

# Entorno de cómputo estadístico en eDatos: Hilos computacionales, ETL y microservicios de procesamiento

**Autores:** Jesús Alberto González Yanes, Esther Torres Medina, Pablo Betancor Ortiz, Carlos Javier Pérez González, Andrés Nacimiento García, José Carlos Gil Carreras

**Organismo de procedencia:** Instituto Canario de Estadística (ISTAC)

**Palabras clave :** ETL, microservicios, R, Python, Kettle, SAS, Jupyter, Postgre.

**Resumen:** El artículo 36 de la Ley 1/1991 de Estadística de la Comunidad Autónoma de Canarias indica que el Instituto Canario de Estadística es el organismo responsable de centralizar, conservar, ordenar, inventariar y custodiar los ficheros de datos para fines estadísticos, así como, los documentos y metadatos necesarios para su interpretación y uso, que sean de titularidad propia o compartida o de uso permitido para el desarrollo de la actividad estadística de interés de la Comunidad Autónoma de Canarias.

Esta capacidad está dirigida a mejorar la eficiencia de la actividad estadística de interés de la Comunidad Autónoma de Canarias y para su desarrollo, durante la ejecución del Plan Estadístico de Canarias 2018-2022, se impulsará la Infraestructura de Datos y Metadatos Estadísticos de Canarias (eDatos) como la infraestructura central que integra datos, metadatos, tratamientos y servicios comunes e interoperables de la estadística de interés de la Comunidad Autónomas de Canarias. La Infraestructura de Datos y Metadatos Estadísticos será el soporte para la difusión abierta e interoperable de los datos publicados por las operaciones del Plan Estadístico de Canarias. A su vez la infraestructura será el canal único para la difusión descentralizada de las estadísticas en las web corporativas del Gobierno de Canarias.

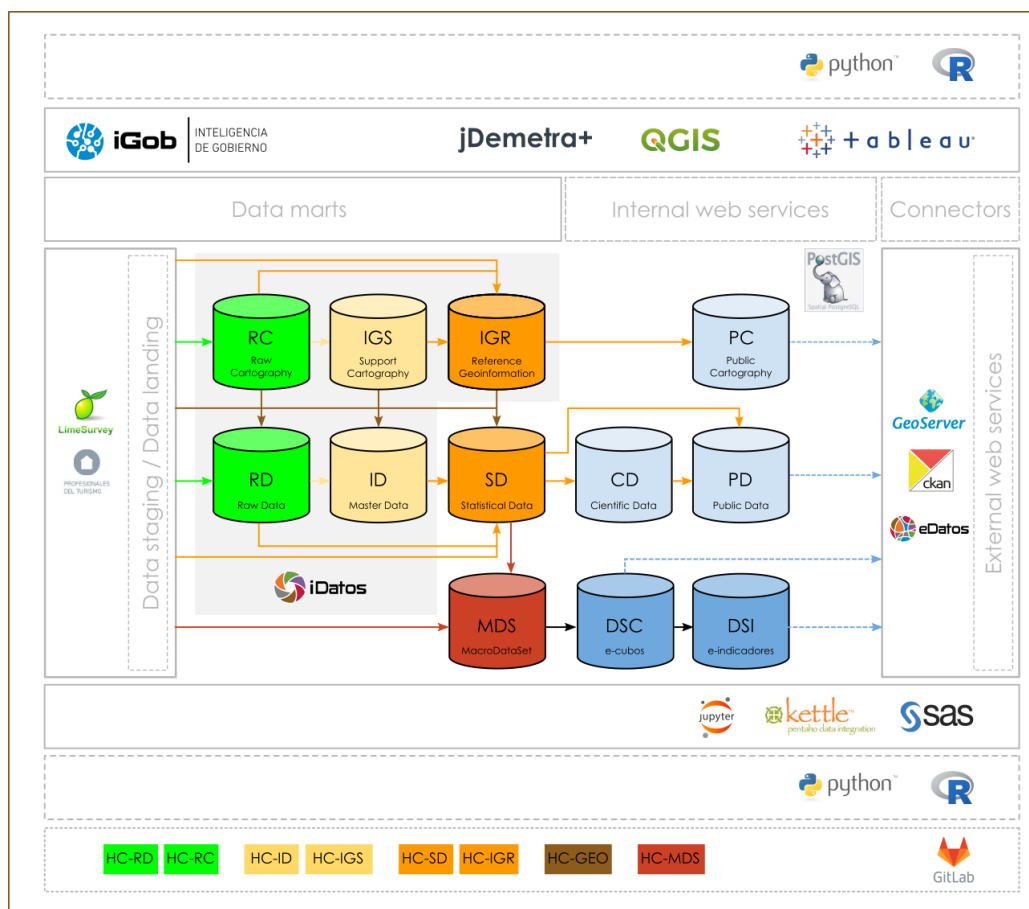
En el desarrollo de la infraestructura se ha establecido una estrategia de catalogación de tipologías de hilos computacionales y procesos ETL asociados. Los hilos computacionales recogen procesos de tratamiento, sobre microdatos o macrodatos, que son generadores de un cambio de naturaleza de los mismos. Los hilos computacionales son procesos de extracción, computación y carga de datos entre entornos del banco de datos o entre esquemas tipo del entorno repositorio. El procesamiento de datos, y el software asociado, se ha estructurado en tres niveles: alto, medio y bajo. Esta estructuración permite la especialización del personal, una mejor distribución de roles, así como el desarrollo a bajo nivel de microservicios horizontales de cómputo. La organización, catalogación y gestión de código de cómputo se gestionará en GIT según operaciones estadísticas e hilos computacionales.

