

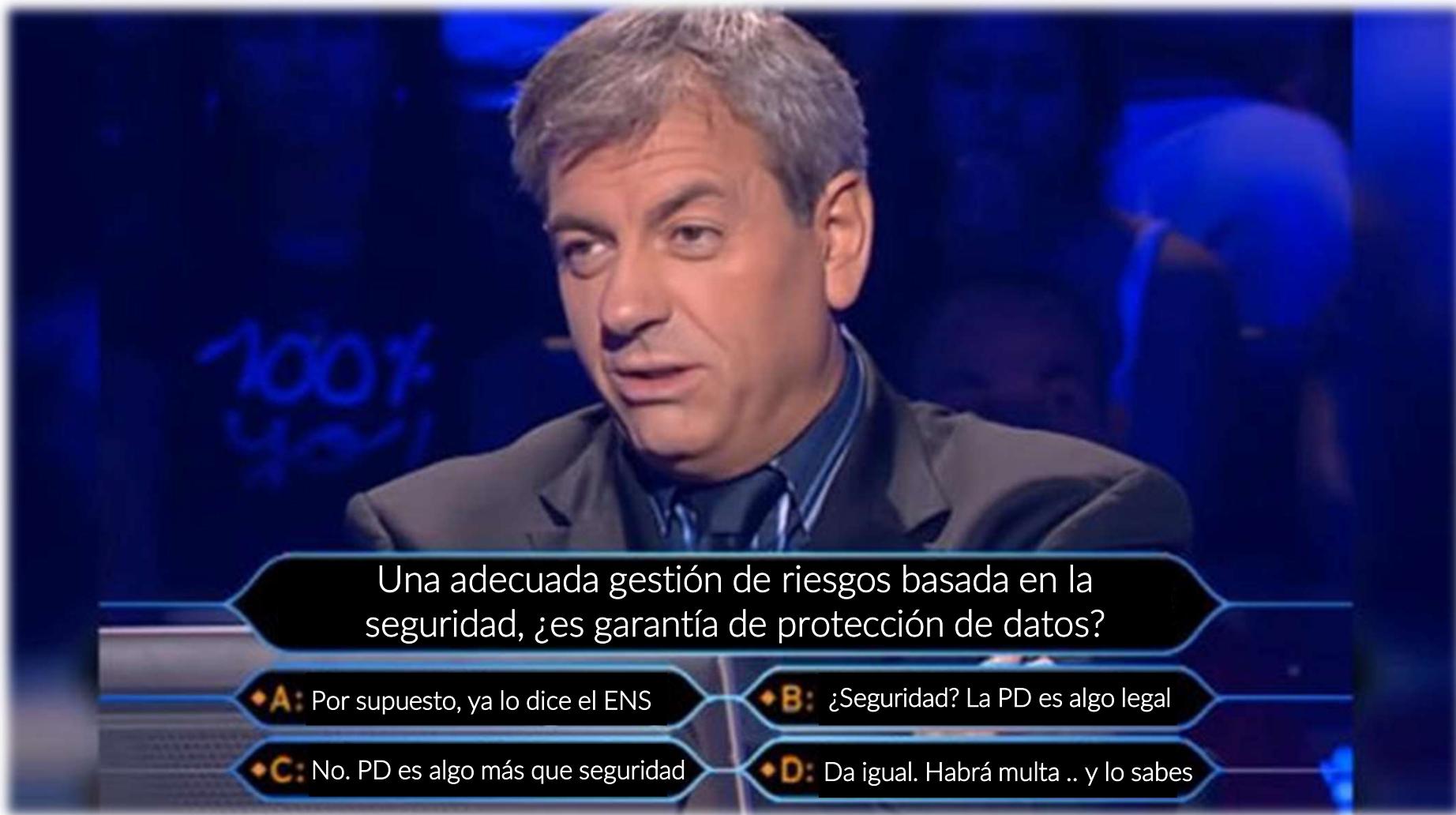
# Una aproximación técnica a la privacidad: la k-anonimización

M. Blanca Muñoz Pérez  
Jefa de Departamento de Estrategia del Dato  
del Ayuntamiento de Madrid



10 marzo 2023

## Comenzamos con la pregunta del millón ...



Una adecuada gestión de riesgos basada en la seguridad, ¿es garantía de protección de datos?

• A: Por supuesto, ya lo dice el ENS

• C: No. PD es algo más que seguridad

• B: ¿Seguridad? La PD es algo legal

• D: Da igual. Habrá multa .. y lo sabes

## Comenzamos con la pregunta del millón ...

Una adecuada gestión de riesgos basada en la seguridad, ¿es garantía de protección de datos?

