



FAIR4Health: la gestión de datos de investigación en salud

Tony Hernández-Pérez (UC3M)

BCSALUT 2019

III JORNADES DE BIBLIOTEQUES I SALUT DE CATALUNYA

Barcelona, 14 | 15 de novembre 2019



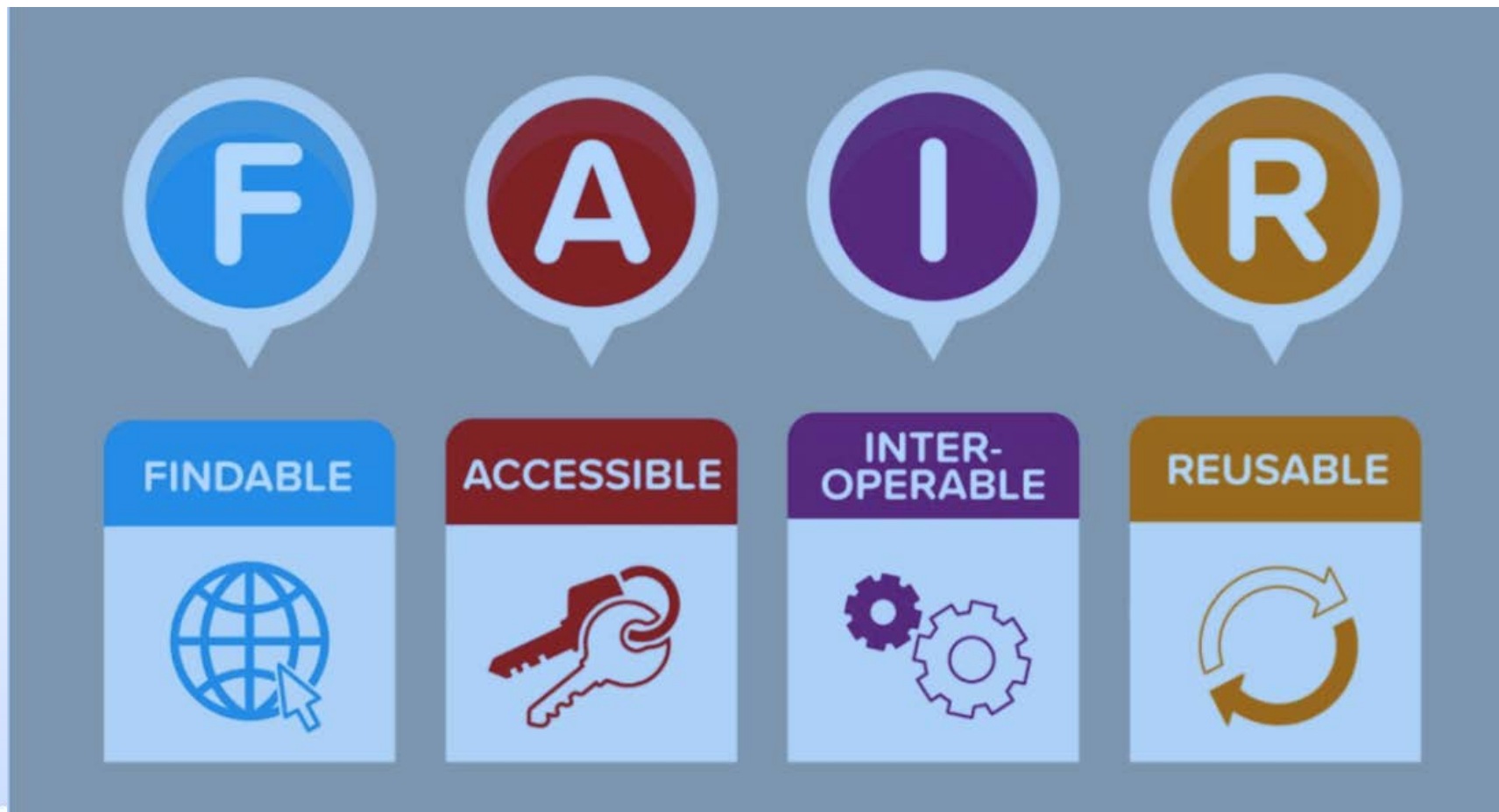
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666



European
Commission



¿Qué son datos FAIR?