<b>I. HILOS DE PROCESAMIENTO EN EL BANCO DE DATOS</b>	<b>2</b>
1.1. Arquitectura del Banco de Datos	2
1.2. Hilos computacionales y de transferencia	5
<b>II. HILOS COMPUTACIONALES</b>	<b>6</b>
2.1. Entornos de cómputo y categorías de hilos computacionales	6
2.2. Estructuración del cómputo por capas	7
2.3. Identificación de tratamientos por categorías de hilo	8
<b>III. MICROSERVICIOS DE PROCESAMIENTO</b>	<b>10</b>
3.1. Entornos de ejecución y documentación	10
3.2. Arquitectura de integración SAS-R-Python	11
3.3. Catálogo de macros SAS	11
3.4. Microservicios disponibles	12

# I. HILOS DE PROCESAMIENTO EN EL BANCO DE DATOS

## 1.1. Arquitectura del Banco de Datos

La arquitectura del banco de datos del Instituto Canario de Estadística (ISTAC) está dirigida al procesamiento supervisado de datos por lotes. Esta arquitectura es similar a las planteadas en otras Oficinas Estadísticas y contrasta con las actuales arquitecturas Lambda o Kappa<sup>1</sup> de procesamiento automático de Big Data tanto en modo batch como en streaming de datos.



- **Entornos de recepción, captura y prealmacenamiento:** Los entornos de recepción, captura y prealmacenamiento de datos son entornos donde se almacenan y catalogan previamente los inputs del banco de datos. Estos entornos pueden ser tecnológicamente variados, desde base de datos SQL, como no-SQL o servidores de ficheros. Además pueden

<sup>1</sup> Julien Forgeat. "Data Processing Architectures – Lambda and Kappa." Ericsson Research Blog. Accessed October 24, 2016.

estar soportados o no por sistemas de gestión (por ejemplo LimeSurvey o Profesionales del Turismo).

- **Entorno repositorio:** El entorno repositorio es un entorno de base de datos, donde se almacenan las cartografías, los microdatos y los macrodatos que son el resultado final de procesos de tratamiento generadores de un cambio de naturaleza de los mismos. El entorno repositorio se organiza en esquemas tipo, según la naturaleza de los ficheros que contiene, y estos esquemas se agrupan en tres tipos de niveles según almacenen cartografías, microdatos o macrodatos.
- **Entornos computacionales:** Los entornos computacionales son entornos para el procesamiento eficiente de datos y su objetivo es el cambio de naturaleza de un fichero de datos. Pueden existir distintos entornos computacionales según los requerimientos de cálculo y los requerimientos tecnológicos exigidos por los programas de computación estadística. A su vez estos entornos pueden también organizarse en esquemas tipo según la fase en la que se aborda un tratamiento.
- **Entornos analíticos:** Los entorno analíticos son entornos para el análisis exploratorio de datos. Estos entornos requieren acceso al entorno repositorio, o también necesitan de copias o transformaciones de datos para su función analítica.
- **Entorno de difusión:** No se trata de un entorno en sí mismo, sino de una capa de servicios web para la difusión de datos y cartografía.