• A: Por supuesto, ya lo dice el ENS

• C: No. PD es algo más que seguridad

• B: ¿Seguridad? La PD es algo legal

• D: Da igual. Habrá multa .. y lo sabes

## Seguridad vs Protección de Datos

---



## Objetivos de protección

### Objetivos de Seguridad



Confidencialidad



Integridad



Disponibilidad

### Objetivos de Privacidad



Desvinculación

(unlinkability)



Transparencia

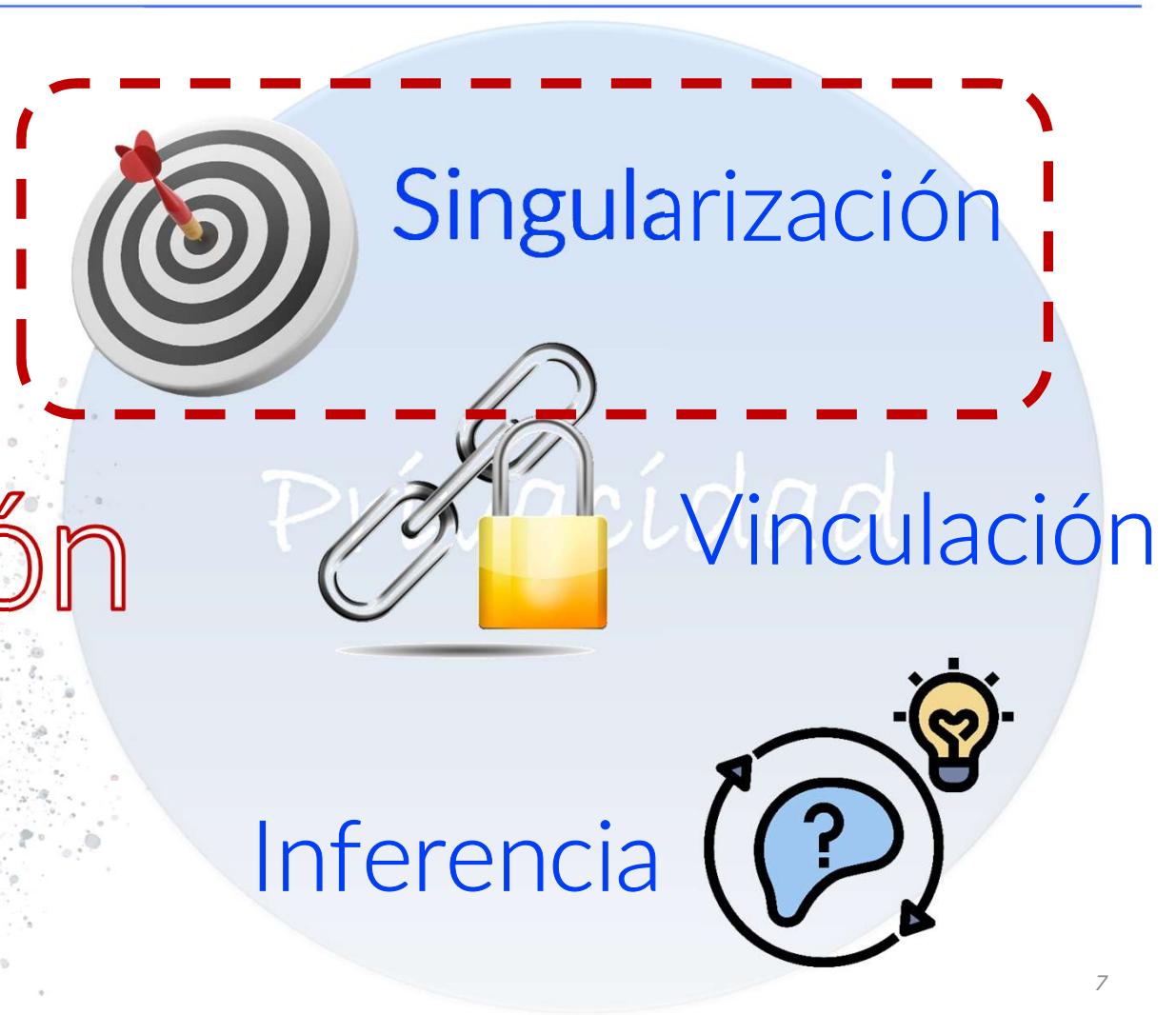
Capacidad de control  
(intervenability)

