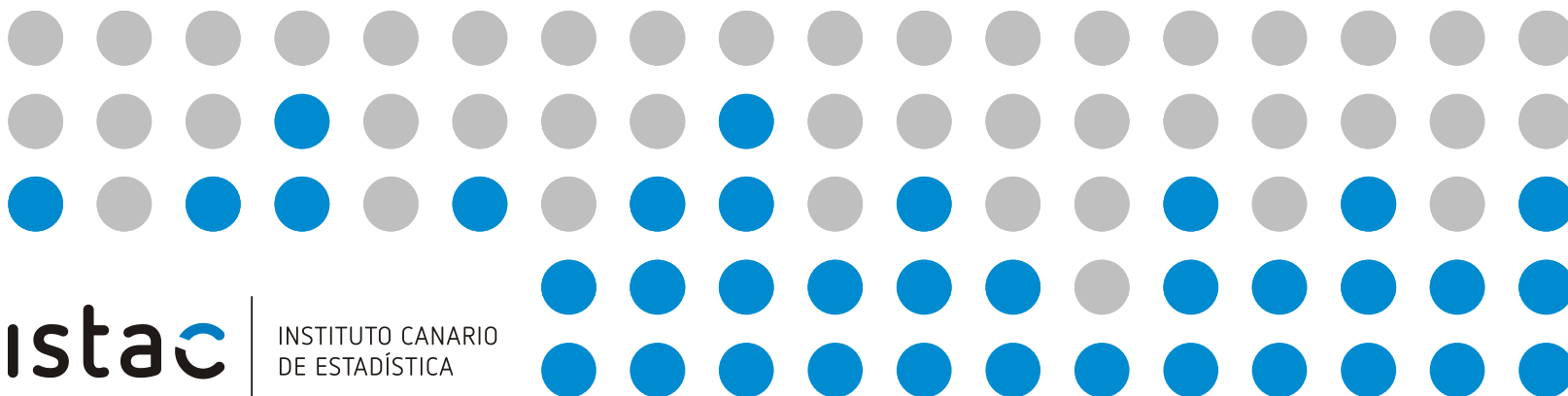


RETOS DE LA ESTADÍSTICA

# TRATAMIENTO DE DATOS

## Capacidades y legislación



**istac**

INSTITUTO CANARIO  
DE ESTADÍSTICA

RETOS DE LA ESTADÍSTICA

# TRATAMIENTO DE DATOS

## Capacidades y legislación

**Jornada sobre retos de la estadística y cartografía pública**

Instituto de Estadística y Cartografía de Andalucía  
11 de diciembre de 2019

**Alberto González Yanes**

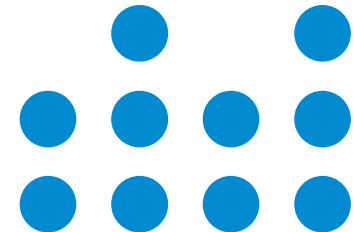
Jefe de Estadísticas Económicas

[jgonyanp@gobiernodecanarias.org](mailto:jgonyanp@gobiernodecanarias.org)

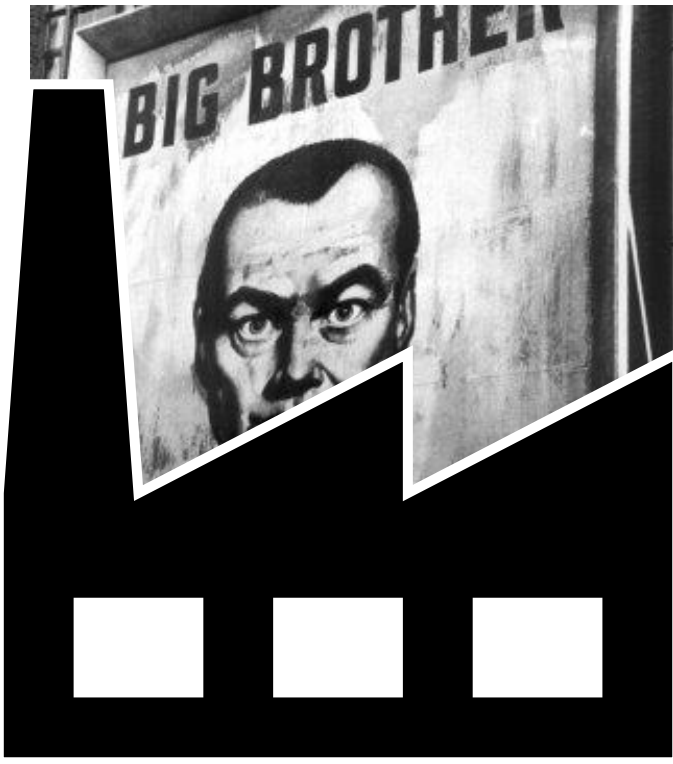
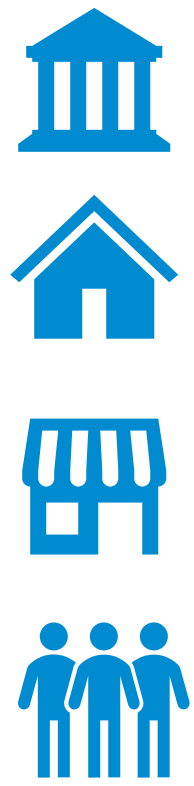
@algoya\_dat