Estos cuatro tipos de entorno se pueden clasificar en dos grupos: Los entornos de transición y los entornos de destino. En los entornos de transición los datos residen provisionalmente para la realización de algún tratamiento estadístico que los cambie de naturaleza, en cambio en los entornos de destino los datos se depositan con su naturaleza ya definitiva. La clasificación de entornos en ambas categorías en la siguiente:

Entornos	Clase de entorno
Entornos de recepción y captura	Transición
Entorno repositorio	Destino
Entornos computacionales	Transición
Entornos analíticos	Destino

Como decíamos, el entorno repositorio se organiza en esquemas-tipo, según la naturaleza de los ficheros que contiene, y estos esquemas se agrupan en tres tipos de niveles según almacenen **cartografías, microdatos o macrodatos**.

Grupo	Tipo de datos	Descripción
<b>Cartografía</b>	Raw Cartography (RC)	Esquemas de cartografía en crudo, con sólo tratamiento de normalización.
	Support Cartography (IGS)	Esquemas de cartografía de soporte. En estos esquemas se almacenan cartografías normalizadas y depuradas, utilizadas para procesamiento y análisis espacial.
	Geographic Information Reference (IGR)	Esquemas de cartografía de referencia. En estos esquemas se almacenan cartografías normalizadas, transformadas y depuradas, para su uso como entidades geográficas de difusión estadística. En este caso se almacenan distintas geometrías e información auxiliar.
	Public Cartography (PC)	Esquemas de cartografía estadísticas de uso público. En estos esquemas se almacenan las cartografías públicas para su uso como entidades geográficas de difusión estadística (IGR) y otras cartografías almacenadas en IGS que puedan ser de interés analítico.
<b>Microdatos</b>	Raw Data (RD)	Esquemas de microdatos en crudo, con solo tratamiento de normalización.  Todo fichero se georreferencia desde su incorporación al Banco de Datos.
	Master Data (ID)	Esquemas de microdatos maestros, donde se almacenan los cuatro grandes directorios (direcciones, edificios y viviendas, población y hogares, empresas y establecimientos).  En el caso que nos ocupa es especialmente relevante el Directorio de Calles y Portales, donde se almacena la información base para la georreferenciación.
	Statistical Data (SD)	Esquemas de microdatos estadísticos de uso interno, no anonimizados y con georreferencias.
	Scientific Data (CD)	Esquemas de microdatos para fines científicos, de uso compartido para uso científico, con eliminación de identificadores y otras anonimizaciones indirectas suaves.  No se incluyen georreferencias, sólo geocódigos.

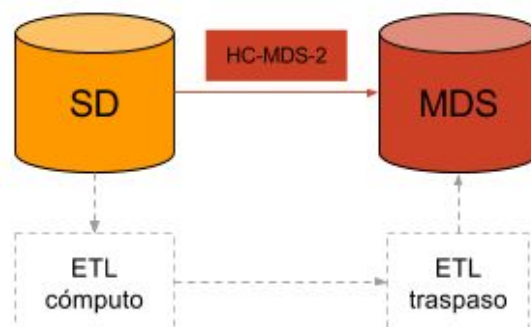
	Public Data (PD)	Esquemas de microdatos públicos, con eliminación de identificadores y anonimizaciones indirectas fuertes.  No se incluyen georreferencias, sólo geocódigos.
<b>Macrodatos</b>	MacroDataSet (MDS)	Esquemas de macrodatos estadísticos de uso interno, con almacenamiento de geocodigos.
	DataSetCube (DSC)	Esquema público de cubos estadísticos (eDatos), con almacenamiento de geocodigos.
	IndicatorsCube (DSI)	Esquema público de indicadores estadísticos (eDatos), con almacenamiento de geocodigos.

