



# DEEPHEALTH

## Hackathon - Course 5

### De-identification of Structured Data

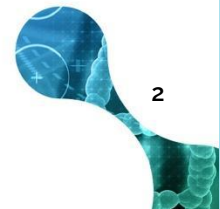


*The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.*



# Index

1. Structured Data
2. De-identification methodologies
3. Deploying Pseudonymisation Techniques
4. Pilot project





DEEPHEALTH

# Structured Data



The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.



# Electronic Health Records (EHR)

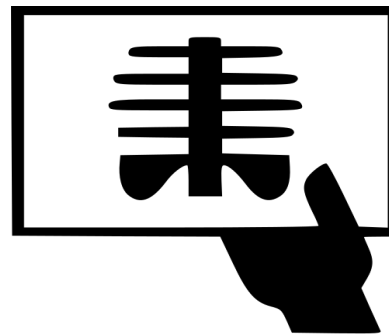
“Systematized collection of patient and population electronically stored health information in a digital format”  
Structured and (mostly) Unstructured Data



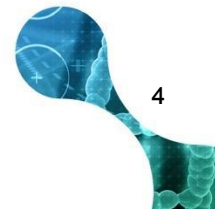
Tests results



Omics Data

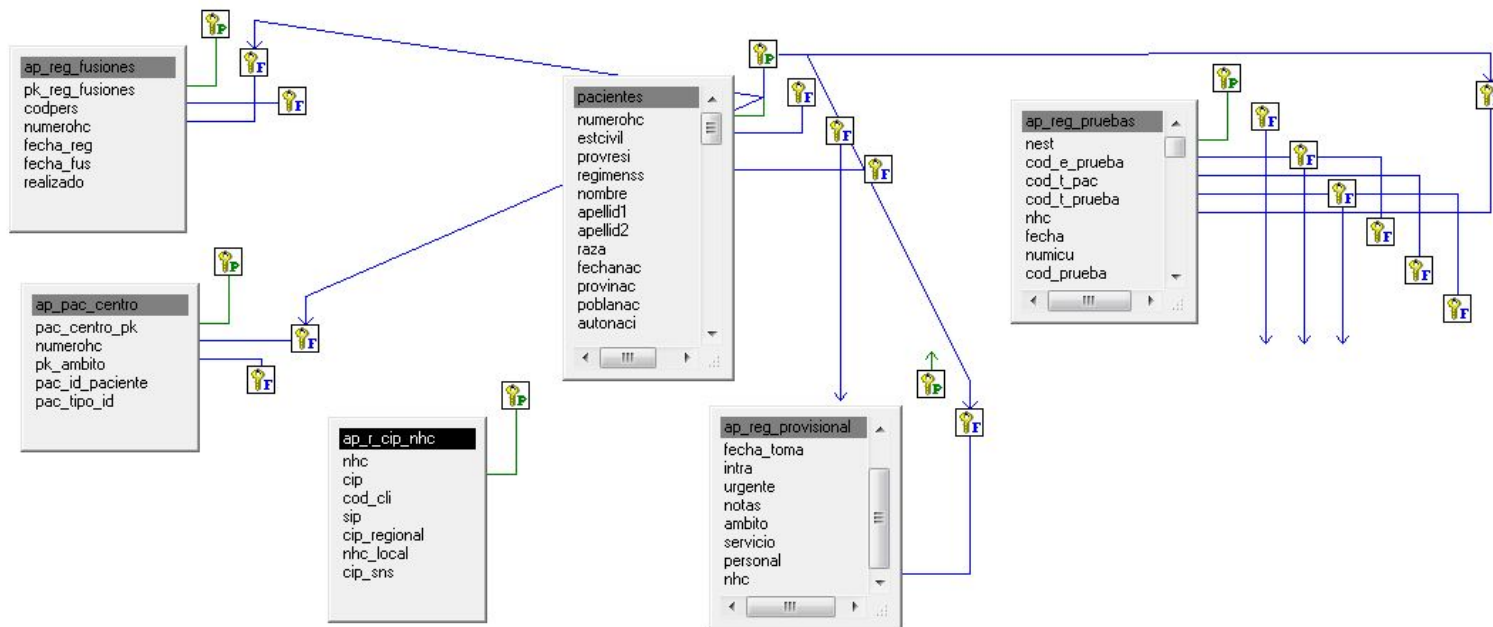


Medical Imaging





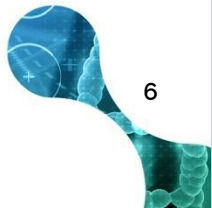
# Structured Data





# Uses of structured information

1. Big Data