# Medidas de protección



## Riesgos asociados a la desvinculación

Desvinculación



## Una película ... (*puede que de miedo*)

---



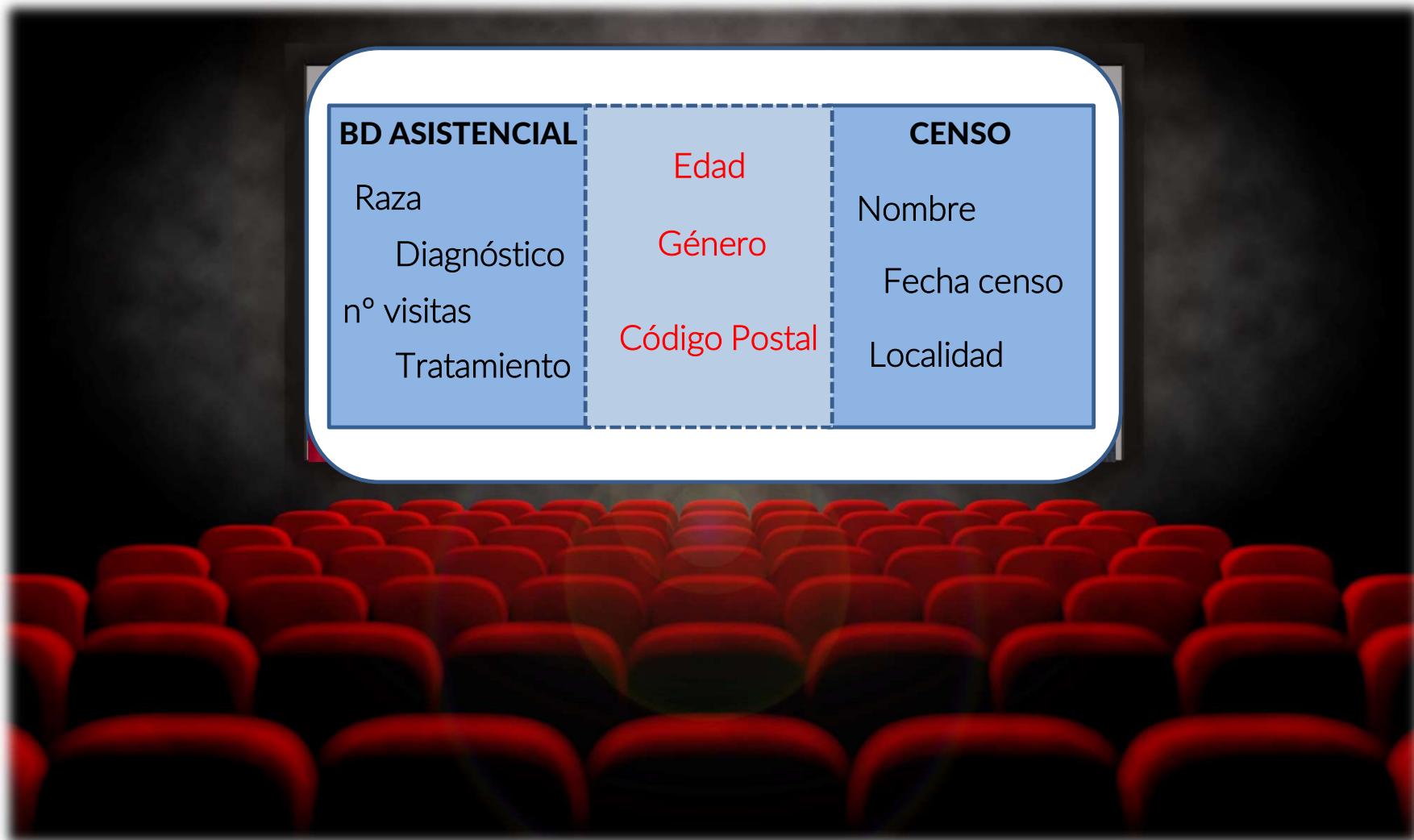
Encuéntrame si puedes

---



Encuéntrame si puedes

---



Encuéntrame si puedes

---



¿Qué falló?

1. La BD sensible contenía **datos demográficos** asociados a los diferentes registros
2. Había una **BD secundaria** con los mismos datos demográficos
3. El sujeto objetivo estaba en los **dos conjuntos** de datos
4. La información demográfica era **única**

