

05. - Data Collection_CU_18_01_infrashp_v_01

June 13, 2023

#

CU18_Infraestructuras_eventos

Citizenlab Data Science Methodology > II - Data Processing Domain *** > # 05.- Data Collection

Data Collection is the process to obtain and generate (if required) necessary data to model the problem.

0.0.1 01. Descarga de datos de infraestructuras

- En esta tarea se descargan los datos actualizados de las infraestructuras sanitarias y de transporte
- En el piloto, se han descargado los datos en formato shp siguientes de la url <https://gestion.comunidad.madrid/nomecalles/DescargaBDTCorte.icm#delimitacionesterritoriales>:
- Salud_ Centros sanitarios. Centros de Atención Integral a Drogodependientes
- Salud_ Centros sanitarios. Centros de especialidades
- Salud_ Centros sanitarios. Centros de salud
- Salud_ Centros sanitarios. Consultorios de salud
- Salud_ Centros sanitarios. Otros centros
- Salud_ Centros sanitarios. Servicio de salud mental de distrito
- Transportes y comunicaciones_ Aeropuertos
- Transportes y Comunicaciones_ Cercanías (estaciones)
- Transportes y comunicaciones_ Helisuperficies
- Transportes y Comunicaciones_ Intercambiadores
- Transportes y comunicaciones_ Metro (bocas)

NOTAS: Los datos de hospitales no se encuentran en esta fuente pero se han obtenido en el caso de uso CU25 y se utilizará el archivo generado ahí.

Table of Contents

Settings

Data Load

ETL Processes

Import data from: CSV, Excel, Tab, JSON, SQL, and Parquet files

Synthetic Data Generation

Fake Data Generation

Open Data

Data Save

Main Conclusions

Main Actions

Acciones done

Acctions to perform

0.1 Settings

0.1.1 Packages to use

No aplica

```
[ ]: # library(xxxx)
```

0.1.2 Paths

```
[ ]: # iPath <- "Data/Input/"  
# oPath <- "Data/Output/"
```

0.2 Data Load

No aplica

0.3 ETL Processes

0.3.1 Import data from: CSV, Excel, Tab, JSON, SQL, and Parquet files

No aplica

0.4 Synthetic Data Generation

No aplica

0.5 Fake Data Generation

No aplica

0.6 Open Data

Los archivos shp se descargaron de la fuente abierta como ficheros comprimidos .zip que se descomprimieron para tener la estructura de carpetas necesaria

0.7 Data Save

No aplica

0.8 Main Conclusions

List and describe the general conclusions of the analysis carried out.

Observaciones generales sobre los datos

- Los datos están en formato shp con las coordenadas y alguna información adicional, casi todo referido a la dirección
- Para cada tipo de infraestructura pueden venir unos campos u otros

0.8.1 Consideraciones para despliegue en piloto

- Los datos de origen fueron obtenidos de <https://gestiona.comunidad.madrid/nomecalles/DescargaBDTCorte> en fichero zip con formato .shp y descomprimido en la carpeta Data/Input
- Si la información geográfica cambia, se debería actualizar el fichero y volver a ejecutar todos los procesos.

0.8.2 Consideraciones para despliegue en producción

- Se deben crear los procesos ETL en producción necesarios para que los datos de entrada estén actualizados

0.9 Main Actions

Acciones done Indicate the actions that have been carried out in this process

- Se han descargado y descomprimido los archivos

Acctions to perform Indicate the actions that must be carried out in subsequent processes

- Se debe consolidar la información en un fichero CSV
- Se deban añadir los datos de los hospitales

0.10 CODE TO DEPLOY (PILOT)

A continuación se incluirá el código que deba ser llevado a despliegue para producción, dado que se entiende efectúa operaciones necesarias sobre los datos en la ejecución del prototipo

Description

- No hay nada que desplegar en el piloto, ya que estos datos son estáticos o en todo caso cambian con muy poca frecuencia, altamente improbable durante el proyecto.

CODE

```
[ ]: # incluir código
```