Force11 (2014) – Wilkinson et al (2016)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666



European
Commission



Principios FAIR

(F) LOCALIZABLE

F1. (meta)datos tienen asignado un identificador global único y persistente.

F2. Los datos están descritos con metadatos.

F3. (meta)datos están registrados o indexados como recursos localizables.

F4. Los metadatos especifican el identificador del dato.

(A) ACCESIBLE

A1. (meta)datos se pueden recuperar con el identificador usando un protocolo de comunicaciones estándar.

A1.1. El protocolo es abierto, gratuito e implementable de forma universal.

A1.2. El protocolo permite procesos de autenticación y autorización.

A2. Los metadatos son accesibles incluso cuando los datos ya no están disponibles

(I) INTEROPERABLE

I1. (meta)datos usan un lenguaje formal, accesible, compartido y de aplicación amplia para la representación del conocimiento.

I2. (meta)datos usan vocabularios que siguen los principios FAIR.

I3. (meta)datos incluyen referencias cualificadas a otros (meta)datos

(R) REUTILIZABLE

R1. (meta)datos tienen una pluralidad de atributos precisos y relevantes.

R1.1. (meta)datos se publican con una licencia de uso clara y accesible.

R1.2. (meta)datos están asociados con su procedencia.

R1.3. (meta)datos cumplen con estándares comunes relevantes en su dominio

FORCE11 - FAIR Data Guiding Principles

<https://www.force11.org/group/fairgroup/fairprinciples>

¿Por qué datos FAIR?



Digitalización –
Obtención datos



Ciencia
transdisciplinar

Sensores – Big Data - Reproducibilidad



Cost of not having FAIR research data

Cost-Benefit analysis for FAIR research data

1200 millones de euros/año
10% aprox del presupuesto H2020
2019 (11.000 mil)

Economía de
datos



GDPR - Directiva
UE 2019/1024

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666



European
Commission



No es fácil: barreras

Culturales

- Falta de conciencia (humanidades, ciencias sociales)
- Pérdida ventaja competitiva

Socio-económicas

- Costes (formación, gestión datos, almacenamiento, recursos técnicos...)
- Incertidumbre (ética, legal...)
- Falta de incentivo (no es publicación)

Técnicas

- Problemas de interoperabilidad técnica (hardw y softw)
- Problemas de interoperabilidad semántica (vocabularios, pid...)
- Problemas de interoperabilidad organizacional (permisos, consentimientos...)

El consorcio

Coordinado por el **Hospital Universitario Virgen del Rocío, del Sistema Andaluz de Salud (SAS)**