2. Datasets for AI
3. Clinical trials
4. Test Environments
5. Training Environments





DEEPHEALTH

# De-identification methodology

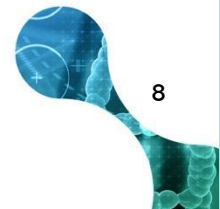


The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.



# De-identification methodology

1. Objectives
2. Identification of sensitive data
3. Location in tables / fields
4. Select masking techniques
5. Creation of auxiliary tables
6. Creating random generation functions
7. Creation of ETL services
8. Referential integrity management





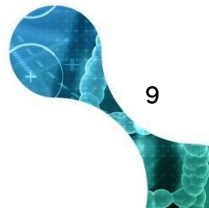
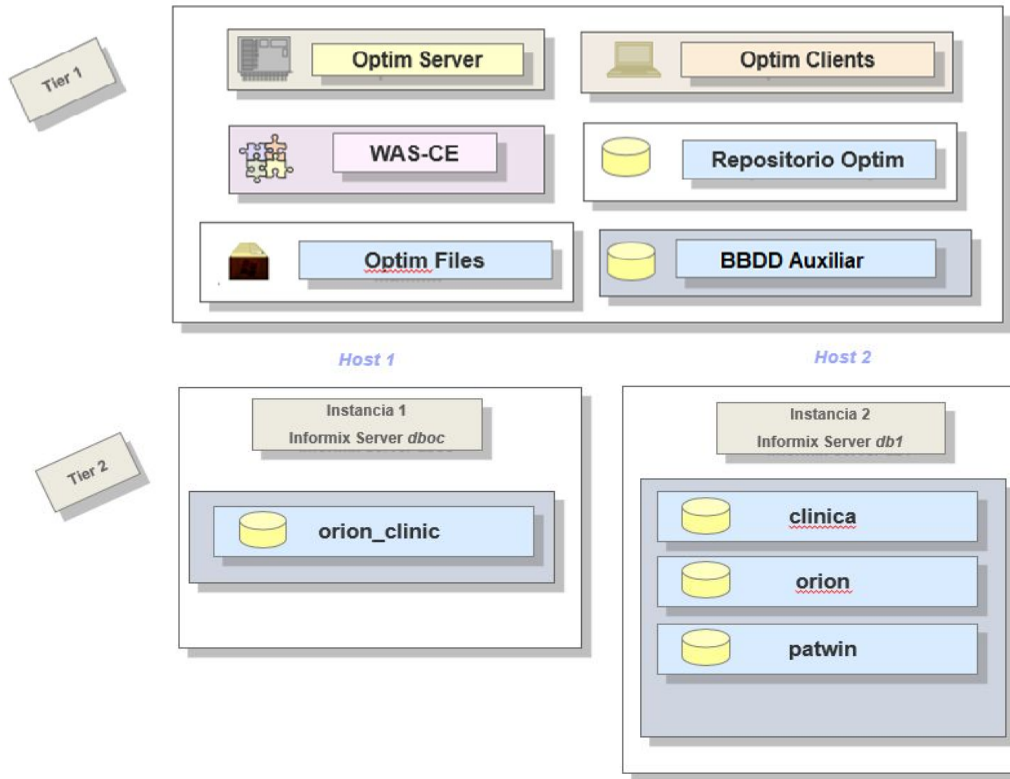


# Technological Solution

Generalitat Valenciana - Optim topology

Optim  
Topology

Data  
sources  
Topology





DEEPHEALTH

# Deploying Pseudonymisation Techniques

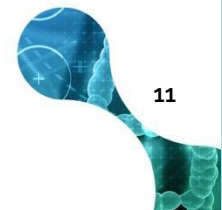


The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.