@istac\_es

**istac** | INSTITUTO CANARIO  
DE ESTADÍSTICA



# Qué es una oficina pública de estadística



ENCUESTA  
REGISTROS

INDUSTRIALIZACIÓN  
INVESTIGACIÓN

INDEPENDENCIA  
INNOVACIÓN

MACRODATO  
MICRODATO





# Reto #1

# Nuevas datificaciones

#SMART\_CITIES

#INTERNET\_OF\_THINGS

#SOCIAL\_DATA

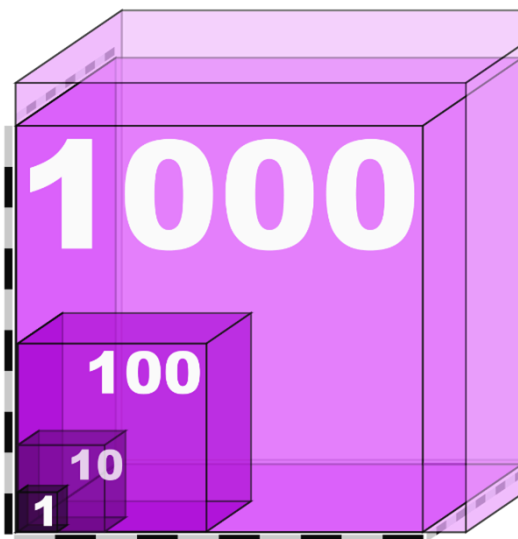
DATIFICACIÓN

#LINKED\_DATA #DATA\_SCIENTIST

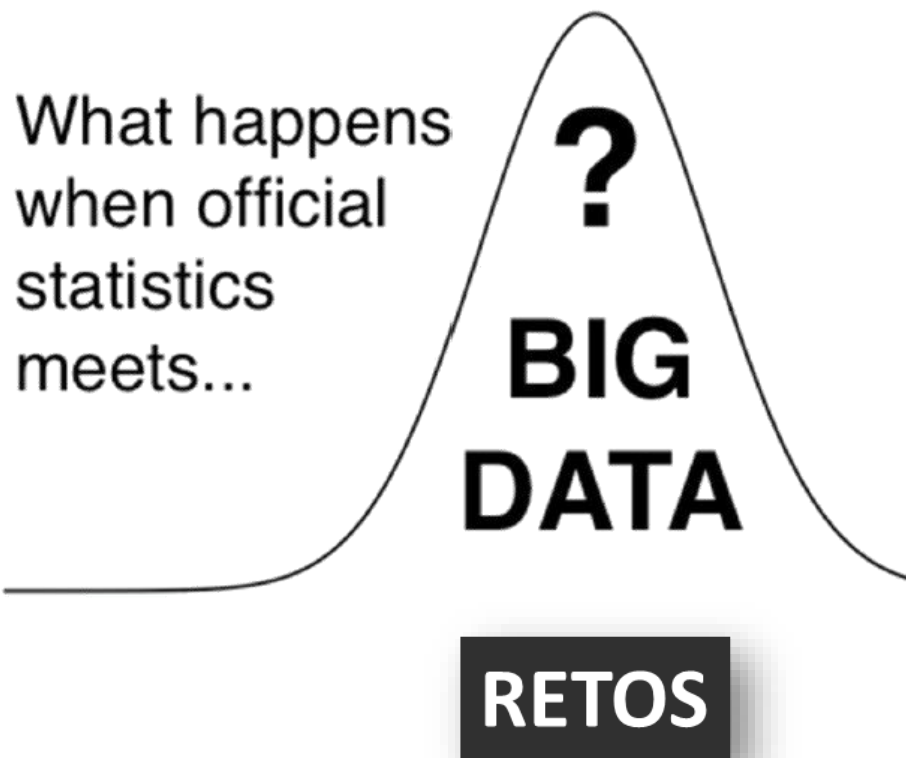
#OPEN\_DATA #BIG\_DATA

#DATA\_DRIVEN\_JOURNALISM

#DATA\_VISUALIZATION

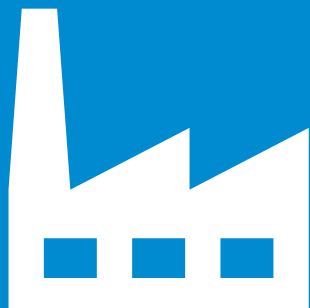


Las características de las fuentes Big Data vienen determinadas por las características clásicas con las que éstas son conocidas (3Vs): Volumen, velocidad y variedad. **Estas características se pueden estudiar desde una perspectiva tecnológica pero también desde un acercamiento estadístico.**



Estrategia Acceso Privacidad Calidad  
Metodología Tecnología Habilidades





# Reto #2

## Industrialización



## De la artesanía

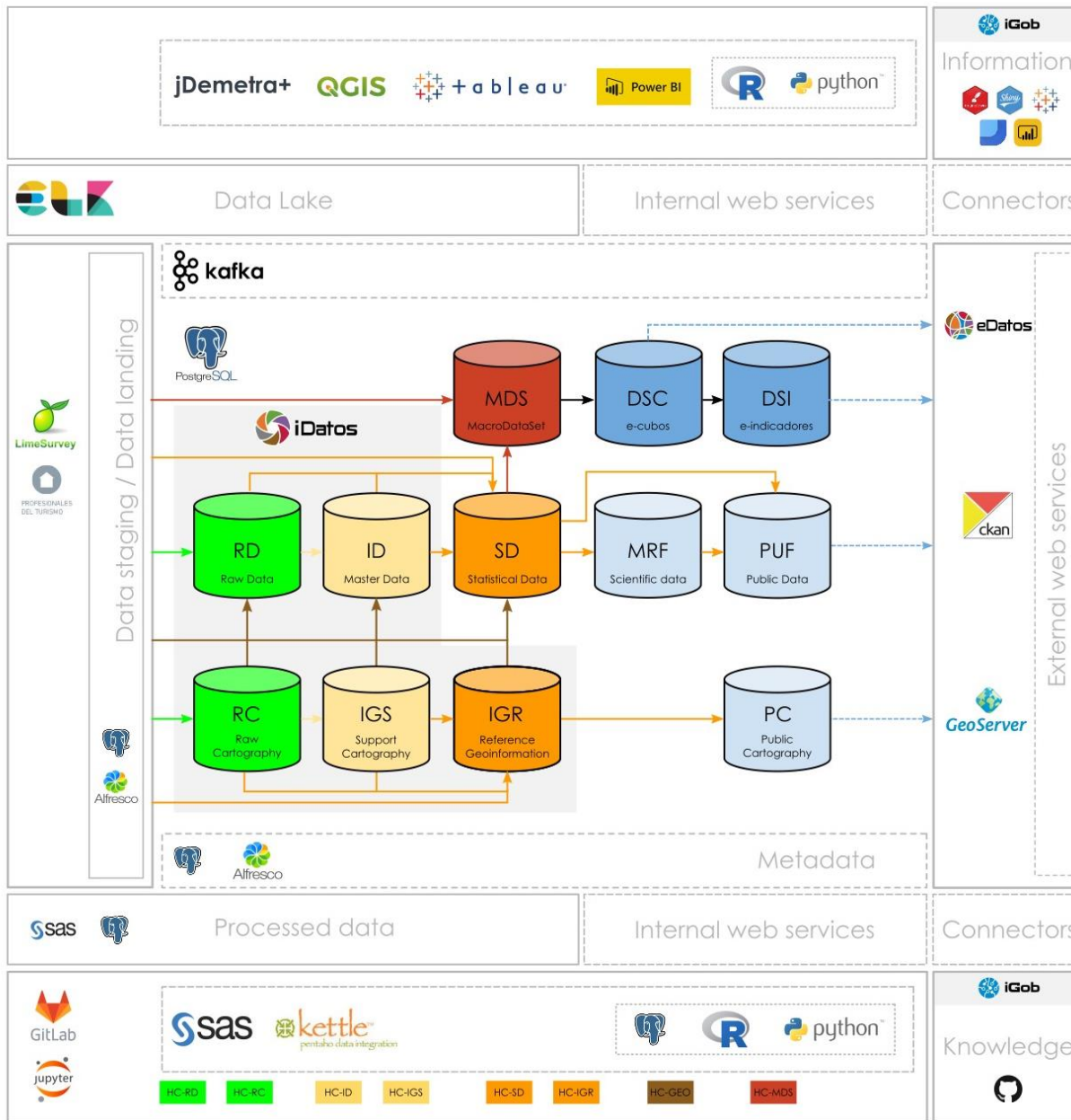


## A la industria





#1 Sistematizando



## 1. RECOGER / OBTENER DATOS (GSBPM-4)

## 2. PROCESAR (GSBPM-5)

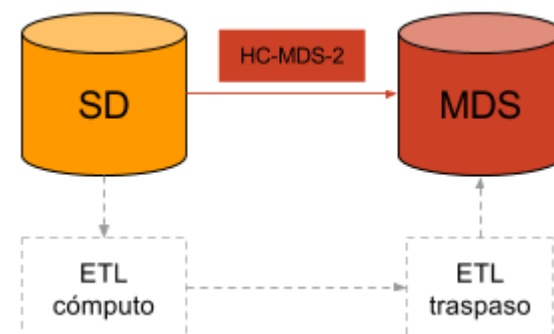
1. Codificación
2. Normalización de códigos
3. Integración
4. Generación de subtotales derivados
5. Generación de estimadores derivados para series temporales
6. Generación de estimadores corregidos de estacionalidad y calendario

