



Islas Canarias  
Del 15 al 19 de noviembre de 2021

istac


INSTITUTO CANARIO  
DE ESTADÍSTICA

# XXI Jornadas de Estadística de las Comunidades Autónomas

## Registro Integrado de Población de Canarias

Noelia Martín Morales, ISTAC  
Rafael Betancor Villalba, ISTAC  
Johana Valera Puche, EVM  
Bencomo Domínguez Martín, EVM

<https://www.linkedin.com/company/istac/> 

@istac\_es 

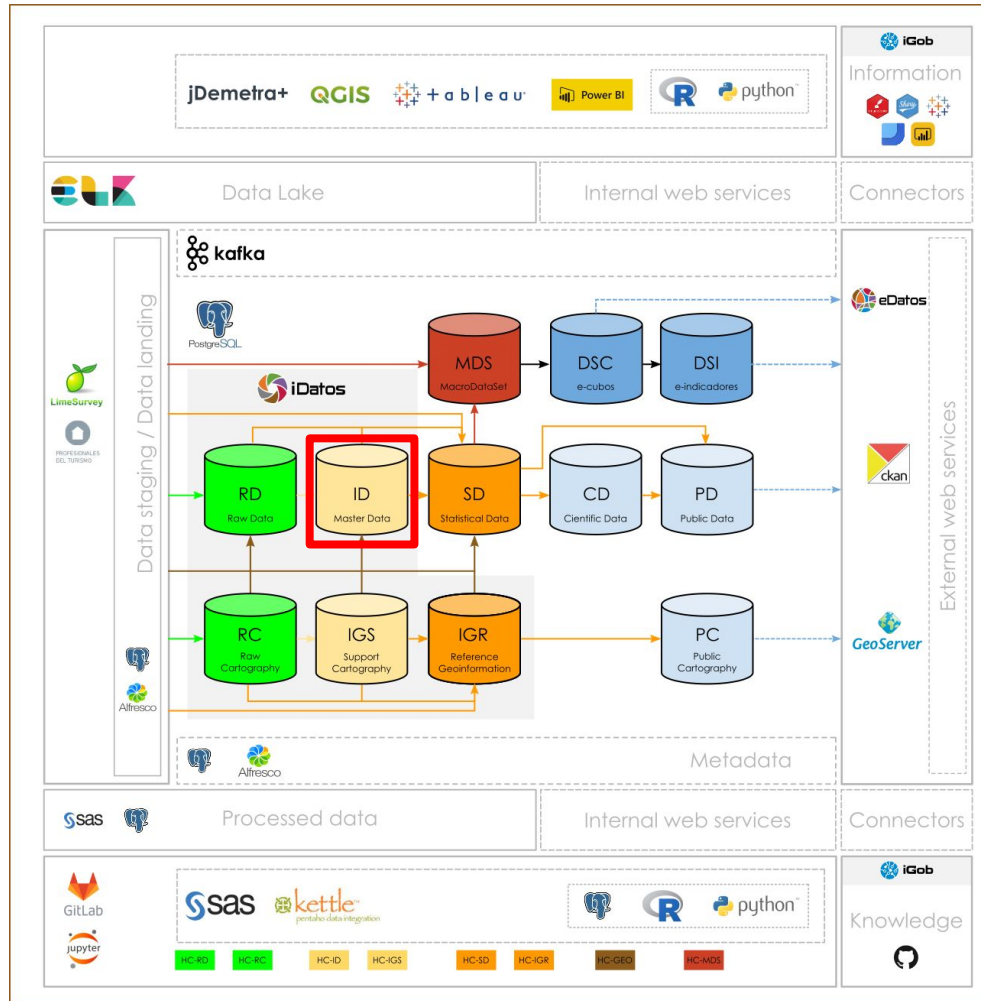
# Índice

- Introducción
- Objetivo
- Metodología
- Resultados
- Casos de uso
- Conclusión

# Introducción

- El artículo 32 de la Ley 1/1991 de Estadística de la Comunidad Autónoma de Canarias indica que se constituirá un banco de datos administrativos para fines estadísticos, que se nutrirá prioritariamente de los ficheros administrativos de Canarias.
- Este banco de datos se elabora a fin de mejorar la eficiencia de la actividad estadística de interés y para su desarrollo se impulsará el Sistema de Datos Integrados (iDatos).

# Introducción: Arquitectura del banco de datos



El proceso de integración se contempla en el Plan Estadístico de Canarias 2018-2022 y en el Programa Estadístico Anual 2021 según el boc-a-2021-073-1949 de 12 de abril de 2021 donde se almacenan los principales directorios dentro del esquema ID:

1. C00047A. Directorio de Calles y Direcciones
2. C00111A. Directorio de Edificios, Viviendas y Locales
3. C00021A. Directorio de Unidades Económicas (DUE)
4. C00063A. Directorio de Población y Hogares