## Concepto importante - *Cuasi-identificadores*

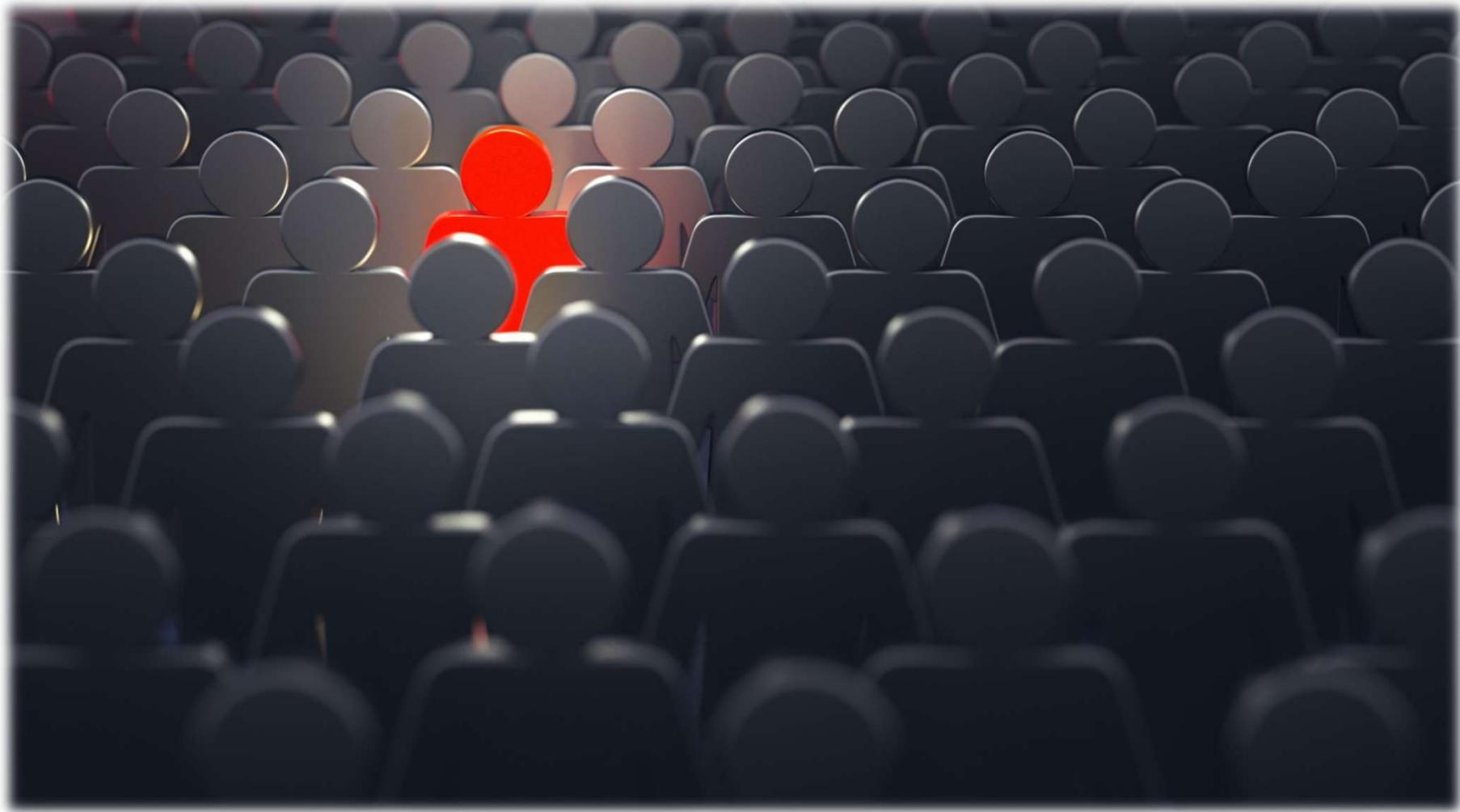


Desidentificación

Anonimización

¿Cómo soy de **único**?

---



# ¿Cómo soy de único?

**How unique am I?**  
Find out how much different you are among the masses.

Fill out the form below to see how unique you are, and therefore how easy it is to identify you from these values.  
Please note that this service is still under development.

Date of Birth	March	10	2003
Gender	<input type="radio"/> Male	<input checked="" type="radio"/> Female	
ZIP Code	11236. This is still under development.		
Date of Birth	March	10	2003
Gender	<input type="radio"/> Male	<input checked="" type="radio"/> Female	
ZIP Code	11236 ZIP code must be 5 digits long.		

**About** **Samples**

**Your Profile**

Gender: Female  
ZIP Code: 11236 (pop. 93877 )

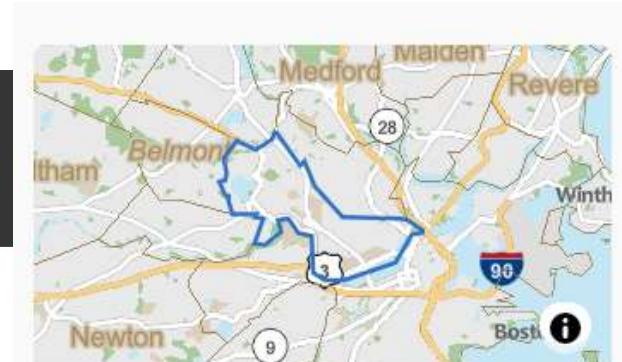
**Date of Birth** 3 / 10 / Easily identifiable by birthdate (about 1).