## 3. ANALIZAR (GSBPM-6)

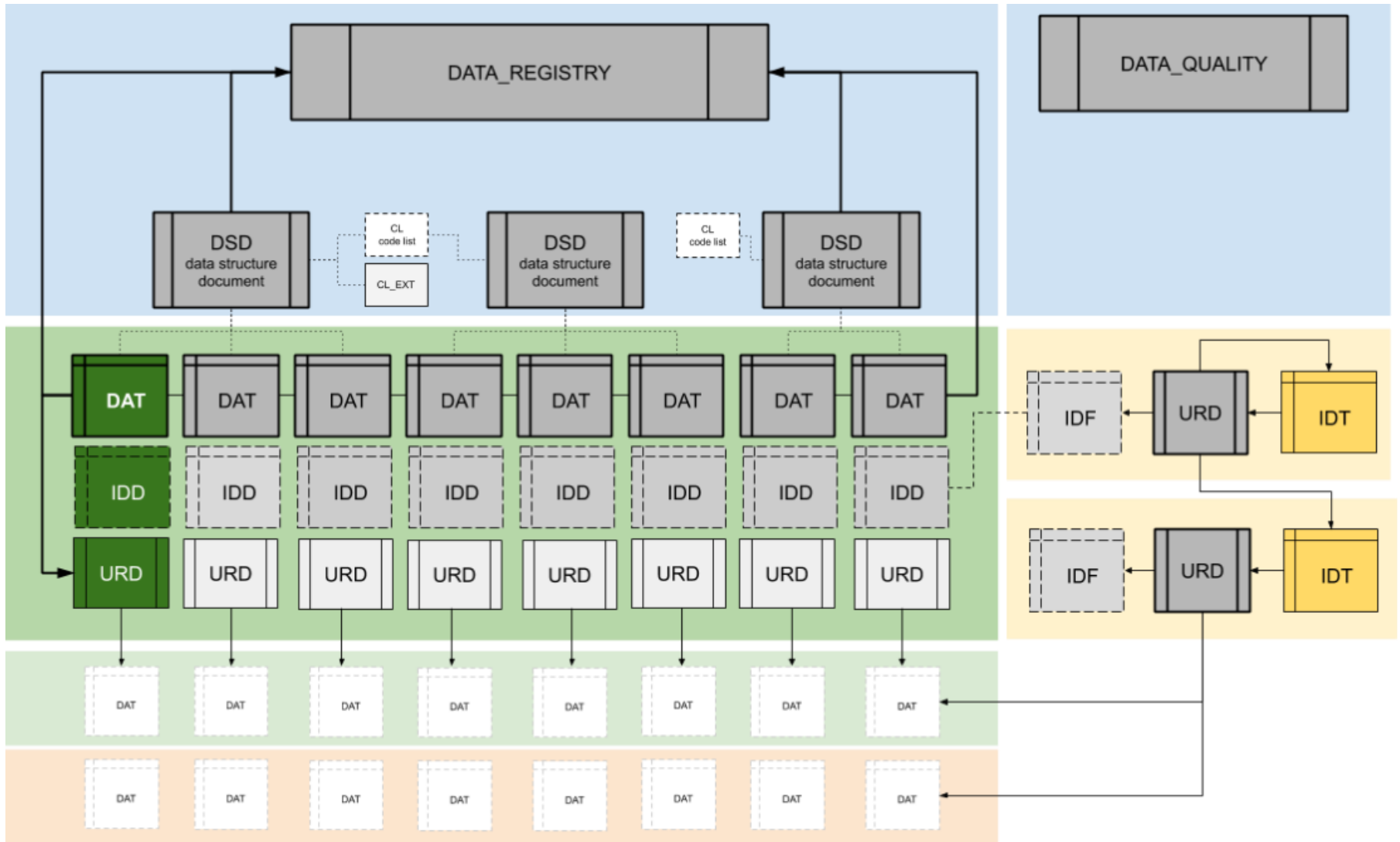
1. Generación de atributos de estado OBS\_STATUS
2. Control de coherencia entre totales y subtotales
3. Detección de outliers en series temporales
4. Generación de atributos de confidencialidad OBS\_CONFIDENCIALITY
5. Cargar en espacio repositorio (ETL de traspaso)
6. Cargar en espacios analíticos (hilo de transferencia)
7. Análisis exploratorio en espacios analíticos (fuera de hilo computacional)

## 4. DIFUNDIR (GSBPM-7)

1. Elaborar datasets cube (DSC)
2. Cargar en espacio repositorio los DSC
3. Cargar en espacios analíticos (hilo de transferencia)
4. Análisis exploratorio en espacios analíticos
5. Cargar en e-Cubos (hilo de transferencia)



