**Documentación técnica de funcionamiento de script conversor de datos de seguimiento a datos agregados**

**ESAVI**

**08-03-2024**

**Repositorio:**

**url:** **https://github.com/HISPColombia/Scripts-ESAVI**

**Finalidad del script:**

El desarrollo del script se ha orientado hacia la conversión de datos de seguimiento de casos en datos agregados de la población. Este proceso comienza con un filtro inicial que se aplica a las unidades organizativas, ya sean municipios o distritos, para contar los casos correspondientes. Luego, se procede a segmentar estos datos en función del periodo de registro de los casos, seguido por la clasificación en rangos de edades y género, y finalmente, por los síntomas reportados, utilizando los códigos de MedDRA.

Es fundamental destacar que esta conversión de datos abre la puerta a una amplia gama de posibilidades para la visualización y generación de informes basados en la información recopilada. Por ejemplo, se puede realizar análisis para determinar cuál es el síntoma más frecuente dentro de un grupo específico de edad, simplemente seleccionando las categorías mencionadas anteriormente. Esto no solo proporciona una comprensión más profunda de los datos, sino que también permite identificar tendencias y patrones relevantes que podrían influir en la toma de decisiones en materia de salud pública.

**Descripción técnica:**

**Lenguaje**

* Python

**Librerías**

* Requests
* Json
* Pandas

**Fuente de datos:**

* PAHO-ESAVI: <https://www.paho-dhis2.org/>

**Fuente receptora:**

* PAHO-DEV-ESAVI: https://dev.paho-dhis2.org/

**Requerimientos mínimos para ejecutar el script:**

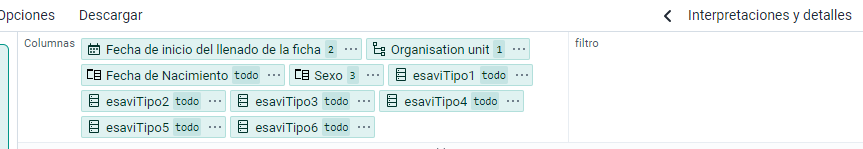
* Python 3.10 en el servidor o contenedor
* Conexión a internet.
* Credenciales de acceso a las plataformas emisor y receptora de datos (si los mismos datos van a una misma instancia solo ocuparía un parámetro de seguridad).
* Permisos para generar tareas programadas en el servidor.

**Secuencia de funcionalidad:**

**1)Recopilación de datos a convertir:**

Según la fuente de datos, es crucial llevar a cabo una consulta óptima en la aplicación Line-listing de la plataforma DHIS2. Para acceder al informe, se debe seleccionar el anexo 1 y realizar el inicio de sesión correspondiente. Este reporte debe incluir los parámetros necesarios para asegurar un agrupamiento adecuado de los datos, tal como se ilustra en la imagen adjunta a continuación.

Principio del formulario



Estos los datos recopilados en para la generación de los agregados

**2) Limpieza de datos vacíos.**

Después de completar la recopilación de datos, se lleva a cabo una revisión exhaustiva por parte del script para detectar cualquier dato vacío. Esta medida tiene como objetivo evitar la generación de procesos innecesarios y garantizar un funcionamiento eficiente del sistema. Si se identifica algún dato vacío que no cumpla con los parámetros establecidos, el script no permitirá realizar el conteo o agrupamiento de estos datos. Esta medida contribuye a evitar el uso innecesario de recursos y asegura que solo se procesen los datos relevantes y completos para el análisis.

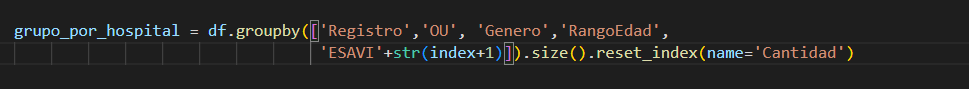


En la figura anterior se puede observar algunos casos que poseen como mínimo un tipo de ESAVI, por lo tanto, algunas columnas quedan vacías, las cuales se deben desagregar en los conteos para evitar que se realicen procesos innecesarios.

**3) Agrupamiento de datos según los criterios de filtrado.**

El script tiene la capacidad de agrupar los datos según los criterios establecidos, como el ítem del reporte. Además, utiliza tecnología basada en la librería de pandas para llevar a cabo la creación o agrupamiento de datos, lo que genera eficiencia en la búsqueda o el agrupamiento de datos. También realiza un conteo de coincidencias dependiendo de los parámetros establecidos, como el lugar del caso, la fecha de nacimiento y el tipo de síntoma (ESAVI).