# Pseudonymisation Techniques

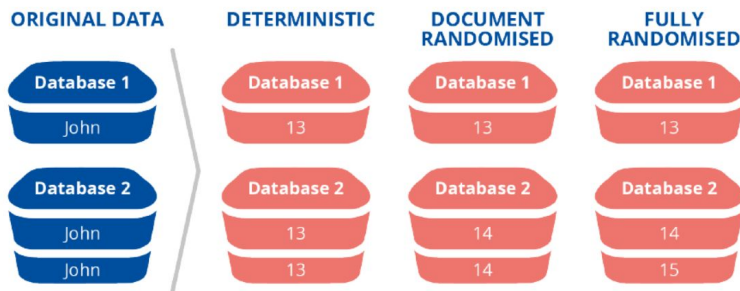
Technique	Pseudonym Generator
Random	The Random Function returns a number selected at random within the range indicated.
Shuffle	The Shuffle Function replaces a source value with another value from the column that is then inserted in a destination column.
Lookup Function	The Lookup Function obtains the value for a destination column from a lookup table, according to the value in a source column.
Hash Lookup Function	The Hash Lookup Function obtains the value for a destination column from a lookup table, according to a hashed value derived from a source column.
TRANS COL Function	The TRANS COL function can mask data that has no inherent format or a format that is not widely known. TRANS COL maintains the format and character type of the source data at the destination.
Affinity	Use the affinity privacy provider to mask data while maintaining the format and character types of the source values.
LUA Functions	Procedure that is used to mask or transform the data in a column





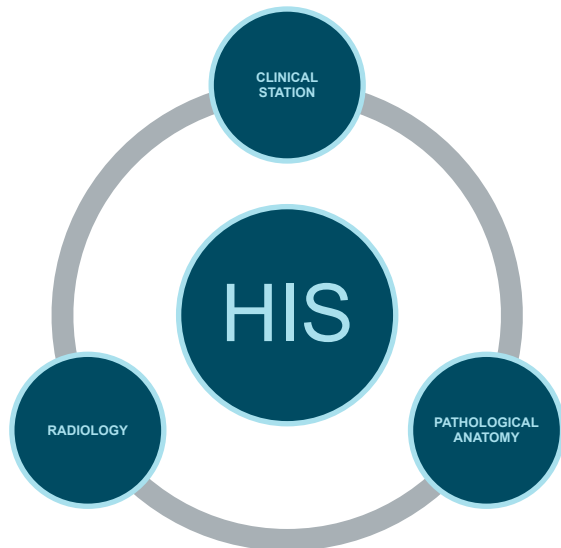
# Application of pseudonymisation policies