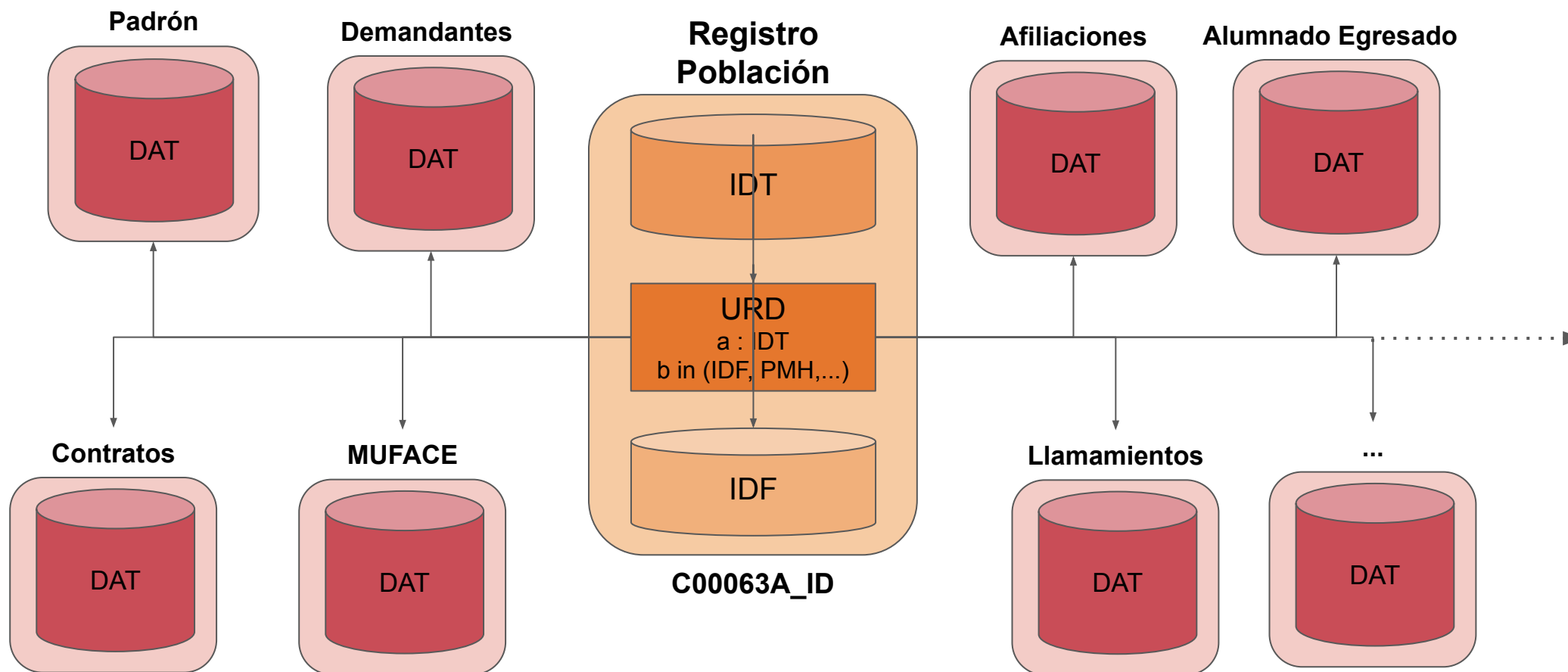
# Introducción: Registro Integrado de Población

Se plantea como una infraestructura estadística integradora de las diferentes fuentes de datos, que permite numerar e identificar a las personas de Canarias según un conjunto básico de variables sociodemográficas.

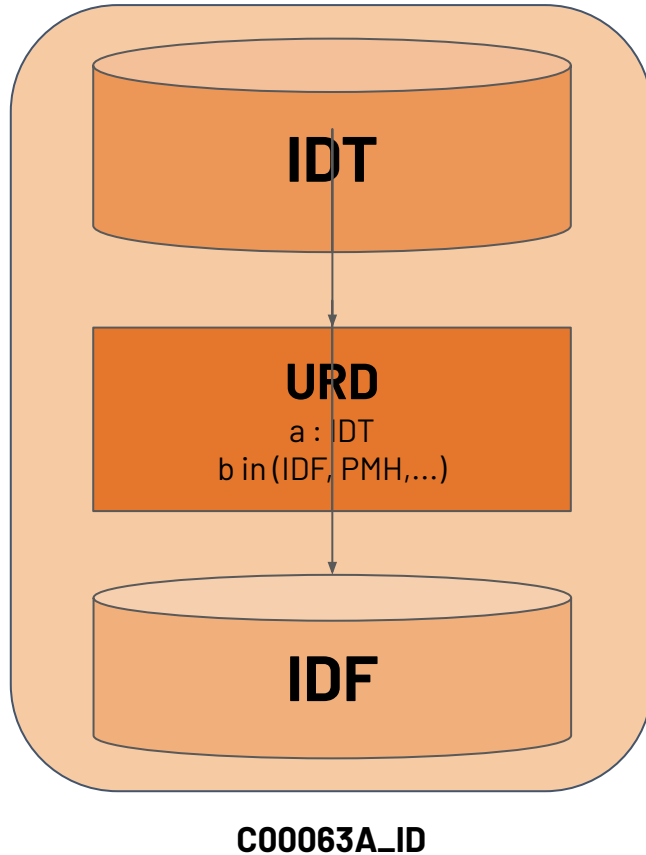
# Objetivo

Integrar las distintas fuentes de información de Canarias que poseen la misma unidad de análisis, en este caso, los datos sobre personas, mediante procesos de diseño, normalización, transformación de campos y programación de enlaces entre los ficheros de microdatos estadísticos según procedimientos determinísticos y probabilísticos (Record Linkage) para identificar a cada unidad (IDT), sus diferentes versiones de escritura (IDF) y almacenar su participación en las distintas fuentes administrativas (DAT) en una tabla de relaciones (URD).

# Metodología: Estructura del Registro de Población



# Metodología: Componentes del Registro de Población



**IDT:** Tabla de identificación única de una persona en la realización del registro

---

**IDF:** Tabla de alimentación del registro a partir de la localización de personas en diferentes microdatos en crudo (RD)