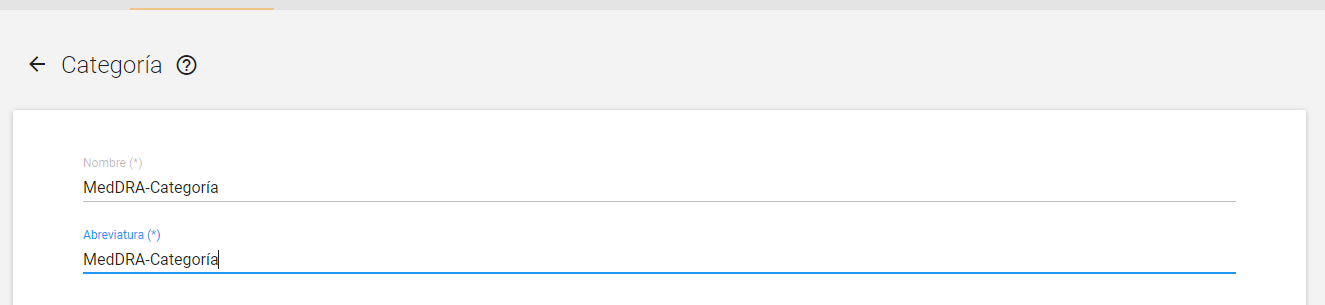


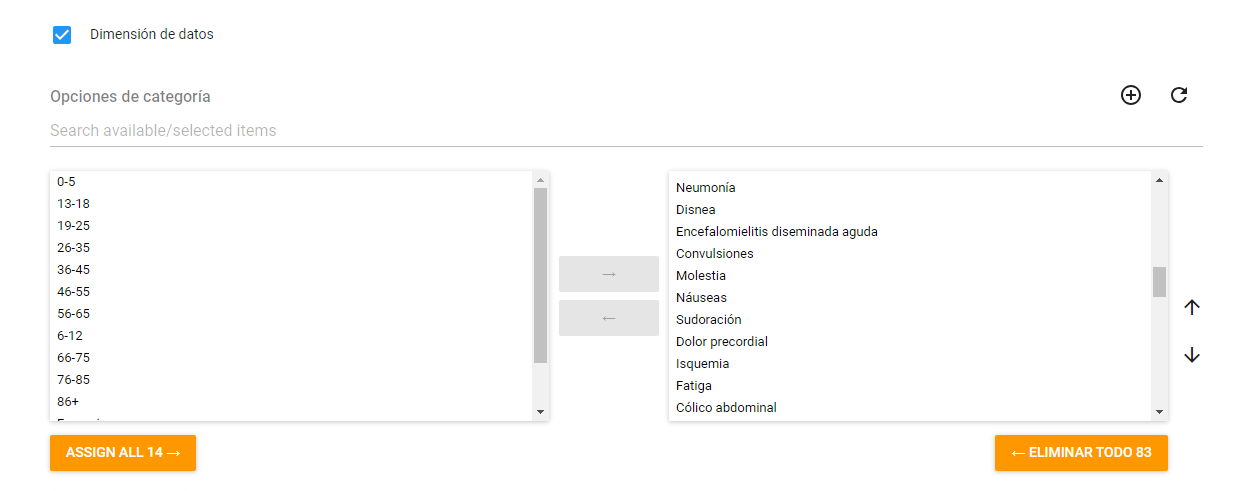


En el anterior fragmentó de código se puede apreciar que columnas se están utilizando para la recepción de los datos (dimensiones a categorizar)

**4) Consulta de categorías existentes.**

Se lleva a cabo la consulta de categorías existentes, generando únicamente la metadata necesaria para el conteo y la agregación de datos. Esta interacción con el sistema DHIS2 permite generar las categorías necesarias para el proceso.