**Your Profile**

Gender: Female  
ZIP Code: 11236 (pop. 93877 )

**Date of Birth** 3 / 10 / 2003  
**Birth Year** 2003  
**Range** 2003 to 2007  
Easily identifiable by birthdate (about 1).  
Lots with your birth year (about 649 ).  
Wow! There are lots of people in the same age range (about 3245 ).

**Submit →**



## Place

### Cambridge city, Massachusetts

Total Population: 118,403

Median Household Income: \$104,746

Bachelor's Degree or Higher: 81.4%

Employment Rate: 66.9%

Total Housing Units: 53,907

Without Health Care Coverage: 2.6%

Total Households: 52,487

Hispanic or Latino (of any race): 10,738

[View Profile](#) >

Fuentes: [Identity • About My Info](#)  
[Zip Codes by Population 2023](#)

Fuente: [Cambridge - Census Bureau Search](#)

¿Cómo soy de único?

# Moratalaz

Población a 1 de enero: 93.810

Fuente: [Portal de datos abiertos del Ayuntamiento de Madrid](#)



## Modelo de la K-anonimidad

---

Basado en la **generalización** y en la **supresión**



“la información que se libera de **cada individuo** incluido en un conjunto de datos **no puede diferenciarse** de la información de otros  $k - 1$  individuos incluidos en ese mismo conjunto de datos”

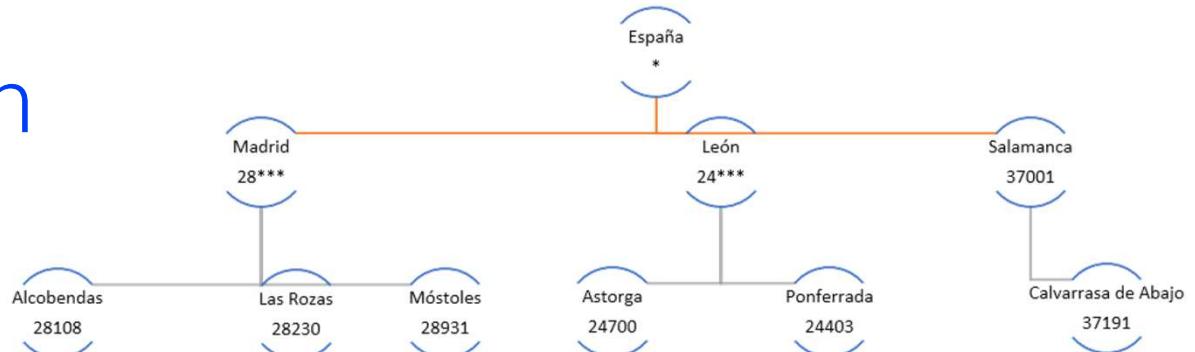
# Modelo de la K-anonimidad

## 1 – Aplicar generalización

- Rangos
- Jerarquías
- Categorizaciones

## 2 – Aplicar supresión

- Valores atípicos



Grupo de generalización	Categorías profesionales
1	Ingenieros y Licenciados. Personal de alta dirección no incluido en el artículo 1.3.c) del Estatuto de los Trabajadores.
2	Ingenieros Técnicos, Peritos y Ayudantes Titulados.
3	Jefes Administrativos y de Taller.
4	Ayudantes no Titulados.
5	Oficiales Administrativos.
6	Subalternos.
7	Auxiliares Administrativos.
18 - 30 AÑOS	
Grupo de generalización	Categorías profesionales
8	Oficiales de primera y segunda.
9	Oficiales de tercera y Especialistas.
10	Peones.
11	Trabajadores menores de dieciocho años, cualquiera que sea su categoría profesional.