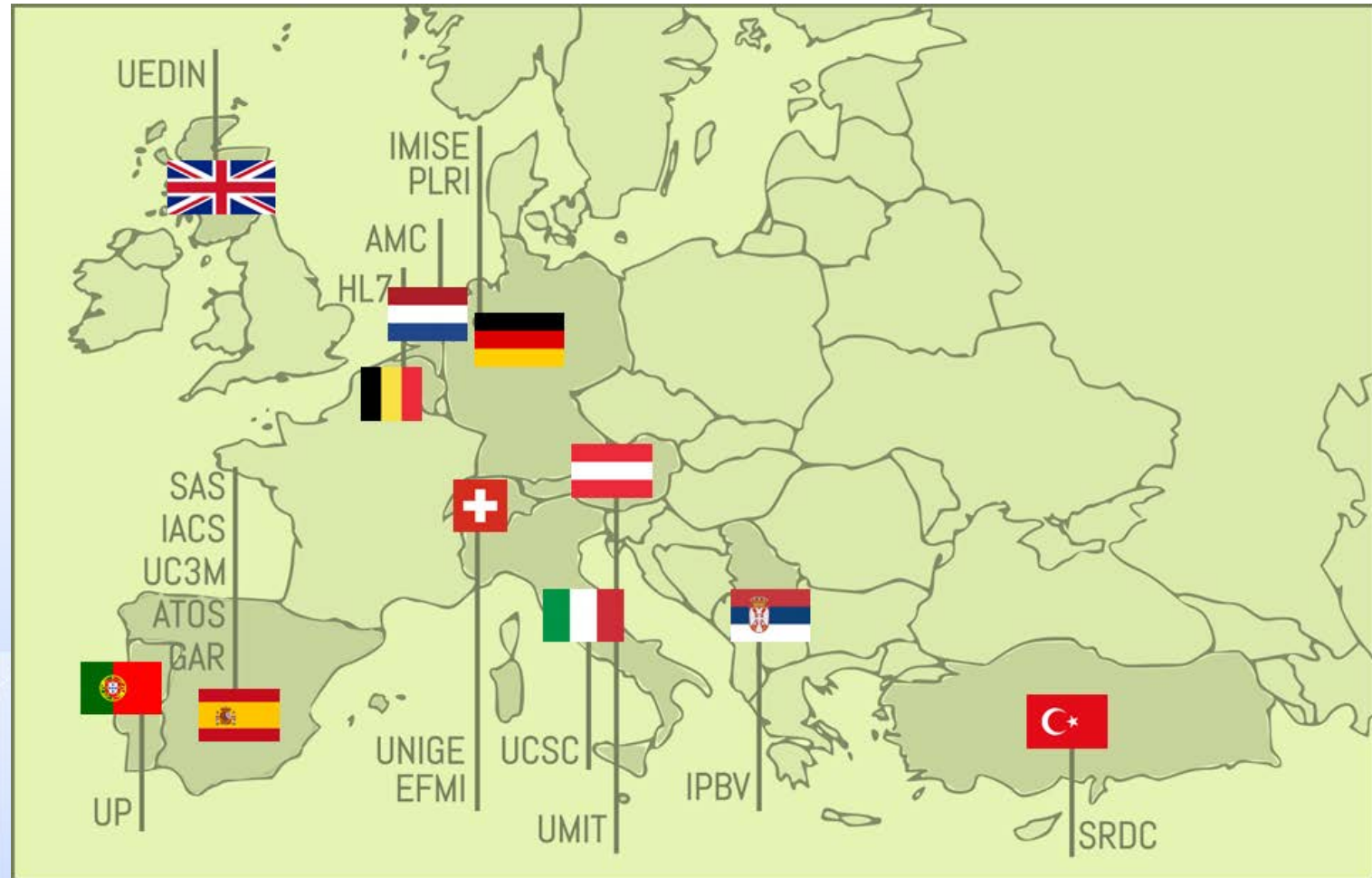
17 socios de **11 países EU y no-EU**

6 organizaciones de investigación en salud

2 universidades con expertos en gestión de datos

4 socios académicos con gran experiencia en informática médica

5 socios del mundo de los negocios (empresas)



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666



European Commission





Objetivo



Facilitar y alentar a la comunidad de investigadores en Salud en la UE **para que “FAIRifiquen” compartan y reutilicen** sus datasets derivados de iniciativas de investigación financiadas con fondos públicos.



ESTRATEGIA DE DIFUSIÓN A ESCALA EUROPEA



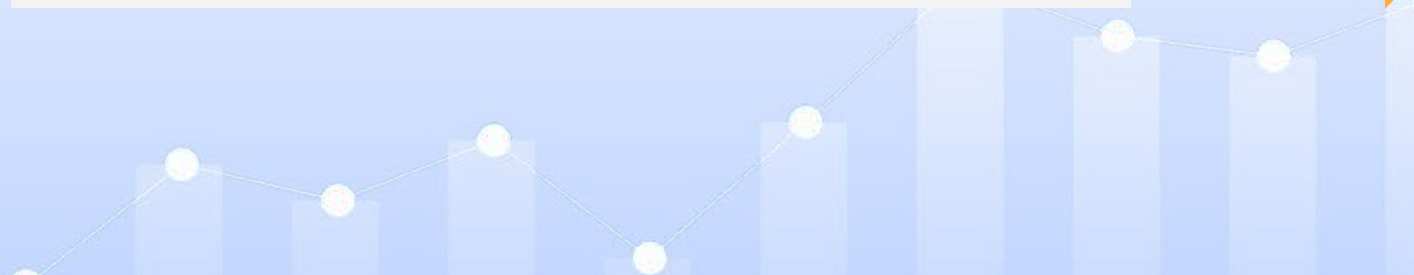
MAPA DE RUTA PARA CERTIFICACIÓN DE DATOS FAIR