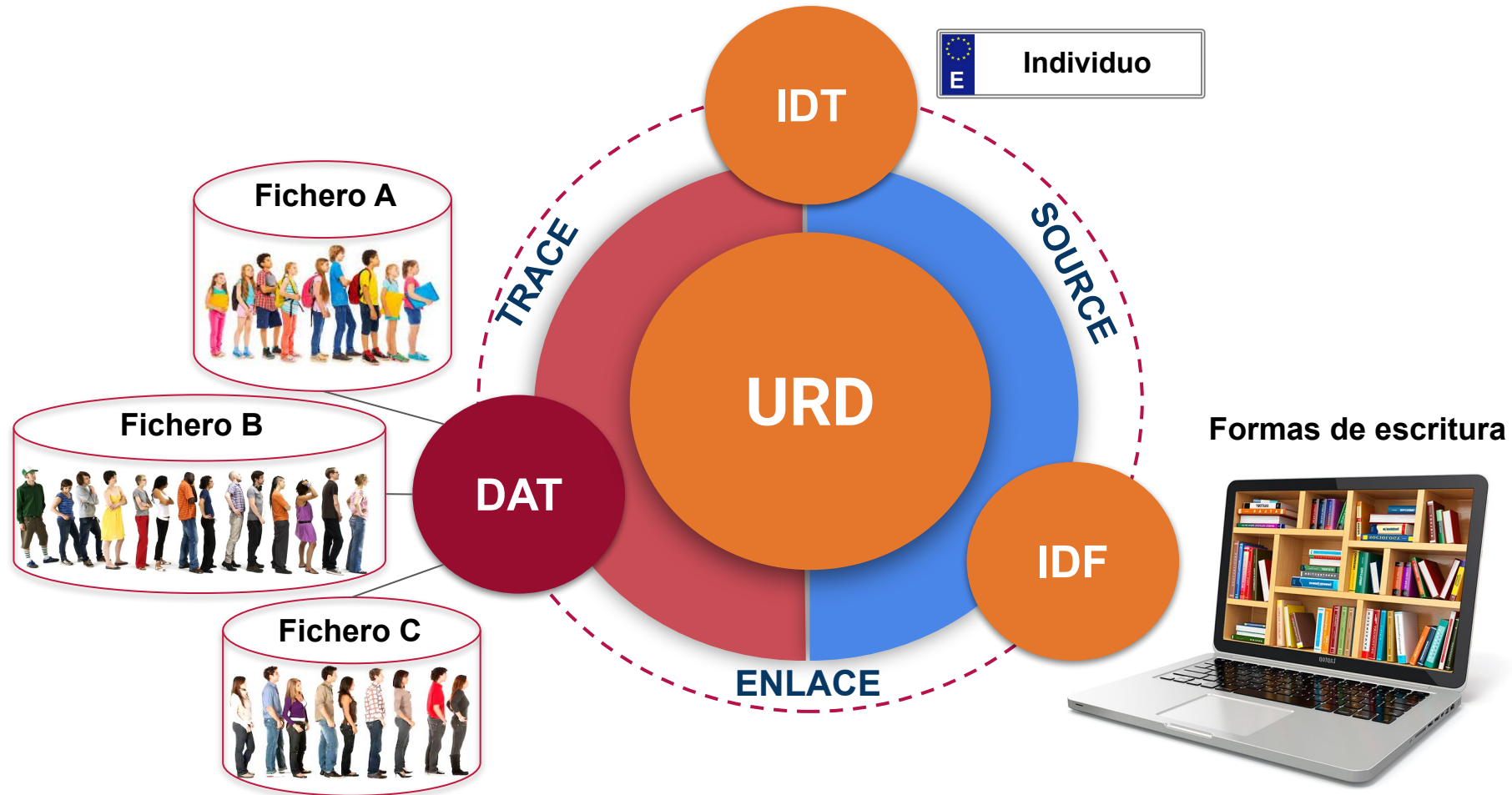
---

**URD:** Tabla de almacenamiento de las relaciones, internas o externas, de las observaciones de una tabla de microdatos con otros microdatos.

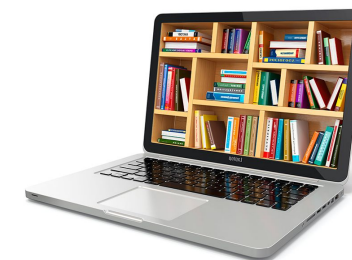
- I) Relaciones entre cada representante único con las diferentes versiones de cada persona
- 
- II) Relaciones entre cada representante único (de las versiones de cada persona) con otros ficheros



# Metodología: Relaciones IDT-IDF-DAT



# Metodología: Variables de la tabla IDF



El IDF tiene el propósito de integrar y almacenar fuentes de información heterogéneas.

Para para ello, se requieren campos normalizados y depurados que representen la **identidad personal**, que hace referencia al conjunto de características propias de una persona y permiten reconocerlo como un individuo diferente a los demás (IDT).

BLOCK	CONCEPT_ID	LABEL
CLAVE	UUID	Identificador Único Universal
	LUID	Identificador Único Local
	STID	Identificador de esquema y tabla
	MARCATIEMPO	Sello de tiempo de creación de la observación
IDENTIFICADORES	NOMEPER_TIPO_IPF	Tipo de documento
	NOMEPER_IDENTIFICADOR_IPF	Número de identificación de la persona física
	NOMEPER_IPF_LIDEN	Letra de control del IPF
	NOMEPER_IPF_LIDEN_RC	Letra de control del IPF (recalculada)
PERSONALES	NOMEPER_NOMBRE_APELLIDOS	Nombre y apellidos de la persona física
	NOMEPER_NOMBRE	Nombre completo de la persona física
	NOMEPER_NOMBRE_1	Primer nombre de la persona física
	NOMEPER_NOMBRE_2	Resto de nombres de la persona física
	NOMEPER_APELLIDOS	Apellidos de la persona física
	NOMEPER_PARTICULA_APELLIDO1	Partícula apellido 1
	NOMEPER_APELLIDO1	Primer apellido de la persona física
	NOMEPER_PARTICULA_APELLIDO2	Partícula apellido 2
DEMOGRÁFICAS	NOMEPER_APELLIDO2	Segundo apellido de la persona física
	NOMEPER_FNAC	Fecha nacimiento
	NOMEPER_SEXO	Sexo
	NOMEPER_CODMUNNAC	Municipio de nacimiento
	NOMEPER_PAISNAC	País de nacimiento

# Metodología: Trabajos previos

Descripción	Nombres y apellidos		
	Entorno de repositorio (IDF)	Entorno computacional	
		Función SOUNDEX (var1=* var2)	Función SOUNDEX_ES
Funciones de depuración: upcase, compress, substr, compbl, tranwrđ y scan	✓	✓	✓
Quitar tildes, diéresis, puntuación, espacios múltiples	✓	✓	✓
Quitar "DE", "DEL", "DE LA", "DE LOS", "DE LAS"		✓	✓
Sustituir "Ñ", "Ç"		✓	
Se contrae el nombre de "MARIA" por "M"		✓	✓