1. Deterministic pseudonymisation
2. Randomised pseudonymisation document
3. Fully randomised pseudonymisation



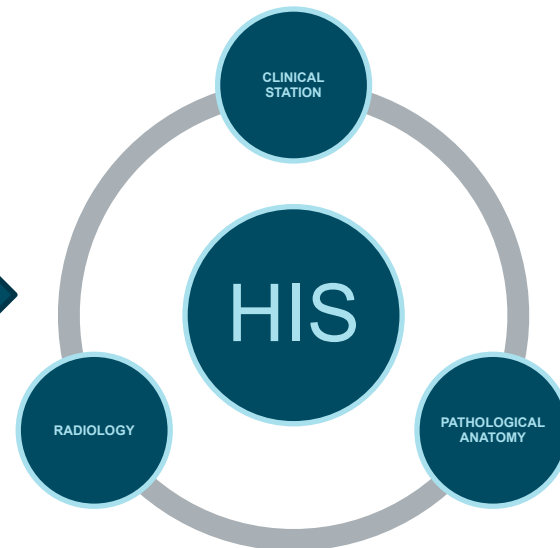
# Pilot Project

## PRODUCTION



PSEUDONYMISATION

## PRE-PRODUCTION



# Original Data

## H Datos administrativos

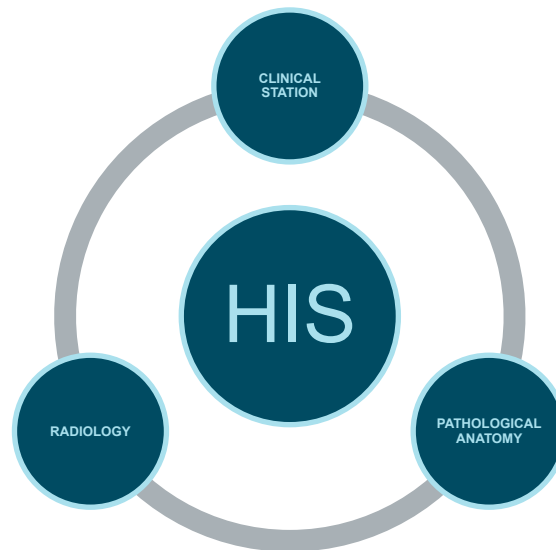
Centro	HOSPITAL GENERAL DE CASTELLON	NHC	33
SIP	7581569		

## 👤 Datos personales

Nombre	CIUDADANO	Nacionalidad	ESPAÑOLA
Apellido 1	FICTICIO	Fecha nac.	01/12/1994
Apellido 2	ACTIVO	Localidad	ALELLA
Sexo	Varón	Prov. nac.	BARCELONA
NIF	92920000T	Pais nac.	ESPAÑA
NSS	05/19999999/71		

## 🏠 Datos primera residencia

AVDA	VALENCIA
Nº	34 Bloque Esc. Piso Puerta
Municipio	LA PLANA/CASTELLO DE LA PLANA
CP	12006
Provincia	CASTELLON
País	ESPAÑA
Teléfono	964323222





# Pseudonymisation

## ORIGINAL DATA



Name: **John**  
Surname: **SMITH**  
Tel: **6548827421**  
Age: **44**

## ASSOCIATION TABLE



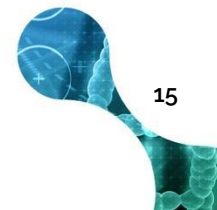
	Pre-P	Post-P
Name	John	aa1f
Surname	SMITH	ac4fb
Tel	6548827421	gri394j2h
Age	44	44

Pseudonyms

## PSEUDONYMISED DATA



Name: **aa1f**  
Surname: **ac4fb**  
Tel: **gri394j2h**



# Pseudonymised Data

## H Datos administrativos

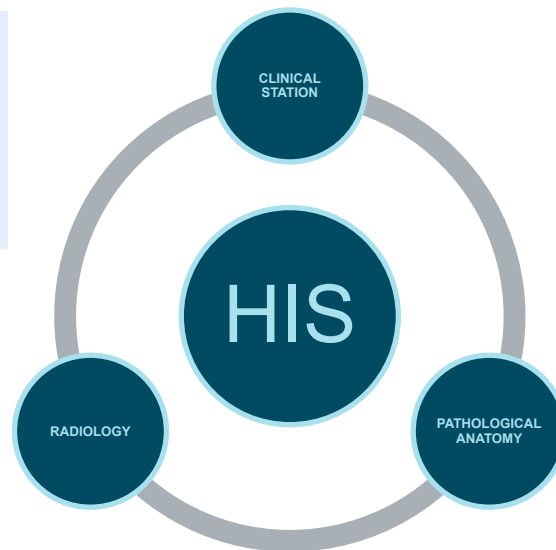
Centro	HOSPITAL GENERAL DE CASTELLON	NHC	2897108
SIP	52601373		

## Datos personales

Nombre	CELEDONIO	Nacionalidad	ESPAÑOLA
Apellido 1	TRINIDAD	Fecha nac.	01/12/1994
Apellido 2	ALBERT	Localidad	CALAF
Sexo	Varón	Prov. nac.	BARCELONA
NIF	95123701W	País nac.	ESPAÑA
NSS	46/10604599/17		

## Datos primera residencia

CALLE	CERVANTES								
Nº	100	Bloque		Esc.		Piso		Puerta	
Municipio	VALENCIA								
CP	46007								
Provincia	VALENCIA								
País	ESPAÑA								
Teléfono	964781569								







DEEPHEALTH

**Thank you for your  
attention!**



*The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.*