## 1.2. Hilos computacionales y de transferencia

Los **hilos computacionales** agrupan procesos de tratamiento de cambio de naturaleza de cartografías, microdatos o macrodatos. Los hilos computacionales son procesos de extracción, computación y carga de datos entre entornos del banco de datos o entre esquemas tipo del entorno repositorio. Un hilo computacional consta de categorías de dos procesos:

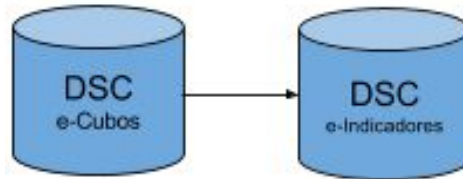
1. **Proceso ETL-computacional:** Proceso de extracción y carga de datos al entorno computacional para la realización de las transformaciones y cálculos estadísticos necesarios. Estos procesos se gestionan desde SAS Guide, siendo diseñados y ejecutados por mantenedores de datos (data steward).
2. **Proceso ETL-traspaso:** Proceso de extracción, transformación y carga desde entornos computacionales a entornos repositorios. En este proceso se realizan tratamientos de descripción y control de calidad de los microdatos o macrodatos. Estos procesos se gestionan desde Kettle, siendo diseñados y ejecutados por auditores de datos.

**Ejemplo de procesos ETL en hilo computacional**



Los **hilos de transferencia** son procesos de extracción y carga de datos o conexiones entre entornos del banco de datos o entre esquemas tipo del entorno repositorio. En los hilos de transferencia no existen cálculos, o son cálculos sencillos generados automáticamente.

**Ejemplo de hilo de transferencia entre e-Cubos y e-Indicadores**



## II. HILOS COMPUTACIONALES

### 2.1. Entornos de cómputo y categorías de hilos computacionales

Los entornos computacionales son entornos de transición para el procesamiento eficiente de datos y su objetivo es el cambio de naturaleza de un fichero de datos. Pueden existir distintos entornos computacionales según los requerimientos de cálculo y los requerimientos tecnológicos exigidos por los programas de computación estadística. A su vez estos entornos pueden también organizarse en categorías según la fase en la que se aborda un tratamiento, esta categorización se realiza en función de la naturaleza de los datos de destino y/o de los procesos involucrados:

Bloque	Categoría	Descripción
<b>NORMALIZACIÓN</b>	HC-RC	Hilo computacional de generación y carga de cartografía en crudo a partir de ficheros en entornos de recepción, captura y prealmacenamiento.
	HC-RD	Hilo computacional de generación y carga de microdatos en crudo a partir de ficheros en entornos de recepción, captura y prealmacenamiento.
<b>SOPORTE</b>	HC-IGS	Hilo computacional de generación y carga de Información Geográfica de Soporte (IGS) a partir de información almacenada en esquema de Raw Cartography (RC).
	HC-ID	Hilo computacional de generación y carga de microdatos maestros (ID) a partir de información almacenada en esquema de Raw Data (RD).
<b>BASE</b>	HC-IGR	Hilo computacional de generación y carga de Información Geográfica de Referencia (IGR) a partir de información almacenada en esquema de Raw Cartography (RC), Support Cartography (IGS)

		o a partir de ficheros en entornos de recepción, captura y prealmacenamiento.
	HC-SD	Hilo computacional de generación y carga de microdatos estadísticos (SD) a partir de información almacenada en esquema de de Raw Data (RD), Data Master (ID) o a partir de ficheros en entornos de recepción, captura y prealmacenamiento.
<b>GEOGRÁFICO</b>	HC-GEO	Hilo computacional de generación de georreferencias de ficheros de microdatos a partir de esquemas cartográficos, datos maestros o fuentes externas.
<b>AGREGACION</b>	HC-MDS	Hilo computacional de generación de ficheros de macrodatos (MDS) y datasetcubes (DSC) a partir de ficheros de microdatos estadísticos (SD), o a partir de ficheros en entornos de recepción, captura y prealmacenamiento.

## 2.2. Estructuración del cómputo por capas

En un **hilo computacional** se recogen los procesos de tratamiento sobre cartografías, microdatos o macrodatos, que son generadores de un cambio de naturaleza de los mismos. Un hilo computacional puede tener asociados distintos procesos ETL que puede ejecutarse en SAS Guide o en Kettle.