# Metodología: Tabla de relaciones URD

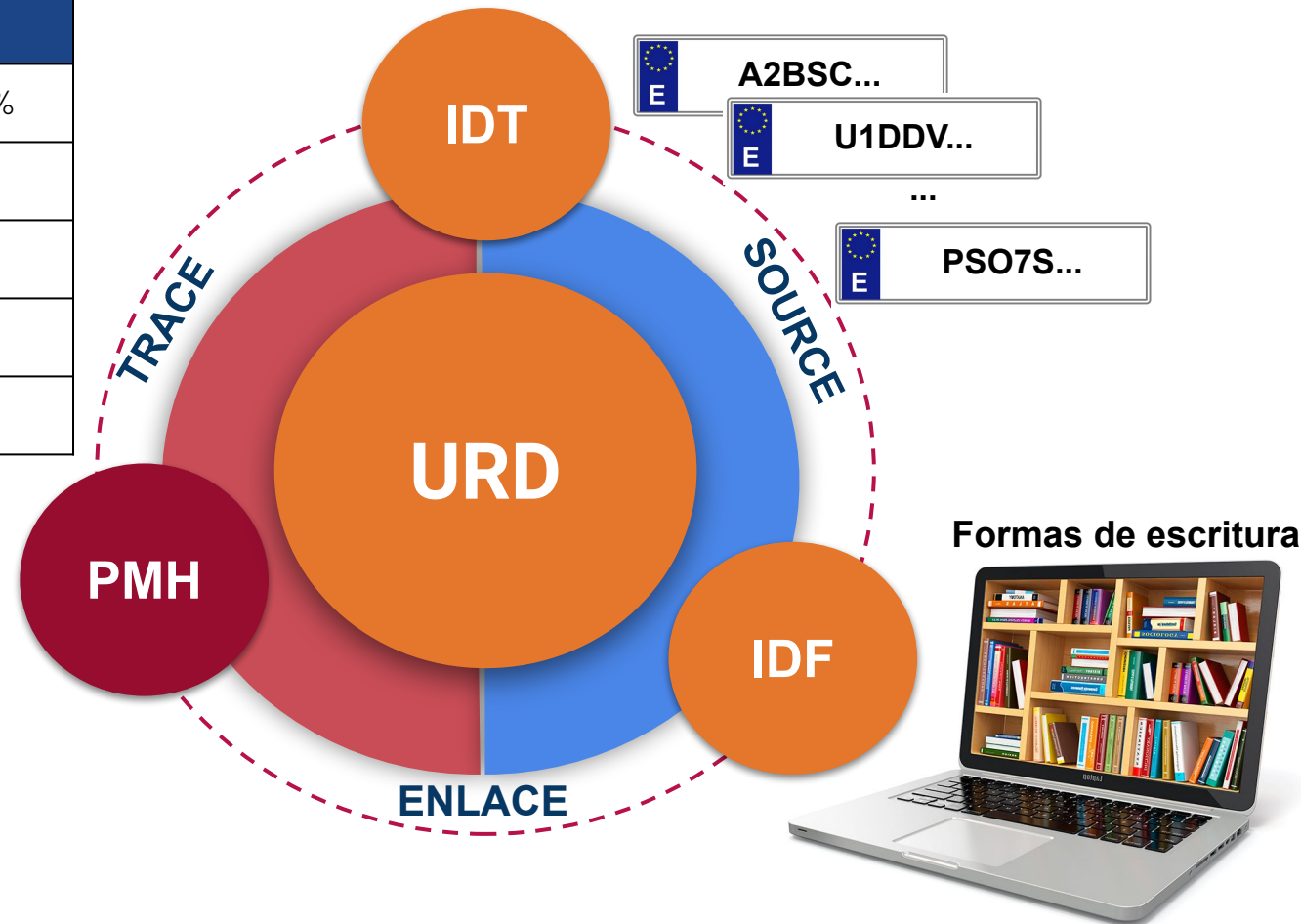
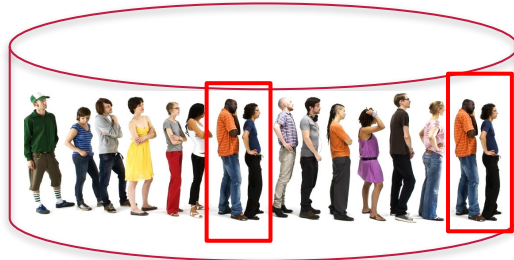
uuid	uuid_a	stid_a	uuid_b	stid_b	rel_description	similitud AND link_description
<b>A</b>	<b>X</b>	IDT Registro Población	<b>X1</b>	IDF Registro Población	DAT Padrón	
<b>B</b>	<b>X</b>	IDT Registro Población	<b>X2</b>	IDF Registro Población	DAT Padrón	
<b>C</b>	<b>X</b>	IDT Registro Población	<b>X1'</b>	DAT Padrón		$(X1' = X1) \text{ AND } A(x:x1)$
<b>D</b>	<b>X</b>	IDT Registro Población	<b>X2'</b>	DAT Padrón		$(X2' = X2) \text{ AND } B(x:x2)$
<b>E</b>	<b>X</b>	IDT Registro Población	<b>X3</b>	IDF Registro Población	DAT Afiliados	
<b>F</b>	<b>X</b>	IDT Registro Población	<b>X3''</b>	DAT Afiliados		$(X3'' = X3) \text{ AND } E(x:x3)$
<b>G</b>	<b>X</b>	IDT Registro Población	<b>X1''</b>	DAT Afiliados		$(X1'' \cong X1) \text{ AND } A(x:x1)$
<b>H</b>	<b>X</b>	IDT Registro Población	<b>X2'''</b>	DAT Demandantes		$(X2''' \cong X2) \text{ AND } B(x:x2)$
<b>I</b>	<b>X</b>	IDT Registro Población	<b>X1''''</b>	DAT Muface		$(X1'''' \cong X1) \text{ AND } A(x:x1)$

# Metodología: Primer fichero - Padrón Municipal de Hab. Enero 2017

Fase	Registros	Obs	% Obs
1	Únicos	2.174.703	99,38%
2	Únicos	12.363	0,56%
	Duplicados	1.244	0,06%
	Triplicados	2	0,00%
<b>Total</b>		<b>2.188.312</b>	<b>100%</b>