PLATAFORMA TECNOLÓGICA



DEMOSTRACIÓN DEL IMPACTO POTENCIAL



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666



European Commission





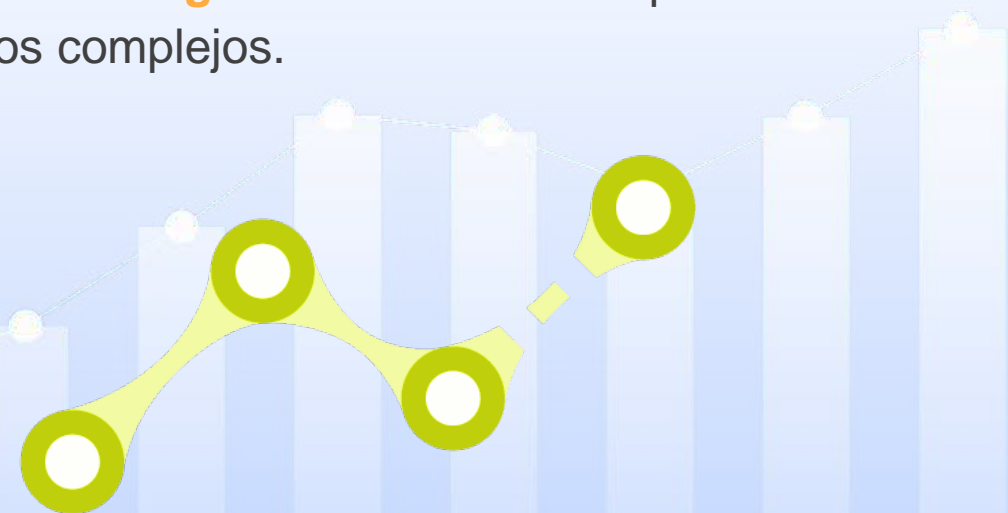
Estudios de casos

Servicios innovadores de eSalud basados en la reutilización de datos FAIR:

#1 Identificar patrones de multimorbilidad y su correlación con el número de drogas prescritas. Utilizando datos demográficos y datos clínicos aplicar técnicas de análisis multivariable, de modelización predictiva, etc para **descubrir los desencadenantes que provocan la aparición de una nueva enfermedad** crónica (riñón, corazón, etc.)



#2 Desarrollar y realizar un proyecto piloto para un **servicio de predicción para el riesgo de reingreso de 30 días** en pacientes crónicos complejos.








This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666



European Commission



Datasets

Name	The EpiChron Cohort	
Owner	SHELTER	
Funding Source	DPIData	
Owner	IBENC	
Funding Source	EPID SURVEY	

Datos de salud y cuidados sociales de **5 millones** de personas

Owner	Virgen del Rocío University Hospital (SAS)
Funding Source	Public funding for healthcare delivery, not for research
Size	2.32 million EHR with information about 18.83 million clinical episodes.


This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666




European Commission




Datasets available to the project


Name	The EpiChron Cohort 
Owner	Instituto Aragonés de Ciencias de la Salud (IACS)
Funding Source	Institute of Health Carlos III for Strategic Actions in Health (Grant Number PI11/1126), the Program INTERREG V-A Spain-France-Andorra of the European Commission (POCTEFA) and private funds.
Size	1.3 million inhabitants approx.


Name	DPIData 
Owner	Hôpitaux Universitaires de Genève (UNIGE)
Funding Source	Public funding for healthcare delivery, not for research
Size	1.3 Mill. patients, 90 Mill. encounters, 40 Mill. order entries, 400 Mill. acts scheduled, 120 Mill. lab results, 30 Mill. documents, 200 Mill. structured clinical facts.