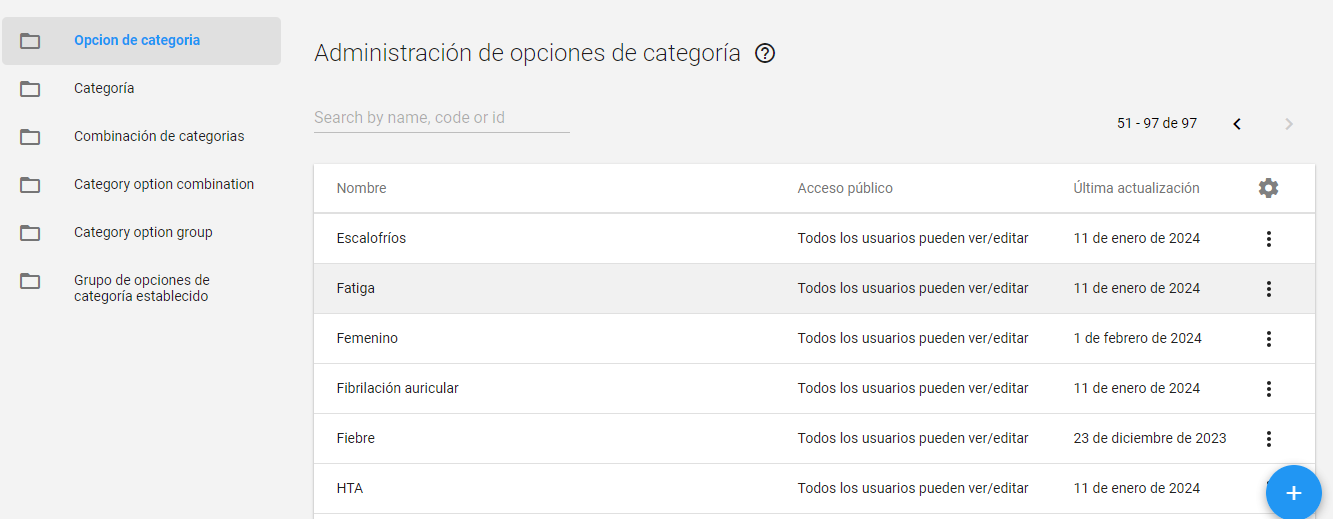




En la figura anterior se puede observar el proceso manual que se debe de realizar cuando se agrega una opción de categoría, (el sistema actual realiza la operación automatizada después de haber generado las opciones de categorías no existentes).

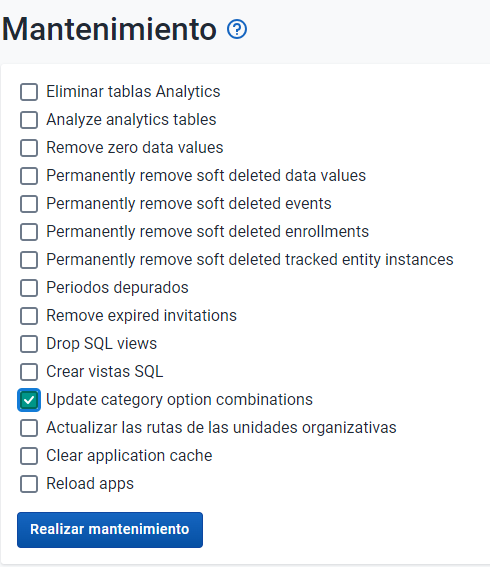
**5) Consulta y creación de opciones de categorías (solo se genera la creación de esta cuando se detecta un síntoma nuevo).**

El script realiza la consulta y creación de opciones de categorías solo cuando se detecta un síntoma nuevo. Este proceso es similar al anterior, ya que realiza la consulta cuando se agrega un nuevo ítem a la lista, generando las opciones de categoría correspondientes cuando se encuentra un nuevo síntoma. Además, se actualizan las categorías según sea necesario para garantizar el éxito de la importación de los datos.



**6) Combinación de categorías con las nuevas opciones creadas.**

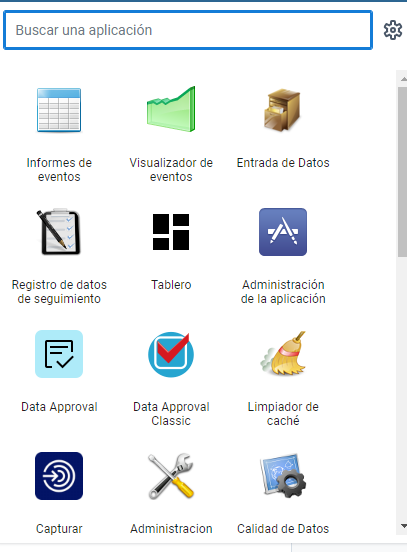
Cuando se realiza una creación o actualización de la categoría debido a la agregación de una nueva opción, el sistema es capaz de realizar la combinación necesaria.



En la figura anterior se presenta el paso manual que se debe de realizar mediante la aplicación “Administración de datos”, lo que hace es hacer la combinación de las posibles nuevas opciones de categoría, con las categorías creadas, no obstante, el script también genera este mismo proceso para evitar dependencias en la carga de datos.