**DEDUPLICADO**

**PMH**

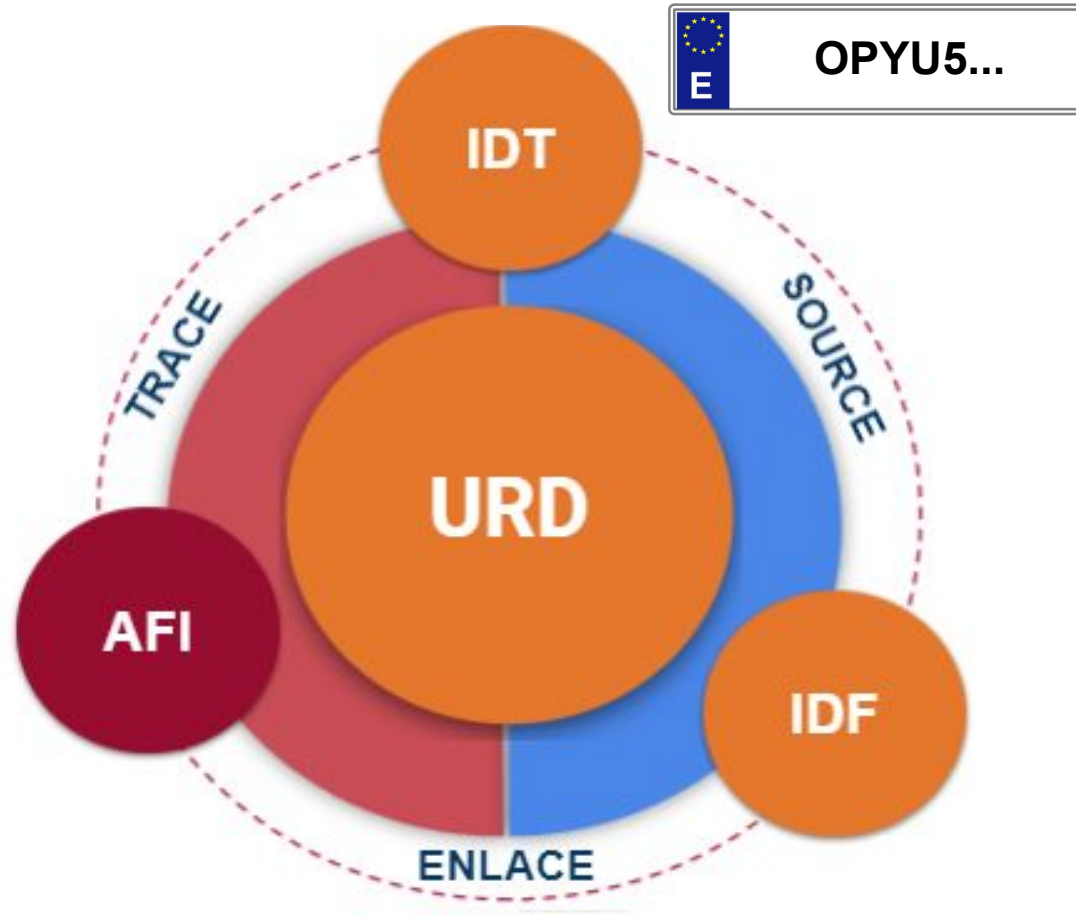


Formas de escritura

# Metodología: Resto de fuentes - Enlaces por igualdad



**MARÍA CONCEPCIÓN**



**Formas de escritura  
previas**



**MARÍA CONCEPCIÓN**



# Metodología: Resto de fuentes - Enlaces por igualdad

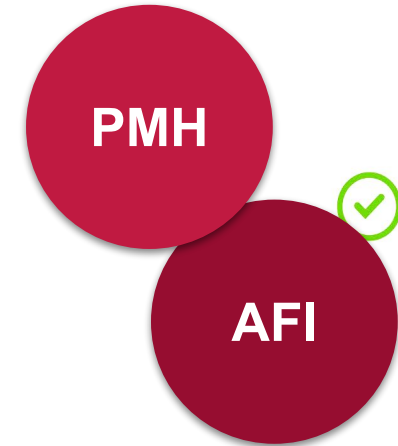


✗ Genera nuevo IDT



IDF1: MARÍA CONCEPCIÓN

✗ Genera nuevo IDF



✓ Genera relación DAT

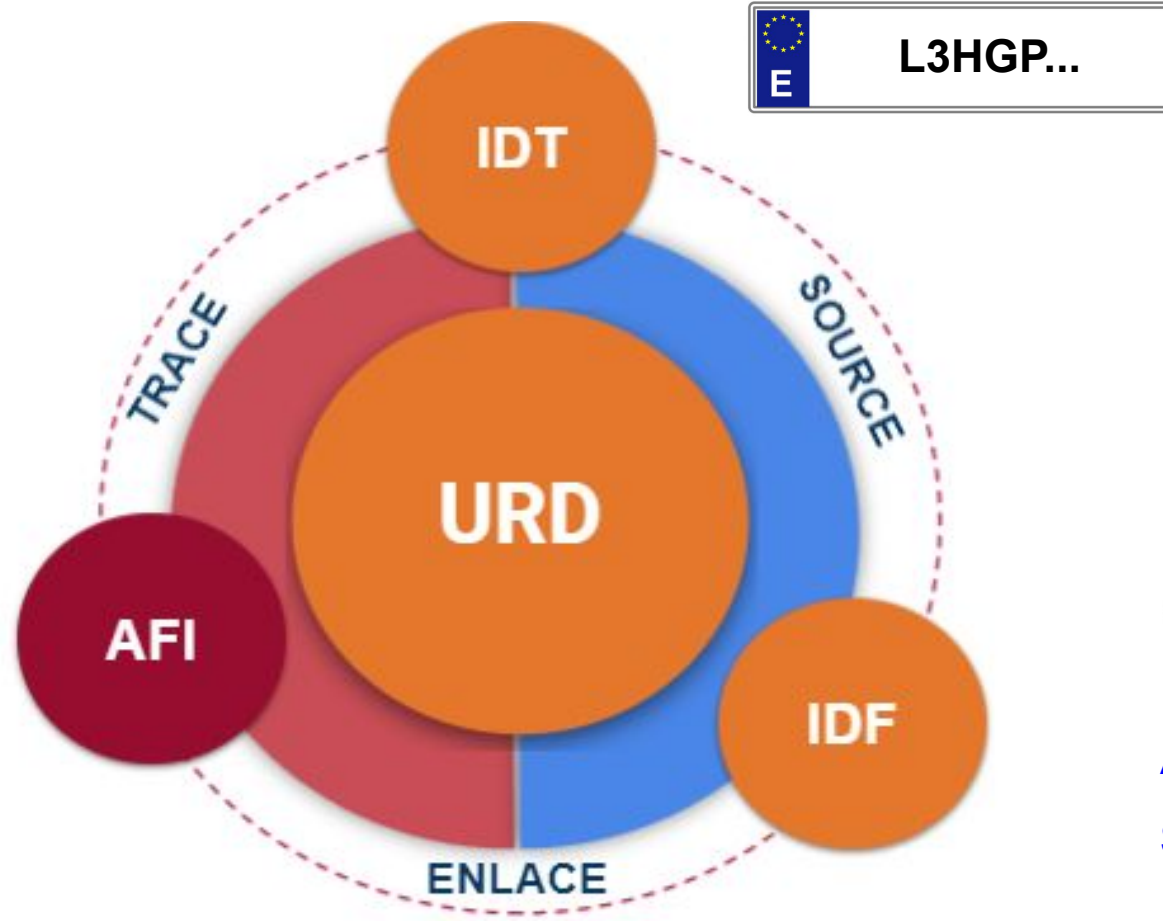


# Metodología: Resto de fuentes - Enlaces por similitud



**ANDREA SCARIONNI**

**SEXO: 1**



**ANDREA SCARIONI**

**SEXO: 2**





# Metodología: Resto de fuentes - Enlaces por similitud

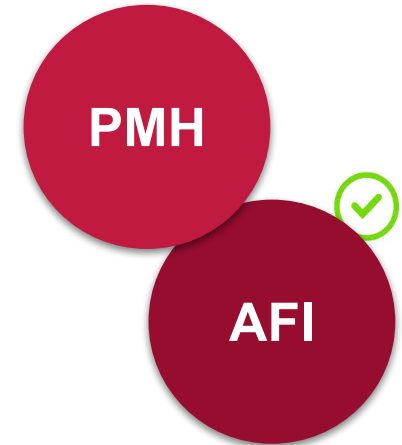


✗ Genera nuevo IDT



IDF1 : ANDREA SCARIONI; SEXO=2  