# Organización conceptual de banco datos



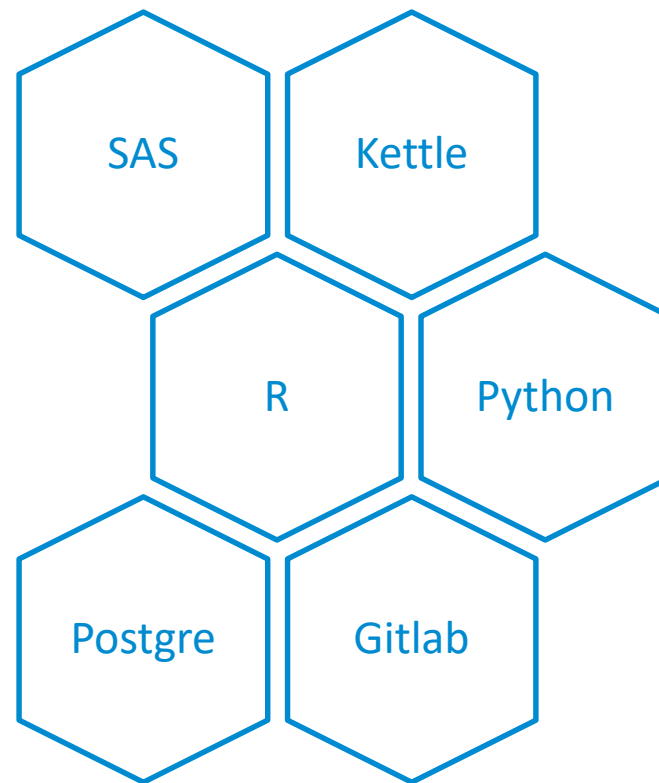
# AUTOMATION



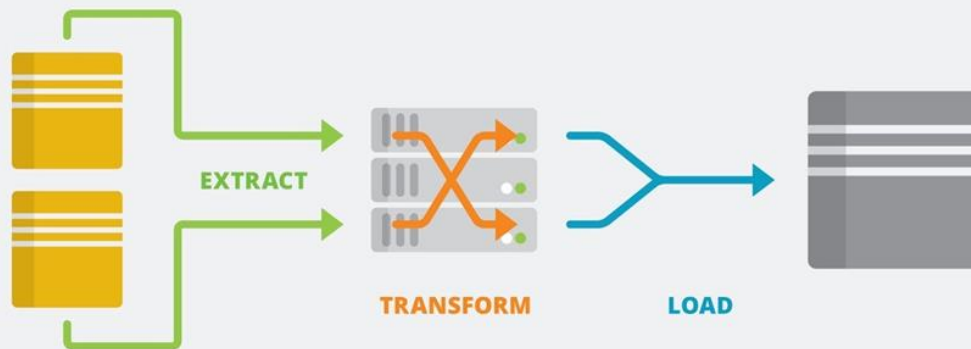
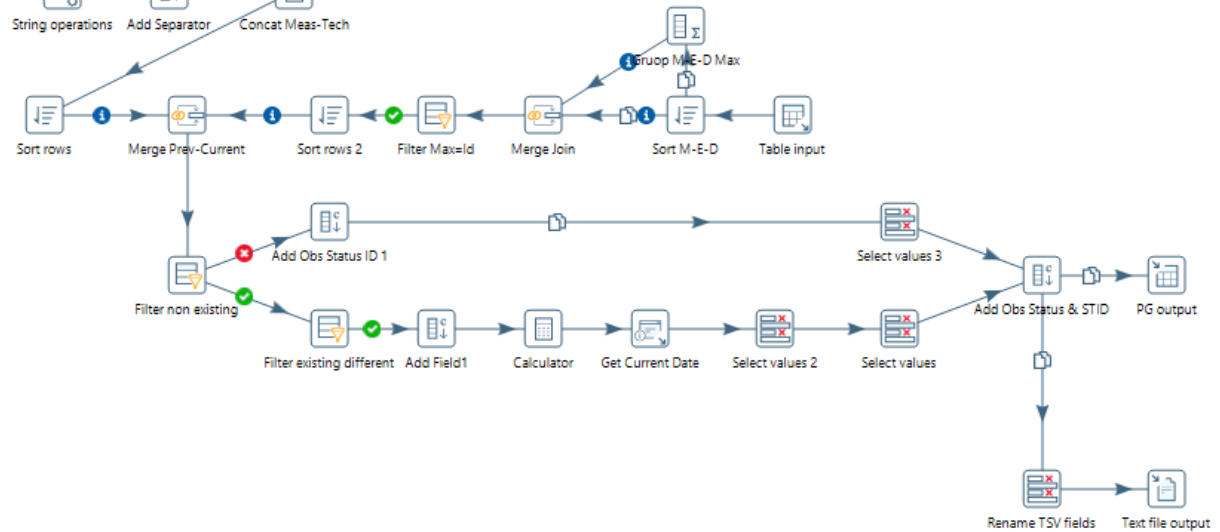
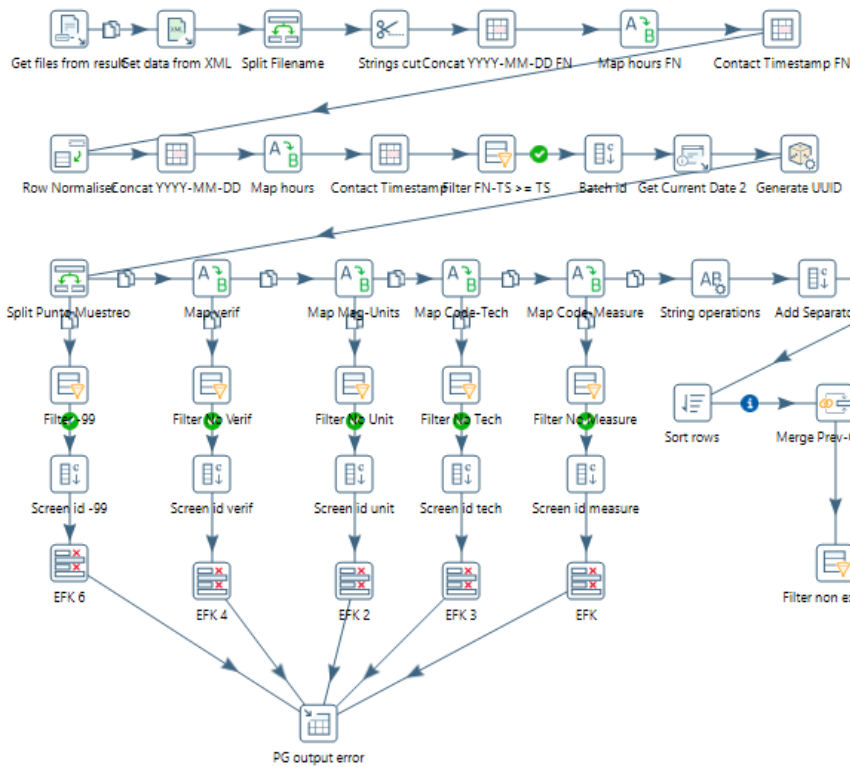
#2 Automatizando



# Organizar el cómputo



# Lanzamiento automático



## Registro de ETLs

[+ Crear nueva ETL](#)

Código

Nombre

Organismo responsable

Tipo

Incluir planificados

 Sí  No

Incluir eliminados

 Sí  No[✕ Reiniciar filtros](#)

CÓDIGO ↕

NOMBRE ▲

ORGANISMO RESPONSABLE ↕

T-ETL-CAH-EST-MEAN-TSV-001

Cálculo de la medida correspondiente

ISTAC

T-ETL-HIST-CA-009\_EXC\_TSV

Carga de datos históricos

ISTAC

J-ETL-CAH-EST-MEAN-001

Datos de configuración para el cálculo de medidas

ISTAC

J-ETL-CAH-EST-MEAN-TSV-001

Datos de configuración y cálculo de medidas

ISTAC

J-ETL-CAH-002

Descarga FTP para datos horarios

ISTAC

T-ETL-CAH-006

Transformación de datos horarios

ISTAC

1 - 6 de 6

Tamaño de página 20 ▼

« « 1 » »

# COETL

Aplicación web que permite gestionar el inventario de los procesos de extracción, validación, normalización, transformación y carga en eDatos