### Ejemplo de ETL de cómputo en un hilo HC-MDS-1

The screenshot displays the SAS Enterprise Guide interface. On the left, the 'Árbol del proyecto' (Project Tree) shows a hierarchy of tasks for 'LECTURA DE EXCEL Canarias'. The main workspace shows a workflow diagram with the following steps: 'Importar datos PLANTILLA ACIDIFICADO', 'Recodificar SECTOR\_ID ACIDIFICADO', 'Importar datos PLANTILLA METALES PESADOS', 'Recodificar SECTOR\_ID METALES PESADOS', 'Unión', 'EMISIONES\_00', 'Apilar columnas', 'EMISIONES\_01', 'Filtro y orden', and 'Añadir a Tabla Emisiones'. Below the workflow, a 'WORKEMISIONES' data preview is visible, showing a table with the following data:

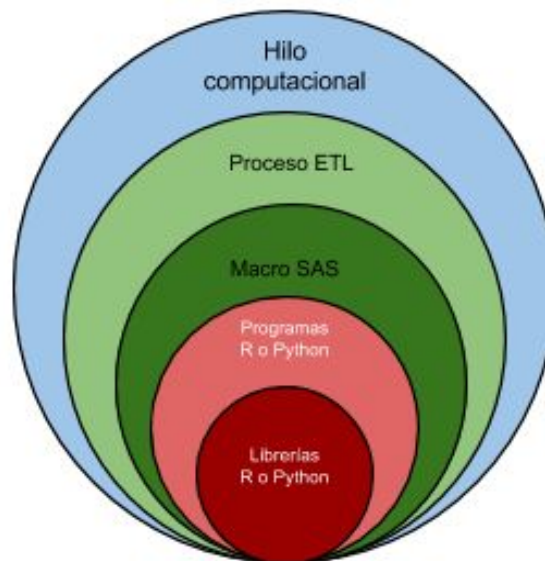
ANNO	SECTOR_ID	ELEMENTO	VALOR
1990		As (kg)	490.013487
1990	01	BC (t)	0
1990	01	CH4 (t)	311.1820482
1990	01	CO (t)	1439.78026
1990	01	CO2 (kt)	3589.023048
1990	01	CO2/NM (t)	708.2023015
1990	01	Cd (kg)	960.202309
1990	01	Cr (kg)	2400.593764
1990	01	Cu (kg)	960.1671771
1990	01	DIOX (g)	0.09823973
1990	01	HAP (kg)	0
1990	01	HCS (kg)	0
1990	01	HCH (kg)	0
1990	01	HFC (kg)	0
1990	01	Hg (kg)	960.1671771
1990	01	N2O (t)	18.01222333
1990	01	NM (t)	0
1990	01	NCx (t)	17119.80068
1990	01	Ni (kg)	33602.3302
1990	01	PCB (kg)	1.505866E-7

Para su uso en diferentes procesos ETL se identifican clases de procesamientos y se automatizan a través de lo que denominamos **microservicios de procesamiento**. Estos microservicios son las siguientes tipologías:



1. **Macro SAS:** Tratamiento normalizado que se pone a disposición de los técnicos estadísticos. Una macro SAS puede hacer uso de programas R o Python, que son transparentes para los técnicos estadísticos, y que también podrían ser llamados desde Kettle.
2. **Programas R o Python:** Un programa R o Python ejecuta procedimientos normalizados llamados desde SAS o Kettle. Son transparentes para los técnicos estadísticos y su ejecución se realiza en los servidores del ISTAC, lo que permite mejorar la velocidad de cómputo. Los programas R o Python utilizan librerías o paquetes desarrollados por terceros, aumentando la eficiencia de la organización.
3. **Librerías y paquetes R o Python:** Las librerías aumentan la capacidad de funciones de cómputo en R o Python.