IDF2 : ANDREA SCARIONNI; SEXO=1

✓ Genera nuevo IDF

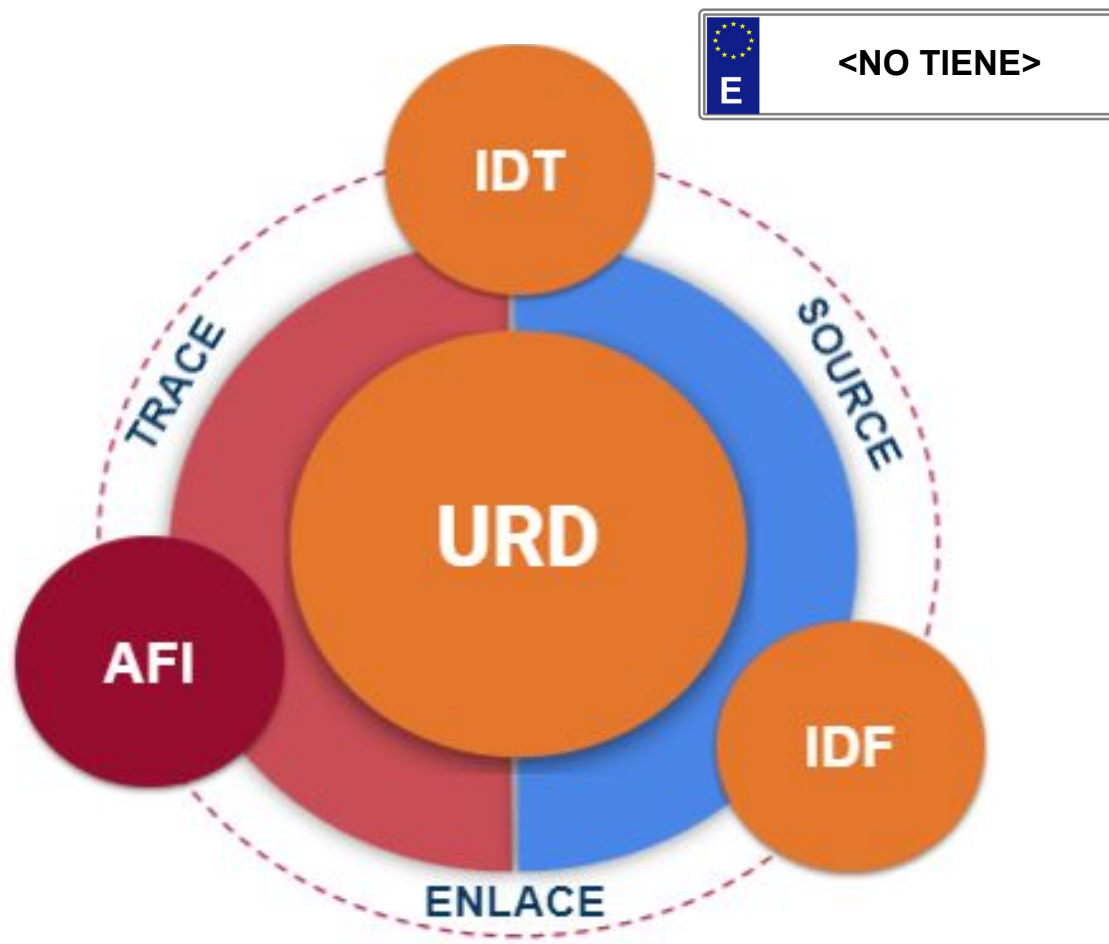


✓ Genera relación DAT

# Metodología: Resto de fuentes - Sin enlaces



THIAGO LOPES



Formas de escritura  
previas



<NO APARECE>

# Metodología: Resto de fuentes - Sin enlaces



IDT



✓ Genera nuevo IDT



IDF

IDF1 : THIAGO LOPES



✓ Genera nuevo IDF



DEM

✓ Genera relación DAT

# Metodología



Genera nuevo IDT



Genera nuevo IDF



Genera relación DAT

Enlaces por igualdad



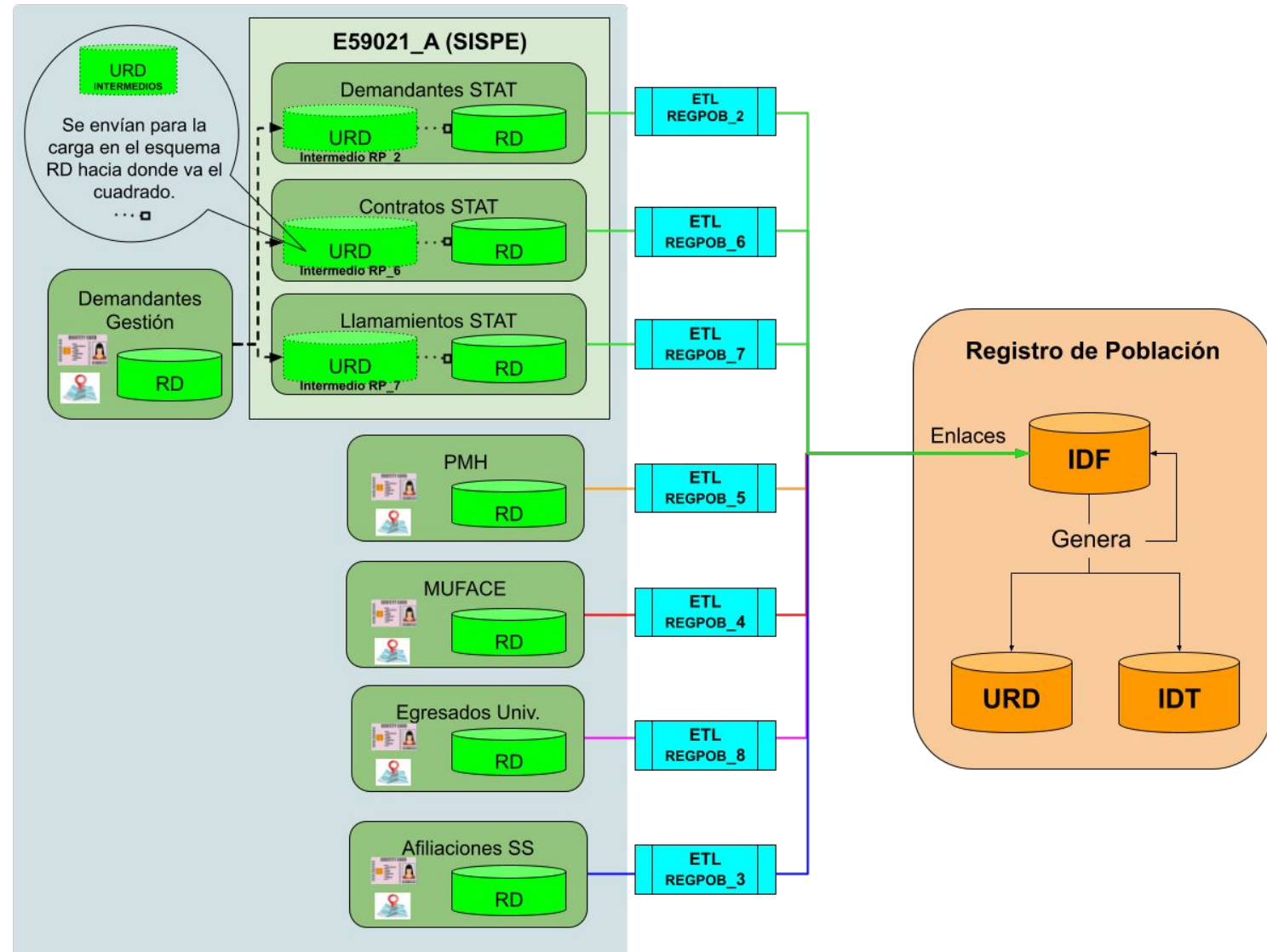
Enlaces por similitud



Sin enlaces



# Metodología: Procesos ETL



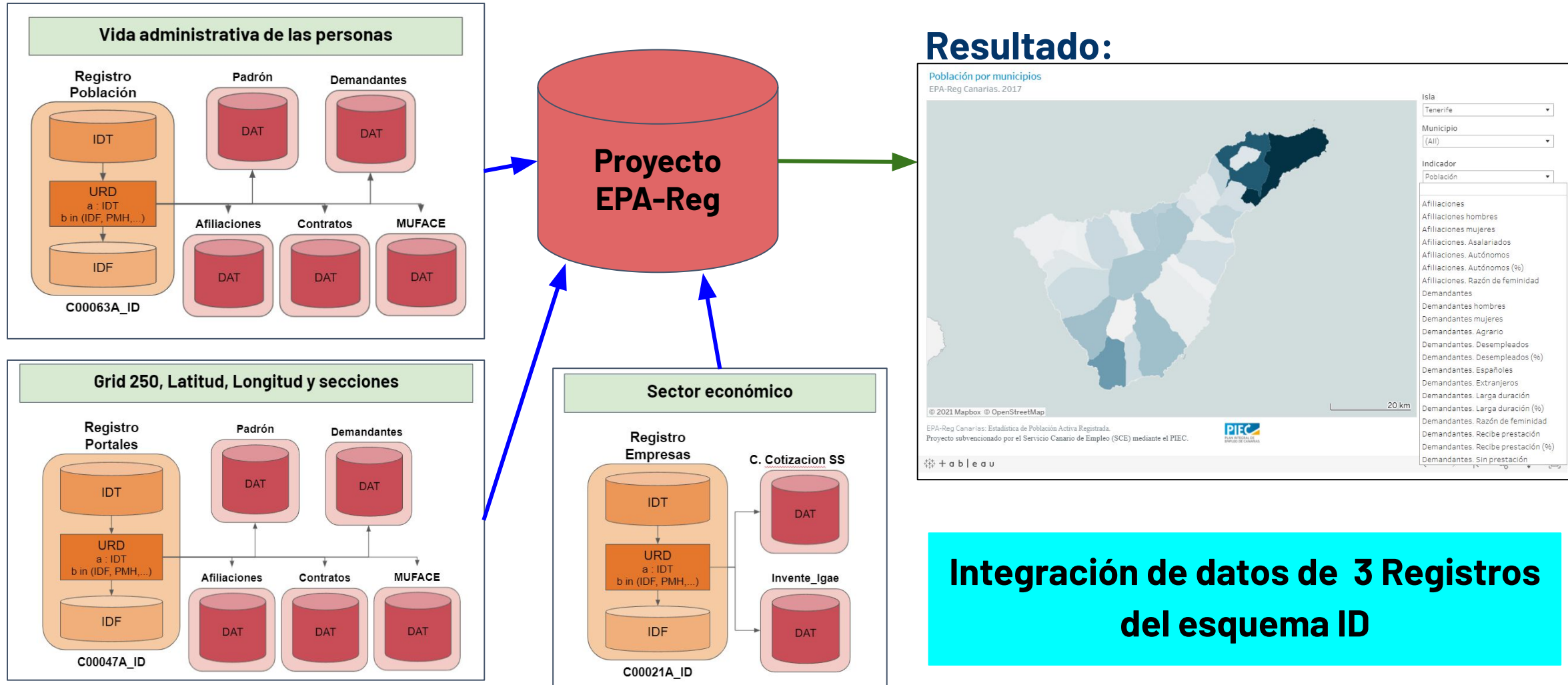
# Resultados: Serie integrada desde 2003-2021(sept)

Fuente	Ficheros	Obs	Nro. IDF	Nro. IDT	Nro. URD
Padrón municipal de hab.	31	67.031.646	6.397.617	3.133.164	73.429.263
Afiliaciones a la SS	54	40.073.268	487.120	102.946	40.560.388
Contratos	129	7.546.002	65.648	49.694	7.611.650
Demandantes	129	45.194.052	371.086	47.266	45.565.138
Afiliaciones a MUFACE	43	1.706.324	15.568	1.416	1.721.892
Egresados universitarios	4	27.125	4.975	242	32.100
Llamamientos	129	154.138	128	86	154.266
<b>Total general</b>	<b>519</b>	<b>161.732.555</b>	<b>7.342.142</b>	<b>3.334.814</b>	<b>169.074.697</b>

Durante los años 2003 y 2021 se han encontrado 3.3 millones de personas distintas en los registros administrativos antes descritos de Canarias (IDT), que se identifican a través 7.3 millones de versiones (IDF) y las relaciones se han almacenado en una tabla URD con más 169 millones de observaciones.