# QUALITY CONTROL



#3 Controlando



## Reto #3

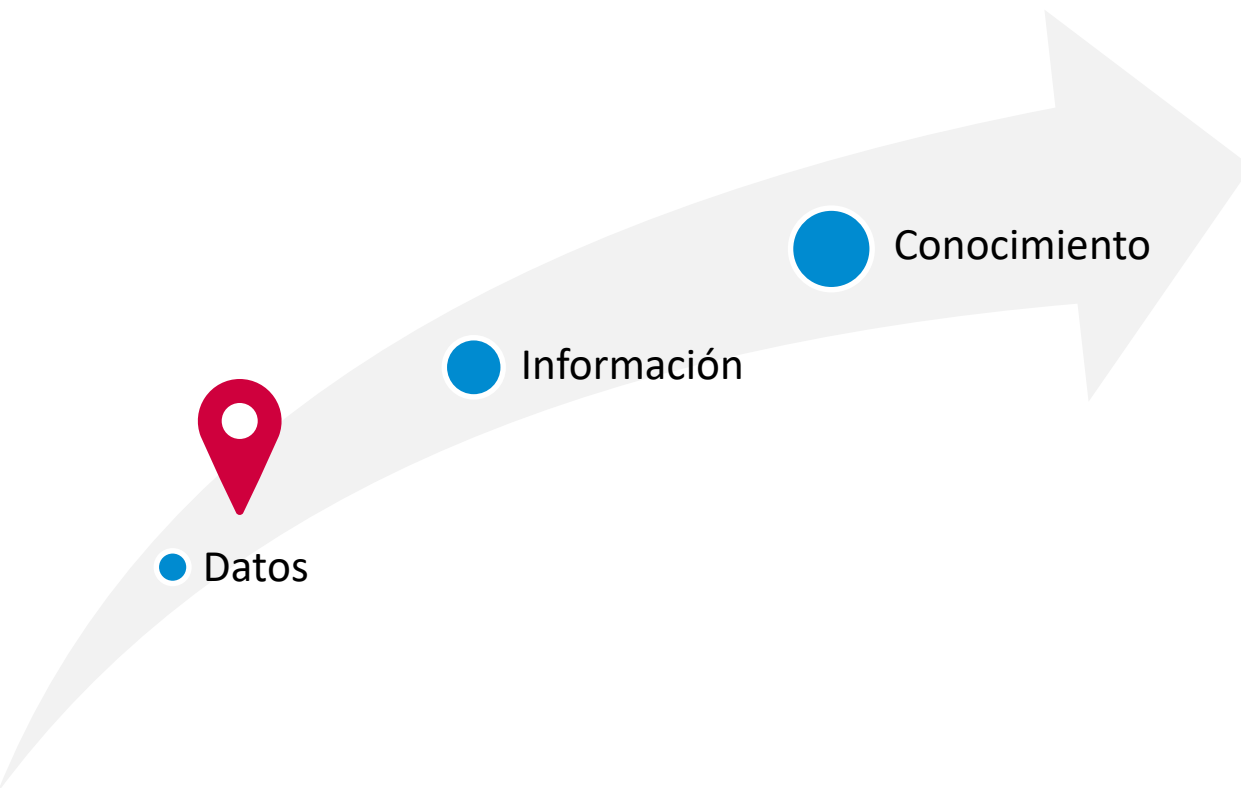
# Inteligencia de gobierno



**iGob**

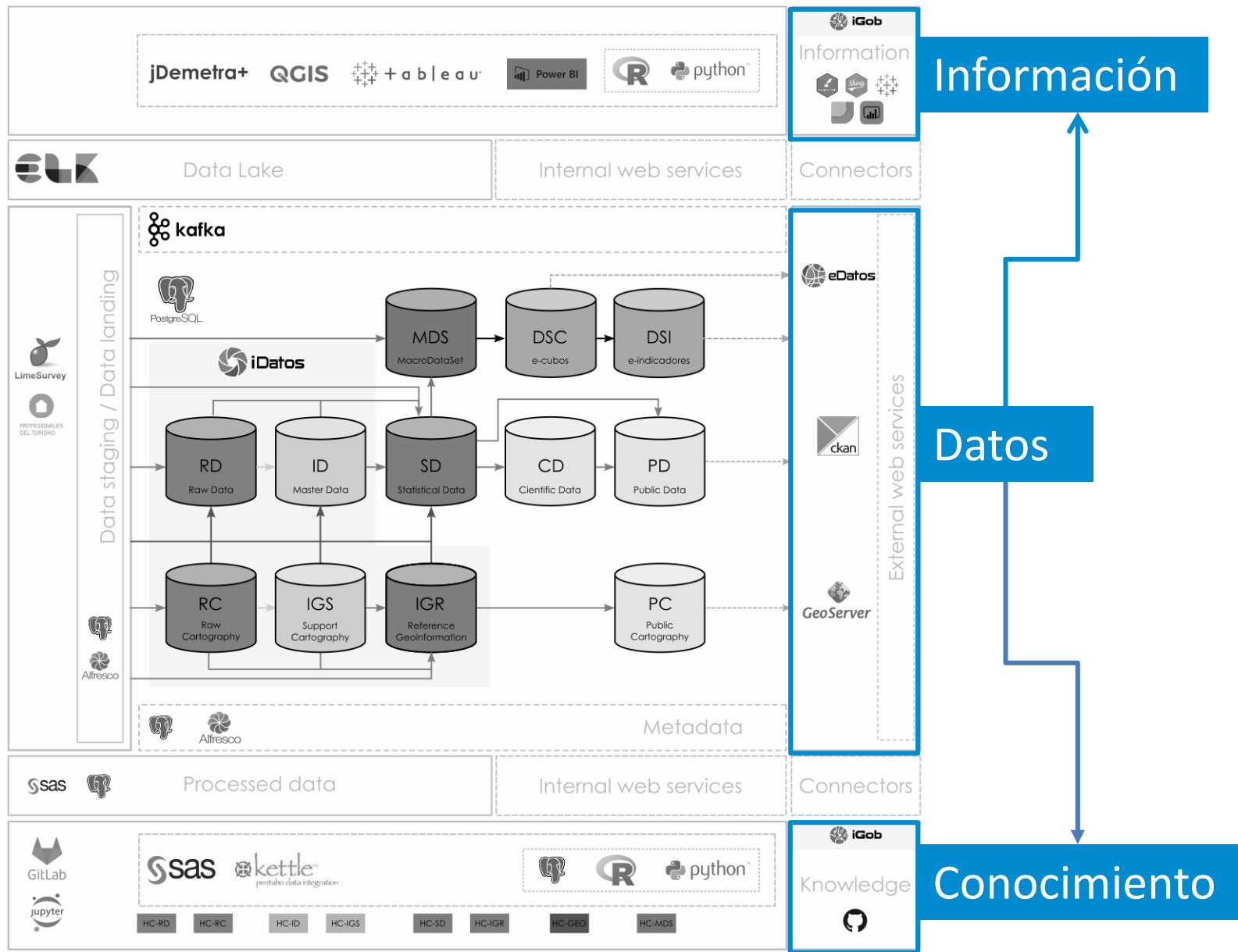
INTELIGENCIA  
DE GOBIERNO

# Apoyo a decisiones usando evidencias estadísticas

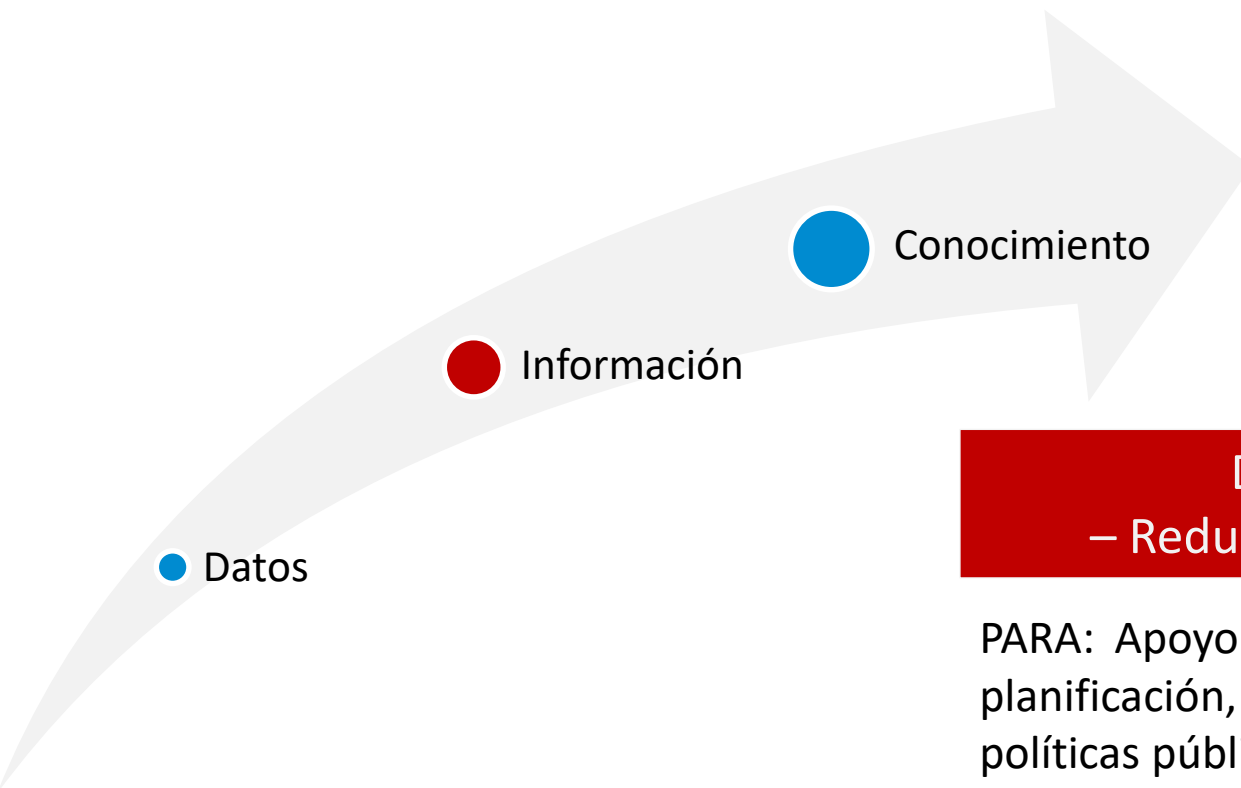


**iGob**

INTELIGENCIA  
DE GOBIERNO







**iGov**

INTELIGENCIA  
DE GOBIERNO