**Esquema de profundidad en capas de ETL-computacional**



### 2.3. Identificación de tratamientos por categorías de hilo

Para la sistemación y normalización de cada categoría de hilo computacional se identifican los tratamientos involucrados, agrupándolos según los procesos y subprocessos del **Generic Statistical Business Process Model (GSBPM)**. Por ejemplo, para el hilo computacional HC-MDS-01 que transita desde los entornos de captura de macrodatos hasta el repositorio de macrodatos privados (MDS) hemos identificado los siguientes tratamientos:

1. RECOGER / OBTENER DATOS (GSBPM-4)
2. PROCESAR (GSBPM-5)
  - a. Codificación

- b. Normalización de códigos
  - c. Integración
  - d. Generación de subtotales derivados
  - e. Generación de estimadores derivados para series temporales
  - f. Generación de estimadores corregidos de estacionalidad y calendario
3. ANALIZAR (GSBPM-6)
- a. Generación de atributos de estado OBS\_STATUS
  - b. Control de coherencia entre totales y subtotales
  - c. Detección de outliers en series temporales
  - d. Generación de atributos de confidencialidad OBS\_CONFIDENCIALITY
  - e. Cargar en espacio repositorio (ETL de traspaso)
  - f. Cargar en espacios analíticos (hilo de transferencia)
  - g. Análisis exploratorio en espacios analíticos (fuera de hilo computacional)
4. DIFUNDIR (GSBPM-7)
- a. Elaborar datasets cube (DSC)
  - b. Cargar en espacio repositorio los DSC
  - c. Cargar en espacios analíticos (hilo de transferencia)
  - d. Análisis exploratorio en espacios analíticos
  - e. Cargar en e-Cubos (hilo de transferencia)

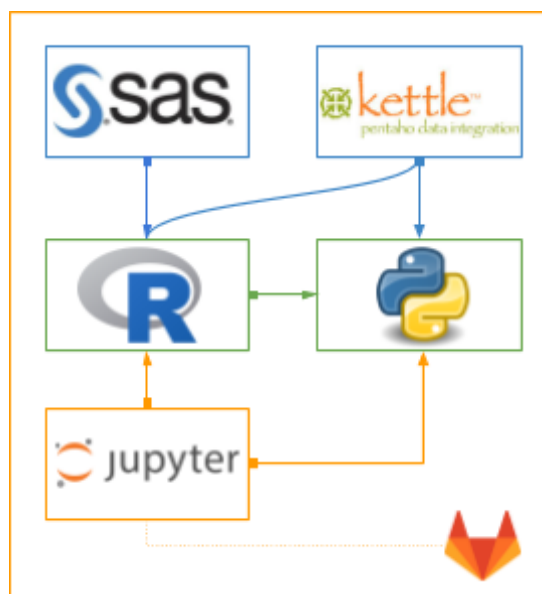
Una vez identificados los posibles tratamientos implicados en una categoría de hilo computacional, se procede a la identificación de cuáles de ellos pueden ser generalizados y apoyados mediante microservicios de procesamiento. Siguiendo con los el caso que nos ocupa, hasta ahora se han identificado los siguientes microservicios:

- 1. RECOGER / OBTENER DATOS
  - a. Macro SAS-R de captura de datos de la API del INE
  - b. Macro SAS-R de captura de datos de ficheros PC-Axis
  - c. Macro SAS-R-PYTHON de captura de datos de AENA
  - d. Macro SAS-R de captura de datos de Eurostat
- 2. PROCESAR
  - a. Librería R de lectura de listas de códigos y conceptos de medida desde el repositorio de metadatos
  - b. Macro SAS-R de lectura de listas de códigos y conceptos de medida desde el repositorio de metadatos
  - c. Macro SAS de estimadores derivados en series
- 3. ANALIZAR
  - a. Macro SAS de detección de outliers en series temporales
  - b. Macro SAS-R de control de secreto estadístico en macrodatos

### III. MICROSERVICIOS DE PROCESAMIENTO

#### 3.1. Entornos de ejecución y documentación

El procesamiento de datos, y el software asociado, se ha estructurado en tres niveles: De workflow (azul), de soporte (verde) y documentación. Esta estructuración permite la especialización del personal, una mejor distribución de roles, así como el desarrollo a bajo nivel de microservicios horizontales de cómputo.