# Casos de uso: Proyecto EPA-Reg



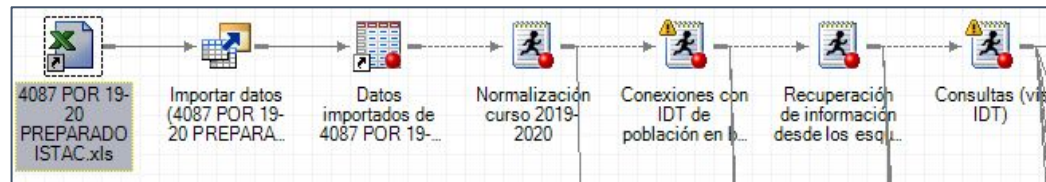
**Integración de datos de 3 Registros del esquema ID**

# Casos de uso: Vida administrativa de alumnado de F.P.

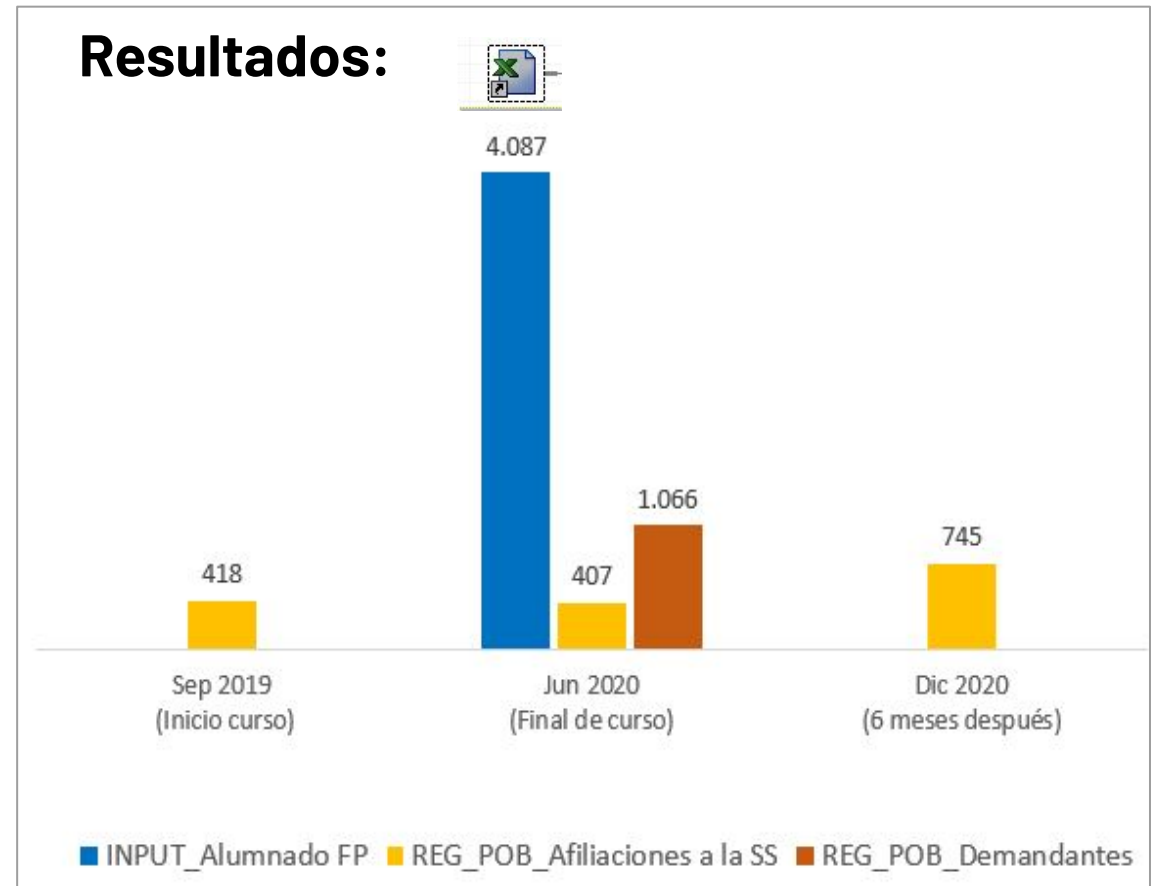
**Consulta:** Inserción laboral de alumnado de F.P. del curso 2019-2020 favorecidos por los fondos sociales europeos de financiación

**Metodología:** Partiendo de un fichero de entrada no anonimizado, se rastrea la presencia de las personas en los ficheros de Afiliaciones a la SS y en Demandantes de empleo para ciertos periodos

## ETL:



## Resultados:





# Casos de uso: Próximamente

- **Marco de población presente:** Con el fin de apoyar otras operaciones estadísticas y haciendo uso del URD se puede construir un marco para Canarias por periodo.
- **Apoyo a encuestas:** Aprovechar el Registro de Población para la validación e imputación de variables así como para aligerar información de los cuestionarios ya que podrían obtenerse haciendo uso del URD y consultando los registros administrativos necesarios.
- **Análisis longitudinal:** Para la investigación de acontecimientos de cohortes.
- **Otras consultas internas / externas.**

# Conclusiones

1. El Registro de Población permite conocer la vida administrativa de las personas y su participación en los distintos ficheros integrados en la tabla de relaciones URD
1. Permite introducir buenas prácticas que hagan más eficiente la producción de estadísticas del ISTAC.
1. La infraestructura del Registro de Población está en constante aprendizaje cada vez que se integra una fuente administrativa ya que aporta nuevas versiones del individuo y a medida que se integren más ficheros o fuentes, el aprendizaje de versiones (IDF) irá decreciendo.