**Datos + Contexto**  
**– Reducción de incertidumbre**

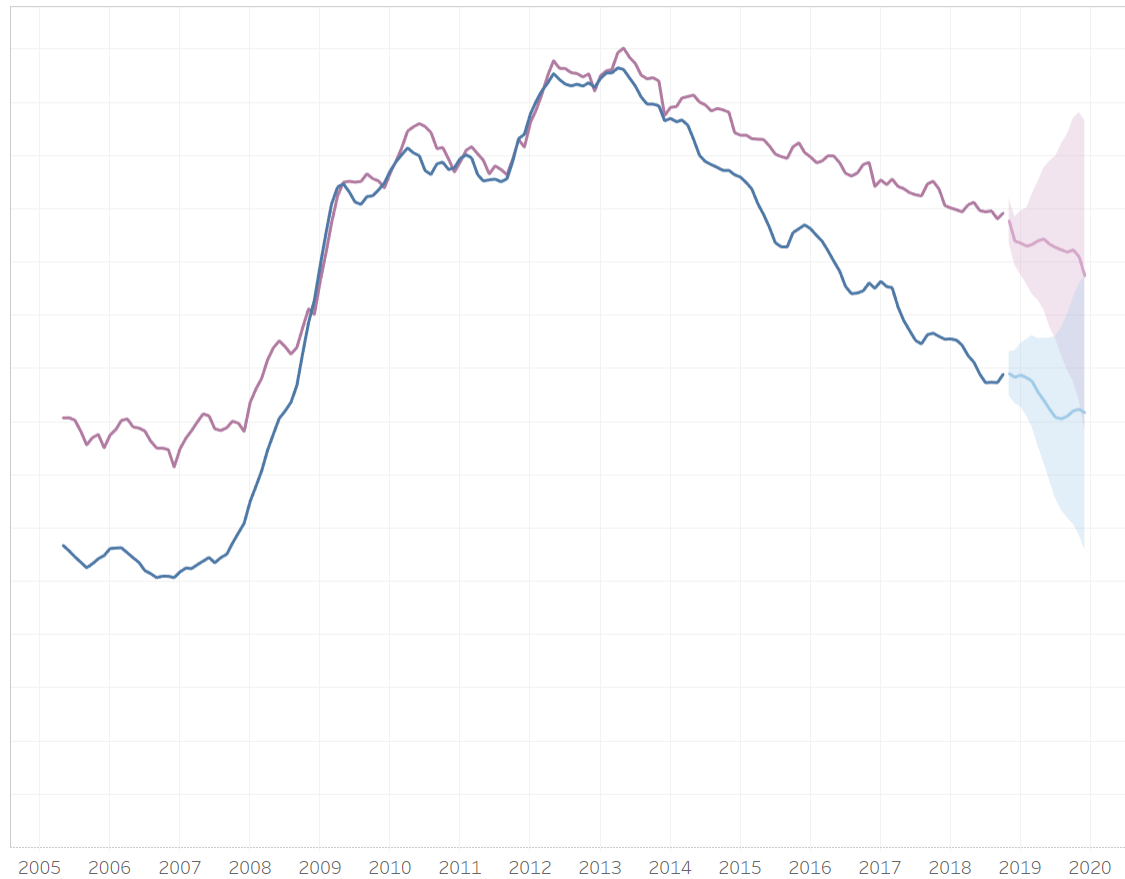
**PARA:** Apoyo a decisiones de alto nivel en planificación, seguimiento y evaluación de políticas públicas

**CON:**

1. Instrumentación numérica, visual, textual
2. Instrumentación metodológica

## INDICADORES DE GÉNERO DE CANARIAS - Paro registrado

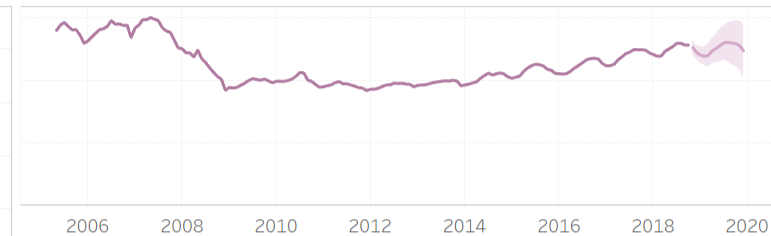
Pronóstico de paro registrado según sexo