**7) Actualización del dataSet recolector de datos agregados**

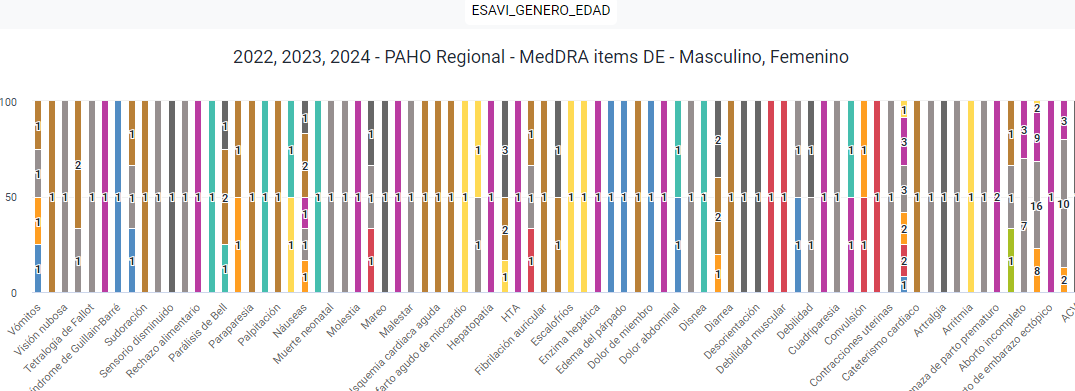
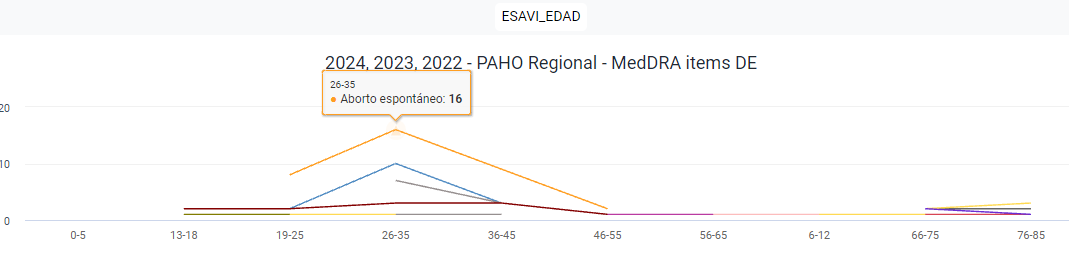
Después de modificar la categoría asociada al elemento de datos del dataset, es necesario limpiar la caché del sistema para evitar posibles problemas relacionados.

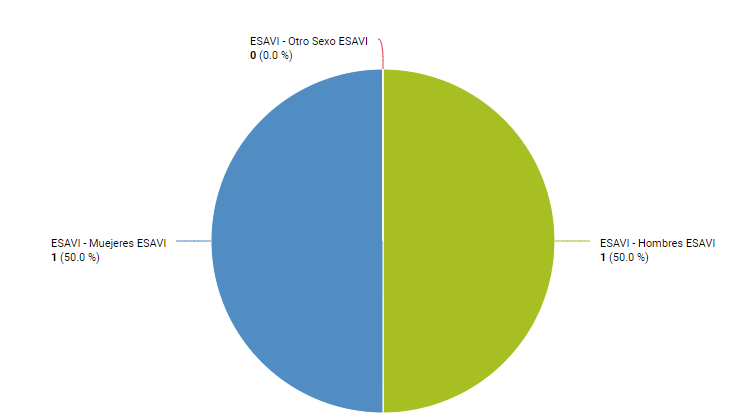


Nuevamente se puede mencionar que esta tarea también es realizada por el script para evitar dependencias en el proceso de carga de datos.

**8) Carga de datos al sistema receptor**

Durante el proceso de carga de datos al sistema receptor, el sistema ya tiene listo el contenedor de los datos y el agrupamiento de los mismos para su carga en la instancia final. Se recomienda tener un usuario capacitado para realizar esta carga de datos en la instancia.





En las figuras anteriores se puede visualizar, algunos métodos de presentación de la información agregada en el sistema optimizando el consumo de indicadores de programa.

**Anexo 1**

1)<https://www.paho-dhis2.org/api/38/analytics/events/query/aFGRl00bzio?dimension=ou%3AMDzTj1yWApH%3Bq8kYW2iwzeU,NI0QRzJvQ0k,oindugucx72%3AIN%3A2%3B1%3B3,lSpdre0srBn.PZxZirhNzgS,lSpdre0srBn.maY0Vi68Fv9,lSpdre0srBn.Sy1uqYvgR3r,lSpdre0srBn.Og99AH5tIQz,lSpdre0srBn.vqf60JfNqsf,lSpdre0srBn.pQJc4VA2SDW&headers=eventdate,ouname,NI0QRzJvQ0k,oindugucx72,lSpdre0srBn.PZxZirhNzgS,lSpdre0srBn.maY0Vi68Fv9,lSpdre0srBn.Sy1uqYvgR3r,lSpdre0srBn.Og99AH5tIQz,lSpdre0srBn.vqf60JfNqsf,lSpdre0srBn.pQJc4VA2SDW&totalPages=false&eventDate=THIS_YEAR,LAST_5_YEARS&displayProperty=SHORTNAME&outputType=EVENT&page=1&includeMetadataDetails=true&stage=lSpdre0srBn&pageSize=1000>