 Lewis Kaplan, MD, FACS, FCCM<sup>1,4</sup>  
 Heatherlee Bailey, MD, FCCM<sup>1,5</sup>  
 Jozef Kesecioglu, MD, PhD<sup>1,6</sup>  
 Maurizio Cecconi, MD, PhD<sup>1,7</sup>  
 Matthew Churpek, MD, MPH, PhD<sup>8</sup>  
 Gilles Clermont, MD<sup>11</sup>  
 Mihaela van der Schaer, PhD<sup>12,13</sup>

**MEASUREMENTS AND MAIN RESULTS:**

AmsterdamUMCdb contains approximately **1 billion clinical data points from 23,106 admissions of 20,109 patients**. The privacy audit concluded that reidentification is not reasonably likely, and AmsterdamUMCdb can therefore be considered as anonymous information, both in the context of the U.S. Health Insurance Portability and Accountability Act and the European General Data Protection Regulation. The ethics audit concluded that responsible data sharing imposes minimal burden, whereas the potential benefit is tremendous.

**CONCLUSIONS:**

Technical, legal, ethical, and privacy challenges related to responsible data sharing can be addressed using a multidisciplinary approach. A risk-based deidentification strategy, that complies with both U.S. and European privacy regulations, should be the preferred approach to releasing ICU patient data. This supports the shared Society of Critical Care Medicine and European Society of Intensive Care Medicine vision to improve critical care outcomes through scientific inquiry of vast and combined ICU datasets.

	Hypothetical Adversary		
	Friendly Researcher	Rogue Researcher	Rogue Insurance Company
<b>Assessment of reidentification risk</b>			
P(access)	1.00	1.00	0.27
P(intention)	0.20	0.10	0.10
<b>Average risk</b>			
P(reidentification)	0.047	0.047	0.009
k-anonymity	89	89	682
l-diversity	26	26	65
<b>Maximum risk</b>			
P(reidentification)	0.50	0.50	0.50
k-anonymity	2	2	2
l-diversity	2	2	2
P final risk	0.01	0.05	0.0002



## Caso éxito





Comunidad de Madrid 
Inicio
Proyectos CPI
Eventos
Documentación
Actualidad
Contacto

# Infobanco

El proyecto **Arquitectura de datos de salud para el aprendizaje, INFOBANCO**, tiene la finalidad de desarrollar una **arquitectura de red regional de datos para el aprendizaje del sistema sanitario, concebido como un repositorio normalizado de datos de salud**, que combine información generada por diferentes fuentes, tanto sistemas clínicos como de investigación y administrativos, y con gran utilidad para la mejora asistencial y la innovación, la atención sanitaria basada en el valor (VBHC), la investigación biomédica y otros usos secundarios.

Esta arquitectura pretende tener un funcionamiento de plataforma que provea servicios a clínicos, gestores e investigadores, haciendo posible la combinación de datos de múltiples fuentes, y deberá estar dotada de herramientas de gobernanza, obtención, transformación, interrogación, visualización y análisis de datos para la obtención de conocimiento y el soporte a la toma de decisiones. Se evaluará la calidad de los datos y la validez y utilidad de esta arquitectura de INFOBANCO.

INFOBANCO permitirá la combinación y explotación conjunta de datos de salud entre distintas fuentes de los centros de la Consejería de Sanidad de Madrid y/o el Servicio Madrileño de Salud (SERMAS), para el uso secundario, procedentes de los diferentes niveles asistenciales (atención primaria, hospitales, emergencias, farmacia), con una perspectiva tanto individual, en el ámbito de la medicina personalizada y de precisión, como poblacional y de salud pública.


Jornada Ciencia Abierta



**Hospital Universitario  
Ramón y Cajal**  
Comunidad de Madrid



**IRYCIS**  
INSTITUTO RAMÓN Y CAJAL DE INVESTIGACIÓN SANITARIA

## Data Management Plan

### Ángel Luis Del Rey-Mejías








# ¡Gracias!

 @areymejias  
 alreymejias@gmail.com

Con la colaboración de:









Y con la financiación de:








Gracias al programa de mentoring del proyecto Open Life Science.





Fundación para la Investigación Biomédica del Hospital Universitario Ramón y Cajal



**Hospital Universitario  
Ramón y Cajal**  
Comunidad de Madrid



**IRYCIS**  
INSTITUTO RAMÓN Y CAJAL DE INVESTIGACIÓN SANITARIA



“If I had an hour to solve a problem I'd spend 55 minutes thinking about the problem and 5 minutes thinking about solutions.”

*Albert Einstein*



“It is not the strongest of the species that survives, nor the most intelligent, but the one most responsive to change.”

-Charles Darwin, 1809

## ¿Preguntas?

Ángel Luis Del Rey Mejías

 @areymejias  
 alreymejias@gmail.com