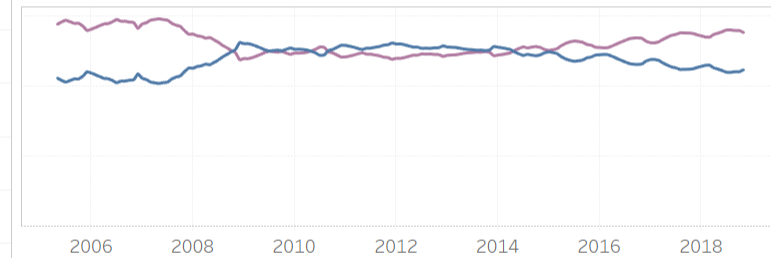
Municipios

(Todo)

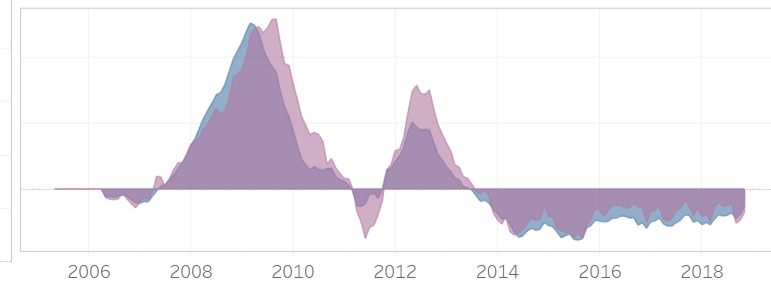
Ratio de feminidad



Porcentaje de paro registrado según sexos



Variación interanual de paro registrado según sexo



## **Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno.**

### **Artículo 8. Información económica, presupuestaria y estadística.**

1. Los sujetos incluidos en el ámbito de aplicación de este título deberán hacer pública, como mínimo, la información relativa a los actos de gestión administrativa con repercusión económica o presupuestaria que se indican a continuación:

(...)

- i) La **información estadística** necesaria para valorar el **grado de cumplimiento y calidad de los servicios públicos** que sean de su competencia, **en los términos que defina cada administración competente.**

● Datos

● Información

● Conocimiento



**iGov**

INTELIGENCIA  
DE GOBIERNO

**Datos + Algoritmos**  
– Reducción de incertidumbre

PARA: Apoyo a decisiones de alto nivel y bajo nivel (microdecisiones públicas)

CON:

1. Perfilados-segmentación
2. Modelos estadísticos
3. Inteligencia artificial
4. Simuladores de políticas públicas



<https://github.com/ISTAC>

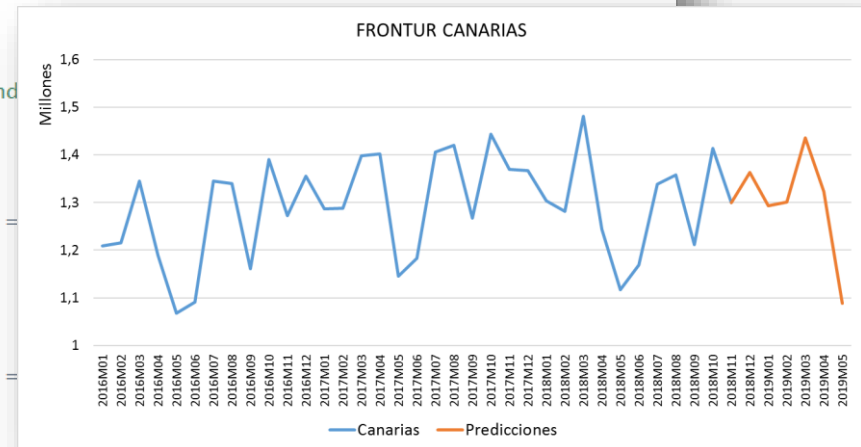
```

#' @title Get forecast
#' @description Executes the Basic Structural Model (BSM) with the inclusion of exogenous variables and seasonal
#' @param X Times series or set of time series to be forecast
#' @param Y Exogenous variables
#' @param n Number of available data
#' @return mylist List of forecast of a time series or set of time series and the value of MAPE for the estimations
#' @example
# get_forecast(myseries,myexogenous_va,n=24)

get_forecast <- function(X,Y,n){

  # CREATION OF OUTPUT VARIABLES
  # Variable for MAPE values
  MAPE <- matrix(NA, ncol = ncol(X), nrow = 1)
  # Variable for forecast values
  Prediction <- matrix(0, ncol = ncol(X), nrow = nrow(X))

  # Loop to execute BSM with the inclusion of exogenous variables and
  if (ncol(X) > 1)
  {for (i in 1:ncol(X))
  {
    ssKi <- SSMmodel(log(as.numeric(X[,i])) ~ SSMtrend(degree = 2, Q =
      SSMseasonal(12, Q = matrix(NA))+
      SSMregression(~Y, Q =matrix(0)),
      H = matrix(NA))
    fitSSKI <- fitSSM(ssKi, inits = c(0,0,0,0), method = "BFGS")
    out <- KFS(fitSSKI$model, filtering = "state")
    dfSSKSIStates <- ts(out$alphahat, start = time(X)[1], frequency =
    for (j in 1:ncol(Y))
    {
      Prediction[,i] <- Prediction[,i] + dfSSKSIStates[,j]*Y[,j]
    }
    Prediction[,i] <- Prediction[,i] + dfSSKSIStates[, ncol(Y) + 1] + dfSSKSIStates[, ncol(Y) + 3]
    Prediction[,i] <- exp(Prediction[,i])
  }
  }
  }
  
```



## Reglamento General de Protección de Datos

### Artículo 22.- Decisiones individuales automatizadas, incluida la elaboración de perfiles

A fin de garantizar un tratamiento leal y transparente respecto del interesado, teniendo en cuenta las circunstancias y contexto específicos en los que se tratan los datos personales, el responsable del tratamiento debe utilizar procedimientos matemáticos o estadísticos adecuados para la elaboración de perfiles, aplicar medidas técnicas y organizativas apropiadas para garantizar, en particular, que se corrigen los factores que introducen inexactitudes en los datos personales y se reduce al máximo el riesgo de error, asegurar los datos personales de forma que se tengan en cuenta los posibles riesgos para los intereses y derechos del interesado y se impidan, entre otras cosas, efectos discriminatorios en las personas físicas por motivos de raza u origen étnico, opiniones políticas, religión o creencias, afiliación sindical, condición genética o estado de salud u orientación sexual, o que den lugar a medidas que produzcan tal efecto. Las decisiones automatizadas y la elaboración de perfiles sobre la base de categorías particulares de datos personales únicamente deben permitirse en condiciones específicas.