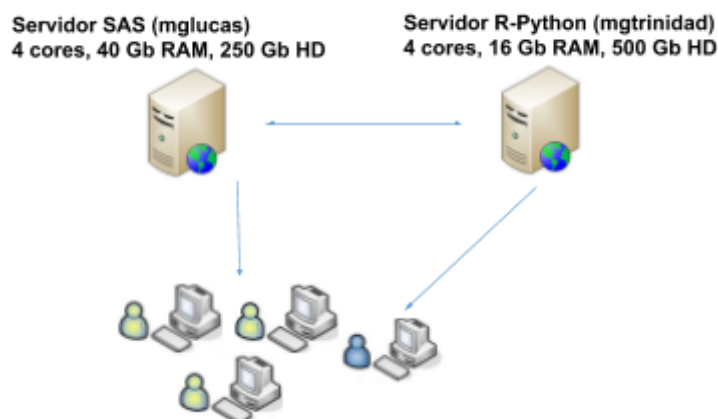
- Los **procesamientos de workflow** se realizan en SAS o Kettle, y estos pueden hacer uso de procesamientos de soporte en R o en Python. En el caso de SAS la conexión con R es directa, y con la conexión con Python es vía R.
- Los **procesamientos de soporte** se realizan en R o Python mediante scripts y librerías que bien se ejecutan directamente o bien son invocadas desde SAS y Kettle. Además estos procesamientos de soporte también están al servicio de los entornos analíticos (QGIS, Tableau o JDemetra+).
- El **procesamiento desde documentación** se realiza desde Jupyter, donde se facilita la ejecución documentada de procesamiento de soporte. Jupyter es una aplicación web de código abierto que permite crear y compartir documentos que contengan códigos, ecuaciones y visualizaciones en vivo. Es compatible con más de 40 lenguajes de programación.

Finalmente el **repositorio de procesamientos** será GitLab. Gitlab es un servicio web de control de versiones y desarrollo de software colaborativo basado en Git. Además de gestor de repositorios, el

servicio ofrece también alojamiento de wikis y un sistema de seguimiento de errores, todo ello publicado bajo una Licencia de código abierto.

### 3.2. Arquitectura de integración SAS-R-Python

Las macros se encuentran ya compiladas y listas para ejecución, ya que se ha procedido a crear un catálogo de macros que SAS carga al principio de la sesión. Este catálogo no se crea en el servidor multigestionado de SAS sino en otro servidor, también multigestionado por la Dirección General de Telecomunicaciones y Nuevas Tecnologías dentro del dominio GOBCAN.



### 3.3. Catálogo de macros SAS

El catálogo de macros SAS facilita la identificación y comprensión de las macros disponibles. Este catálogo se gestiona y documenta en la Wiki del ISTAC, soportada por Atlassian Confluence. Para consultar desde SAS la ayuda de las macros disponibles se realiza mediante `%help()`, o si se desea la ayuda de una particular `%help(<macro>)`.

#### Llamada de ayuda en SAS

TestHelp -

Programa Log Resultados - SAS Report Resultados - HTML

Actualizar Exportar Enviar a Publicar Propiedades

**Ayuda de macros SAS-ISTAC**

macro	hyperlink
GetRates	<a href="http://confluence/c/confluence/pages/viewpage.action?pageId=4228328#TIMECHANGE.Macrodectiulomedidasdevariaciónseries-getrates">http://confluence/c/confluence/pages/viewpage.action?pageId=4228328#TIMECHANGE.Macrodectiulomedidasdevariaciónseries-getrates</a>
GetRatesSet	<a href="http://confluence/c/confluence/pages/viewpage.action?pageId=4228328#TIMECHANGE.Macrodectiulomedidasdevariaciónseries-getrates">http://confluence/c/confluence/pages/viewpage.action?pageId=4228328#TIMECHANGE.Macrodectiulomedidasdevariaciónseries-getrates</a>
GetOutliers	<a href="http://confluence/c/confluence/pages/viewpage.action?pageId=4227785#TIMEOUTLIER.Macrodetectiendeobservacionesatipicasenseries-getoutliers">http://confluence/c/confluence/pages/viewpage.action?pageId=4227785#TIMEOUTLIER.Macrodetectiendeobservacionesatipicasenseries-getoutliers</a>
PlotOutliers	<a href="http://confluence/c/confluence/pages/viewpage.action?pageId=4227785#TIMEOUTLIER.Macrodetectiendeobservacionesatipicasenseries-plotoutliers">http://confluence/c/confluence/pages/viewpage.action?pageId=4227785#TIMEOUTLIER.Macrodetectiendeobservacionesatipicasenseries-plotoutliers</a>