Name	SHELTER 
Owner	Università Cattolica del Sacro Cuore (UCSC)
Funding Source	7 th Framework Programme of the EU
Size	4156 nursing home residents followed for 1 year (Follow up at 6 and 12 months)


Name	IBENC 
Owner	Università Cattolica del Sacro Cuore (UCSC)
Funding Source	7 th Framework Programme of the EU
Size	2884 home care clients followed for 1 year

Datasets available to the project

Name	FRAILSURVEY 
Owner	Porto4Ageing / Elisio Costa (UPorto)
Funding Source	Scale AHA - Support to scaling up of innovations in Active and Healthy Ageing. European Commission
Size	34 x 1500 + 39 x 500

Name	SPRINTT 
Owner	Università Cattolica del Sacro Cuore (UCSC)
Funding Source	Innovative Medicine Initiative (IMI)
Size	1500 older adults followed for 1 year

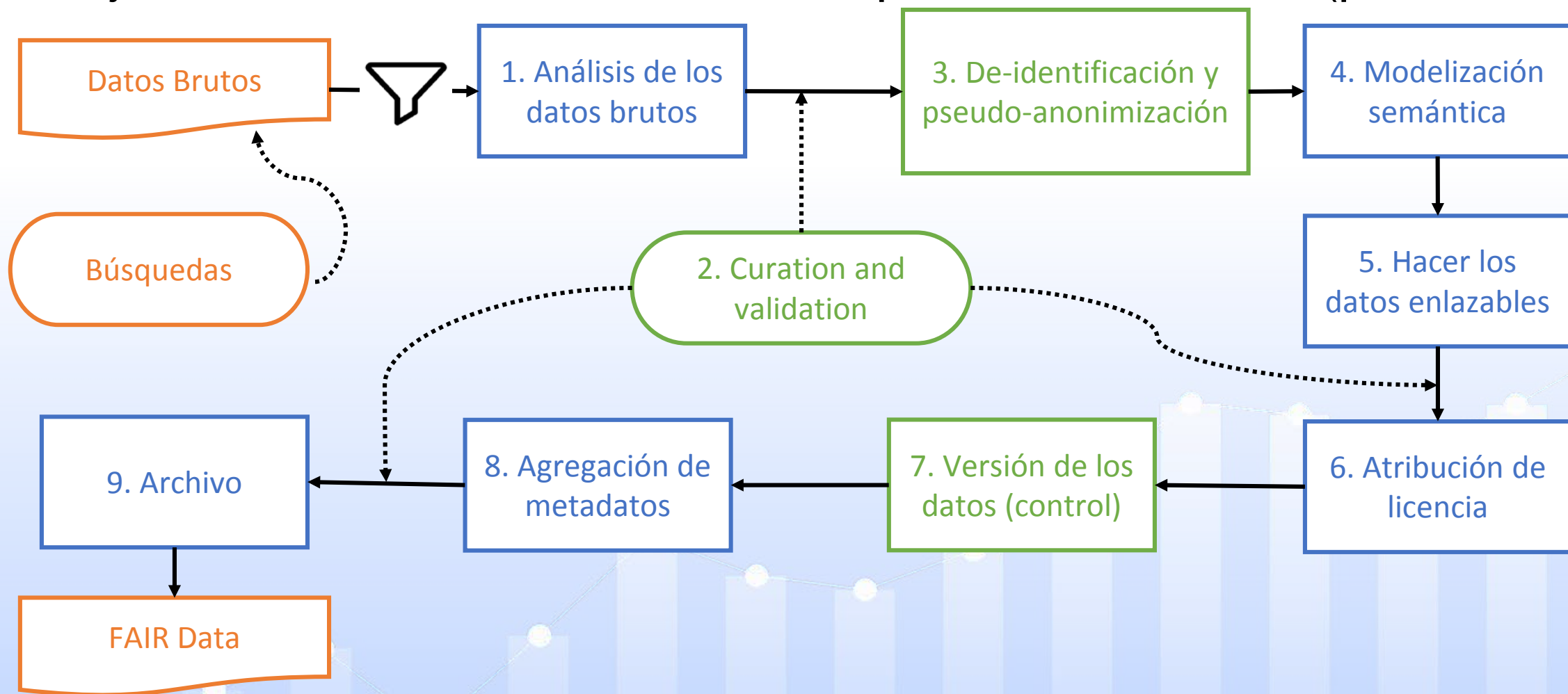
Name	Diraya AE HUVR 
Owner	Virgen del Rocío University Hospital (SAS)
Funding Source	Public funding for healthcare delivery, not for research
Size	2.32 million EHR with information about 18.83 million clinical episodes.