18 - 30  
AÑOS



30 - 35  
AÑOS



35 - 50  
AÑOS

# Aterrizando la k-anonimidad

Base de Datos sin anonimizar

Paciente	Código Postal	Edad	Género	Enfermedad
Patrick	28942	22	Hombre	Cardiovascular
Sandra	28015	10	Mujer	Cáncer
Mateo	28032	36	Hombre	Respiratoria
Ana	28233	46	Mujer	Sano
Sebastian	28932	23	Hombre	Respiratoria
Alicia	28223	47	Mujer	Cáncer
Jorge	28007	42	Hombre	Hígado
Elisa	28230	54	Mujer	Cardiovascular
José	28911	18	Hombre	Sano
Luis	28005	38	Hombre	Sano

Paciente	Código Postal	Edad	Género	Enfermedad
Patrick	28942	22	Hombre	Cardiovascular
José	28911	18	Hombre	Sano
Sebastian	28932	23	Hombre	Respiratoria
Sandra	28015	10	Mujer	Cáncer
Luis	28005	38	Hombre	Sano
Mateo	28032	36	Hombre	Respiratoria
Jorge	28007	42	Hombre	Hígado
Ana	28233	46	Mujer	Sano
Alicia	28223	47	Mujer	Cáncer
Ana	28233	46	Mujer	Sano

# Aterrizando la k-anonimidad

Base de Datos sin anonimizar

Paciente	Código Postal	Edad	Género	Enfermedad
Patrick	28942	22	Hombre	Cardiovascular
Sandra	28015	10	Mujer	Cáncer
Mateo	28032	36	Hombre	Respiratoria
Ana	28233	46	Mujer	Sano
Sebastian	28932	23	Hombre	Respiratoria
Alicia	28223	47	Mujer	Cáncer
Jorge	28007	42	Hombre	Hígado
Elisa	28230	54	Mujer	Cardiovascular
José	28911	18	Hombre	Sano
Luis	28005	38	Hombre	Sano

Paciente	Código Postal	Edad	Género	Enfermedad
Patrick	289xx	15-24	Hombre	Cardiovascular
Sebastian	289xx	15-24	Hombre	Respiratoria
José	289xx	15-24	Hombre	Sano
Sandra	28015	10	Mujer	Cáncer
Jorge	280xx	35-44	Hombre	Hígado
Mateo	280xx	35-44	Hombre	Respiratoria
Luis	280xx	35-44	Hombre	Sano
Ana	282xx	45-54	Mujer	Sano
Alicia	282xx	45-54	Mujer	Cáncer
Elisa	282xx	45-54	Mujer	Cardiovascular

# Aterrizando la k-anonimidad

Base de Datos sin anonimizar

Paciente	Código Postal	Edad	Género	Enfermedad
Patrick	28942	22	Hombre	Cardiovascular
Sandra	28015	10	Mujer	Cáncer
Mateo	28032	36	Hombre	Respiratoria
Ana	28233	46	Mujer	Sano
Sebastian	28932	23	Hombre	Respiratoria
Alicia	28223	47	Mujer	Cáncer
Jorge	28007	42	Hombre	Hígado
Elisa	28230	54	Mujer	Cardiovascular
José	28911	18	Hombre	Sano
Luis	28005	38	Hombre	Sano

Base de Datos 3- anónima

Código Postal	Edad	Género	Enfermedad
289xx	15-24	Hombre	Cardiovascular
289xx	15-24	Hombre	Respiratoria
289xx	15-24	Hombre	Sano
280xx	35-44	Hombre	Respiratoria
280xx	35-44	Hombre	Sano
280xx	35-44	Hombre	Hígado
282xx	45-54	Mujer	Sano
282xx	45-54	Mujer	Cáncer
282xx	45-54	Mujer	Cardiovascular

Riesgo *Intrínseco* = 1 / min (tamaño de las clases de equivalencia) = 1/k

# Aterrizando la k-anonimidad

Base de Datos sin anonimizar

Paciente	Código Postal	Edad	Género	Enfermedad
Patrick	28942	22	Hombre	Cardiovascular
Sandra	28015	10	Mujer	Cáncer
Mateo	28032	36	Hombre	Respiratoria
Ana	28233	46	Mujer	Sano
Sebastian	28932	23	Hombre	Respiratoria
Alicia	28223	47	Mujer	Cáncer
Jorge	28942	24	Hombre	Hígado
Elisa	28230	54	Mujer	Cardiovascular
José	28911	18	Hombre	Sano
Luis	28005	38	Hombre	Sano

Paciente	Código Postal	Edad	Género	Enfermedad
Patrick	289xx	15-24	Hombre	Cardiovascular
Sebastian	289xx	15-24	Hombre	Respiratoria
José	289xx	15-24	Hombre	Sano
Jorge	289xx	15-24	Hombre	Hígado
Sandra	28015	10	Mujer	Cáncer
Mateo	280xx	35-44	Hombre	Respiratoria
Luis	280xx	35-44	Hombre	Sano
Ana	282xx	45-54	Mujer	Sano
Alicia	282xx	45-54	Mujer	Cáncer
Elisa	282xx	45-54	Mujer	Cardiovascular