#### Documentación en Confluence

TIMEOUTLIER. Macro de detección de observaciones atípicas en series  
Creado por Alberto González Yanes, modificado por última vez en jul 15, 2017

**En esta página:**

- Descripción general de la macro
- Medidas de observaciones atípicas
  - Medida basada en cuartiles
  - Medida basada en unidades tipificadas (z-scores)
- Variables de clasificación de las observaciones atípicas
- Ayuda para el uso de la macro
  - GetOutliers: Detección de observaciones atípicas
    - Descripción
    - Utilización
    - Argumentos
    - Detalles
    - Ejemplo
  - PlotOutliers: Representación gráfica de las series que contienen outliers
    - Descripción
    - Utilización
    - Argumentos
    - Ejemplo
  - FilterOutliers: Filtrar una serie temporal de un conjunto de datos
    - Descripción
    - Utilización
    - Argumentos
    - Ejemplo
- Test de uso de la macro sobre varios MDS de ejemplo
- Referencias

### 3.4. Microservicios disponibles

Bloque	Capa	Microservicio	Descripción
<b>Captura de datos</b>	Jupyter-Python	getAENA	Código Python, con documentación y ejecución desde Jupyter, que permite la realización de web-scraping a los datos publicados por AENA.
	Python	getShops	Código Python que permite la extracción de datos de empresas y establecimientos de Big Data.
	Librería R	INEbaseR	Librería de extracción de datos de la API del INE. Permite extraer operaciones y series de datos de forma estructurada y eficiente. La librería permite: <ul style="list-style-type: none"> <li>- Obtención de datos estructurados</li> <li>- Obtención de las series estadísticas mediante lenguaje de consulta estructurada</li> <li>- Análisis de datos</li> </ul>
	Librería R	istacr	Librería en R para la extracción de datos y metadatos de las APIs del ISTAC. La librería permite: <ul style="list-style-type: none"> <li>- Acceso a indicadores</li> <li>- Acceso a variables y listas de códigos</li> <li>- Recodificación</li> <li>- Acceso a información cartográfica</li> </ul>
	Paquete Python	istacpy	Paquete en R para la extracción de datos y metadatos de las APIs del ISTAC. La librería permite: <ul style="list-style-type: none"> <li>- Acceso a indicadores</li> <li>- Acceso a variables y listas de códigos</li> <li>- Recodificación</li> <li>- Acceso a información cartográfica</li> </ul>
<b>Tratamiento de datos</b>	Macro SAS	CALMAR	Macro SAS desarrollada por el INSEE para el calibrado de muestras con información auxiliar.

	Macro SAS	MacroPX	Macro de generación de tablas en formato PX a partir de tablas SAS.
	Macro SAS-R	GEOCODE	Macros de geocodificación. Métodos incluidos: <ul style="list-style-type: none"> <li>- getgeocode: Obtención de los geocódigos de una georreferencia.</li> </ul>
<b>Series temporales</b>	Macros SAS	TIMEOUTLIER	Macros de detección de observaciones atípicas en series. Métodos incluidos: <ul style="list-style-type: none"> <li>- GetOutliers: Detección de observaciones atípicas</li> <li>- PlotOutliers: Representación gráfica de las series que contienen outliers</li> <li>- FilterDataPlot: Filtrar una serie temporal de un conjunto de datos</li> </ul>
	Macros SAS	TIMECHANGE	Macros de cálculo de medidas de variación en series. Métodos incluidos: <ul style="list-style-type: none"> <li>- GetRates y GetRatesSet: Cálculo de indicadores/tasas/índices</li> <li>- MergeData: Fusión de datasets</li> <li>- CompleteFieldID: Crea o incrementa con nuevos valores una variable ID</li> <li>- CompleteFieldSTATUS: Crea o actualiza con nuevos valores una variable STATUS</li> <li>- MergeDataMDS: Fusión de MDS datasets</li> <li>- CheckCompleteTS: Comprobación de completitud de las series de datos</li> </ul>