Name	SmokeFreeBrain 
Owner	Institute for Pulmonary Diseases of Vojvodine (IPBV)
Funding Source	SmokeFreeBrain project (GA 681120), H2020, European Commission
Size	60 patients x 53 observations with follow-up until 6 months.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666



El flujo de FAIRificación workflow, adaptado de GO FAIR (provisional)



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666



European
Commission



El proceso de FAIRificación explicado 1/3

Datos brutos

Extraídos por los investigadores de datos clínicos, dispositivos móviles, datos poblacionales, datos genómicos, etc. Es necesario controlar las búsquedas.

Análisis de los
datos brutos

Estructuras de los datasets, tipos de campo, nombres de campo, cardinalidad, etc.

Curación y
validación de datos

Comprobar valores esperados, extraer conceptos clínicos, diagnósticos, medicación, etc. En general, incrementar la calidad del dataset.

de- identificación y
Anonimización

Una u otra opción dependiendo, entre otras cosas, de la frecuencia periódica o no de la actualización de los datos.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666



European
Commission



El proceso de FAIRificación explicado 2/3

Modelización
semántica

Describiendo el significado de las entidades y relaciones en el dataset, de forma precisa, sin ambigüedades y accionada de forma automatizada (ontologías, vocabularios...)

Hacer los datos
enlazables

Aplicando el modelo semántico definido en el paso anterior.

Atribución de
licencias

Licencias Creative u opciones de licencias “ad-hoc”, licencias que permitan el acceso a los datasets bajo demanda y tras aprobación de un comité.

**Versión de los
datos**

Identificación de cada subconjunto y version de los datos utilizados (pid, sellos de tiempo, ordenaciones estables, unicidad de las búsquedas...) de RDA WG sobre Data Citation

El proceso de FAIRificación explicado 3/3

Agregación de metadatos

Que permitan declarar la procedencia del dataset. Y, lógicamente, que contenga metadatos descriptivos, administrativos y estructurales. Gran parte se pueden encontrar en el esquema de metadatos para la investigación clínica de ECRIN (European Clinical Research Information Network)

Archivo

Repositorios internos o externos con sellos CoreTrustSeal

Poniendo attlist cada vez que defines un atributo:

Datos FAIR

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666



European Commission



Algunos requisitos para el eje 1 de FAIR: Findability

1. Crear un **catálogo de datos** disponibles sobre datos de salud (describiendo el contenido de las colecciones de datos y su localización)
2. Crear o reutilizar **vocabularios** para describir datos.
3. Proporcionar **identificadores** persistentes.
4. Crear **catálogo de metadatos** de los elementos de datos contenidos en el catálogo de datos.

Algunos requisitos para el eje 2 de FAIR: Accesibilidad

1. Establecer y gestionar **repositorios** para almacenar y **archivar** datos a largo plazo.
2. Establecer una Comisión de Acceso a Datos para comprobar y aprobar las **propuestas de uso de datos**.
3. Establecer servicios de contactos para **asesorar** a los investigadores clínicos sobre compilación, compartición, protección de datos, etc.

Algunos requisitos para el eje 3 de FAIR: Interoperabilidad

1. Promover el **uso de estándares** aceptados por las distintas comunidades científicas (HL7, IHE, CDISC, ISO)
2. Ofrecer servicios de **terminología para enlazar datos** con conceptos de las diferentes terminologías médicas
3. Establecer servicios de anotación para **enriquecer datos con metadatos**.

Algunos requisitos para el eje 4 de FAIR: Reutilización

1. Proporcionar información y **asesorar** sobre las bases legales para la reutilización de datos.
2. Proporcionar información y **asesorar** sobre las diferentes licencias de datos.
3. Especificar la **procedencia de las colecciones de los datos**.
4. Conocer y **facilitar herramientas para el uso y la visualización de datos complejos**.

GRACIAS

Gràcies

Transformació de les biblioteques de ciències de la salut: reptes i noves competències

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666



European
Commission