Riesgo *Intrínseco* = 1 / min (tamaño de las clases de equivalencia) = 1/k

# Aterrizando la k-anonimidad

Base de Datos sin anonimizar

Paciente	Código Postal	Edad	Género	Enfermedad
Patrick	28942	22	Hombre	Cardiovascular
Sandra	28015	10	Mujer	Cáncer
Mateo	28032	36	Hombre	Respiratoria
Ana	28233	46	Mujer	Sano
Sebastian	28932	23	Hombre	Respiratoria
Alicia	28223	47	Mujer	Cáncer
Jorge	28942	24	Hombre	Hígado
Elisa	28230	54	Mujer	Cardiovascular
José	28911	18	Hombre	Sano
Luis	28005	38	Hombre	Sano

Base de Datos 2 - anónima

Código Postal	Edad	Género	Enfermedad
289xx	15-24	Hombre	Cardiovascular
289xx	15-24	Hombre	Respiratoria
289xx	15-24	Hombre	Sano
289xx	15-24	Hombre	Hígado
280xx	35-44	Hombre	Respiratoria
280xx	35-44	Hombre	Sano
282xx	45-54	Mujer	Sano
282xx	45-54	Mujer	Cáncer
282xx	45-54	Mujer	Cardiovascular

Riesgo *Intrínseco* = 1 / min (tamaño de las clases de equivalencia) = 1/k

## ¿Ese es el riesgo reidentificación?

Controles de privacidad y seguridad implementados	Motivación y recursos del atacante	Probabilidad de reidentificación
Alto	Bajo	0.03
	Medio	0.05
	Alto	0.1
Medio	Bajo	0.2
	Medio	0.25
	Alto	0.3
Bajo	Bajo	0.4
	Medio	0.5
	Alto	0.6
Ninguno	Bajo	1.0
	Medio	1.0
	Alto	1.0

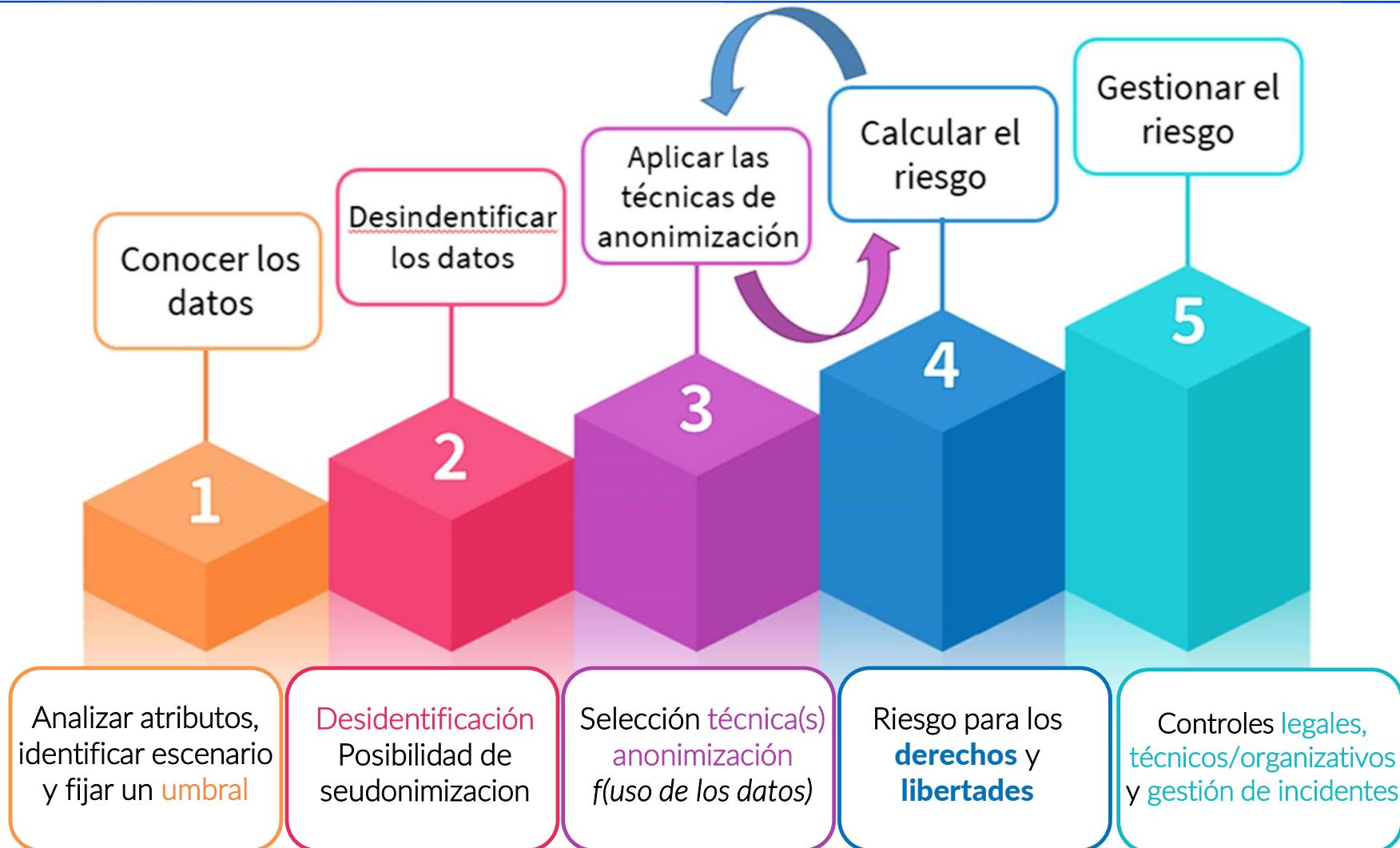
## Riego Escenario:

- Riego reconocimiento accidental
- Riesgo materialización brecha
- Riesgo ataque deliberado

$$\text{Riesgo}_{\text{Extrínseco}} = \max(\text{Riesgo}_{\text{Escenario}})$$

$$\text{Riesgo}_{\text{Global}} = \text{Riesgo}_{\text{Intrínseco}} \times \text{Riesgo}_{\text{Extrínseco}}$$

# Gestión integral del riesgo



## Limitaciones k-anonimidad

---



## Limitaciones k-anonimidad

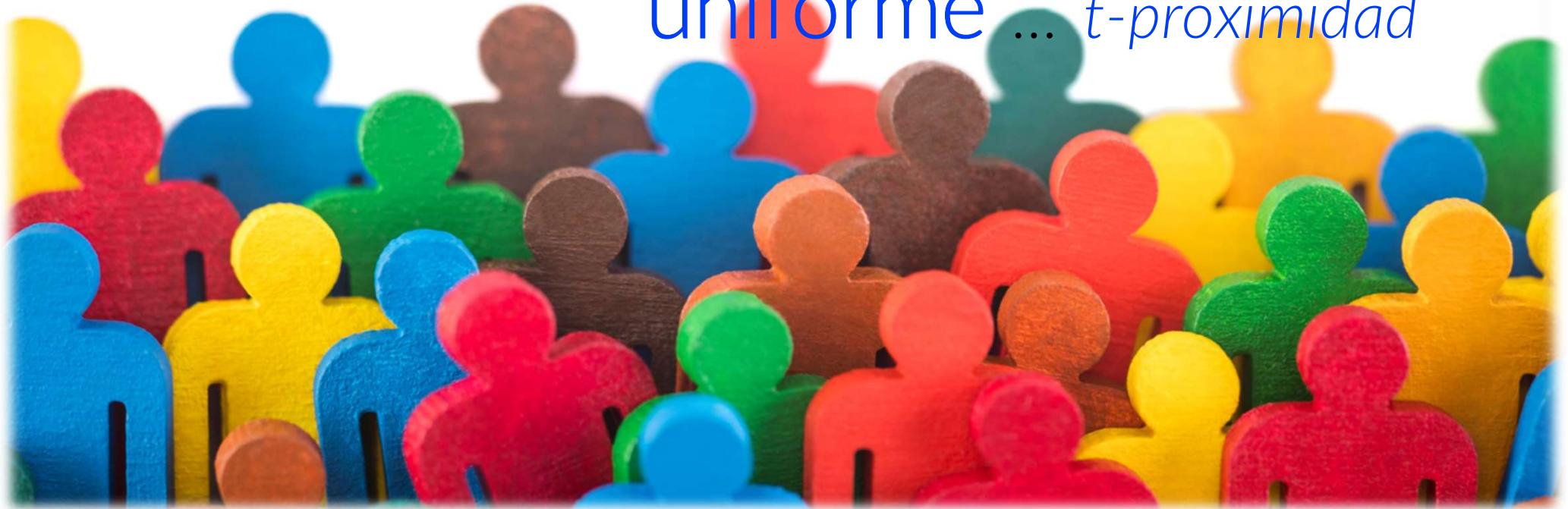
---



## Limitaciones k-anonimidad

---

Introducir **diversidad** ...  $l$ -diversidad  
**uniforme** ...  $t$ -proximidad



pero eso es la trama de una segunda temporada ....



# ¿alguna pregunta?

Recursos:

- [Diez malentendidos relacionados con la anonimización](#)
- [Guía básica de anonimización](#)
- [Herramienta básica de anonimización](#)