## RGPD – Elaboración y evaluación de perfiles

Transparencia - Métodos estadísticos - Control del sesgo algorítmico



Garantizar un tratamiento leal y transparente respecto del interesado



Utilizar procedimientos matemáticos o estadísticos adecuados para la elaboración de perfiles



Impedir efectos discriminatorios en las personas

## ESTRATEGIA ESPAÑOLA DE I+D+I EN INTELIGENCIA ARTIFICIAL

### RECOMENDACIÓN 5: IA PARA EL USO DE DATOS DE LAS AAPP

Esta Estrategia debe servir para fomentar el establecimiento de una base para que la ingente cantidad de datos abiertos y públicos generados, y que no se limitan a la I+D+I, sean modelados, analizados, utilizados y explotados de forma segura y ética por la propia administración que los genera o por otras administraciones, con objeto de obtener el máximo partido de los datos abiertos utilizando técnicas de IA, mejorando la eficacia y la eficiencia de sus procesos, creando plataformas de colaboración pública-privada y reduciendo costes. Asimismo, se deben identificar los cauces para que las empresas privadas que utilizan datos abiertos puedan compartir parte de sus datos con la Administración pública con el objetivo de ofrecer mejores servicios públicos, garantizando el bienestar social y la seguridad individual y pública. Para conseguir minimizar los riesgos que supone un uso de datos abiertos del sector público sin procesos de certificación, se recomienda crear un **Instituto Nacional de Datos** que permita planificar y definir una gobernanza sobre datos procedentes de las AAPP nacionales, regionales y locales.



## ESTRATEGIA ESPAÑOLA DE I+D+I EN INTELIGENCIA ARTIFICIAL

### Instituto Nacional de Datos



Crear un **Instituto Nacional de Datos** que permita planificar y definir una gobernanza sobre datos procedentes de las AAPP nacionales, regionales y locales



Para que la ingente cantidad de datos públicos sean modelados, analizados, utilizados y explotados de forma segura y ética por la propia administración



Con objeto de obtener el máximo partido de los datos abiertos utilizando técnicas de IA, mejorando la eficacia y la eficiencia de sus procesos, creando plataformas de colaboración pública-privada y reduciendo costes



Facilitando que las empresas privadas que utilizan datos abiertos puedan compartir parte de sus datos con la Administración pública con el objetivo de ofrecer mejores servicios públicos

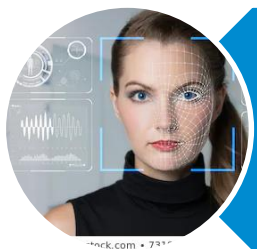


**Audit**

## Nuevas funciones de la estadística pública regional



Evaluación cuantitativa de políticas públicas



Elaboración y auditoría de algoritmos



Nodo de interoperabilidad de datos para fines analíticos

## Nuevas funciones de la estadística pública regional



Son órganos independientes



Tienen autoridad pública



Ejercen la competencia estadística



Tienen capacidades técnicas y metodológicas



Tienen capacidad jurídica para acceso, integración y tratamiento de datos administrativos dispersos

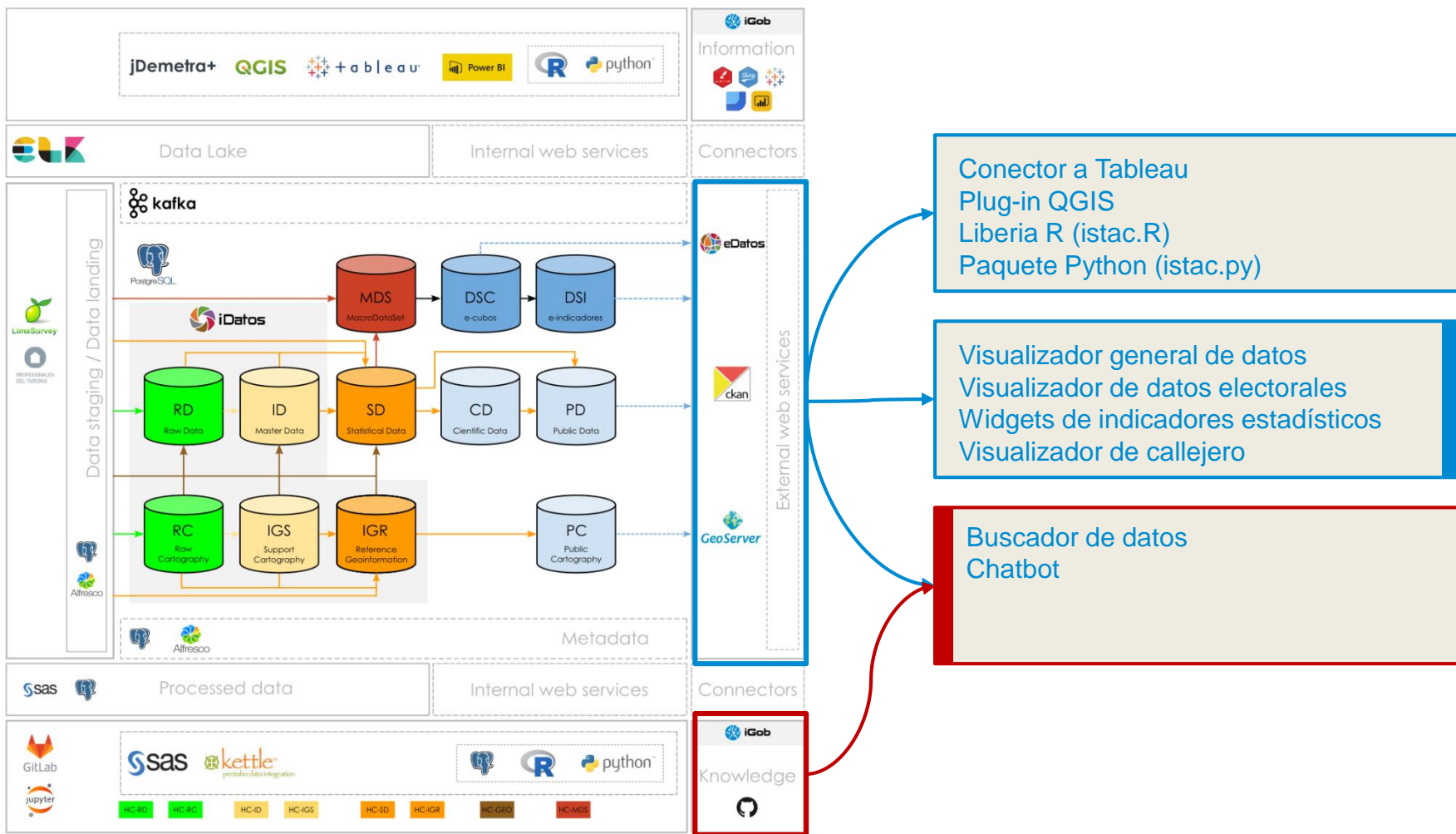


Están sujetas a la publicidad de resultados



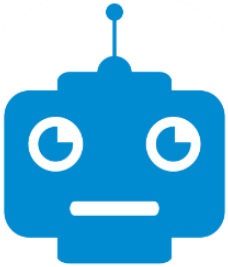
Están sujetas al secreto estadístico

**NO DEBERÍAN DUPLICARSE LOS ORGANISMOS PÚBLICOS Y SOLAPARSE COMPETENCIAS**





**Buscador de datos**



`@istac_bot`



**Georreferenciación**



**Predicción**

# GRACIAS POR SU ATENCIÓN

**Más información:**

[www.gobiernodecanarias.org/istac](http://www.gobiernodecanarias.org/istac)

[www.slideshare/istac](http://www.slideshare/istac)

[@istac\\_es](https://twitter.com/istac_es)

