

Deliverable 3 : Data Management Plan

Proyecto: Multi-scale Spatio-Temporal Analysis of Research Data. Agencia Estatal de Investigación, MICIU/AEI/10.13039/501100011033/, grant no. PID2023-152301OB-I00.

Contacto: [Luz García](#)

Última modificación: 04/07/2025

Fuentes: El proyecto latex editable de este documento se encuentra en overleaf, en: <https://es.overleaf.com/project/683b36a53d762138cca8622d>

1. Introducción

El proyecto **MuSTARD**: *Multi-scale Spatio-Temporal Analysis of Research Data* [1] desarrolla un marco computacional innovador para promover la transformación digital de sectores estratégicos de Salud, Clima y Seguridad. Se van a generar diversos tipos de datos, compartidos entre los miembros del proyecto y de los proyectos colaboradores asociados, incluyendo informes, publicaciones científicas, bases de datos, software en abierto y otros. Hay que destacar que los datos relativos a las aplicaciones de sensado remoto corresponden a proyectos colaboradores, y por extensión MUSTARD está sujeto a la gestión que se haga de los datos desde esos otros proyectos.

Los datos generados por el proyecto se pueden dividir en las siguientes categorías:

1. **Publicaciones de resultados de investigación** del proyecto en revistas científicas de calidad y presentaciones en congresos. La recopilación de publicaciones y presentaciones generadas se visibiliza en la web del proyecto <https://codas.ugr.es/mustard/en/results>
2. **Publicaciones divulgativas** sobre los resultados del proyecto. Estas publicaciones se visibilizan desde la web del proyecto <https://codas.ugr.es/mustard/en/results>
3. **Marcos computacionales y herramientas software** para el análisis espacio-temporal multi-escala de datos. La recopilación de herramientas generadas se visibiliza en la web del proyecto <https://codas.ugr.es/mustard/en/results>. Las herramientas están disponibles de manera abierta (garantizando la accesibilidad) en el github del proyecto <https://github.com/orgs/CoDaSLab/repositories>

4. **textbf**Documentación de uso interno sobre el desarrollo del proyecto. Disponible en el proyecto de github <https://github.com/users/JesusGarSan/projects/5> y en el DRIVE compartido por todos los miembros del mismo. Se garantiza la seguridad en el acceso a los datos mediante un sistema de permisos para acceso a los datos.

5. **Base de datos** para la aplicación de las herramientas computacionales. Las bases de datos que se usan en el desarrollo del proyecto son las siguientes

DB1. Base de datos para el estudio de medicina de precisión en cáncer colo-rectal.

DB2. Base de datos de registros sismo-volcánicos en el entorno del Volcán de la Palma.

DB3. Base de datos de monitorización del tráfico urbano mediante sensorización acústica distribuida (DAS).

DB4. Base de datos de imágenes satelitales para el estudio del cambio climático.

DB5. Base de datos de datos de movilidad geo-localizados.

Estas bases de datos están íntimamente relacionadas con los siguientes objetivos específicos del proyecto:

SO2.1. Application of ASCA to available spatio-temporal omics in colorectal cancer samples.

SO2.3. Spatio-temporal models for the study of Climate Change in Sierra Nevada.

SO2.4. Exploratory data analysis in multi-sensor volcano-seismic activity in Canary Islands.

SO2.5. Exploratory data analysis in urban mobility models from geolocated mobility data
SO2.5. Exploratory data analysis in urban mobility models from geolocated mobility data

SO2.6. Exploratory data analysis in urban mobility models from DAS data

Se describirán en detalle en las siguientes secciones del documento.

2. Base de datos de muestras de cancer colo-rectal

2.1. Resumen de los datos

Se realizará un estudio piloto prospectivo en 15 pacientes con diagnóstico de cáncer colorrectal. El reclutamiento y control clínico y seguimiento de los pacientes se realizará por parte de la Dra. Conde en el Hospital Universitario Virgen de las Nieves (Granada). Se intentará reclutar igual número de mujeres y de hombres, si bien somos conscientes de la existencia de una mayor prevalencia en hombres con respecto a las mujeres. A todos los participantes se les informará del proyecto y se obtendrá el consentimiento firmado antes de su inclusión en el estudio.

Se puede acceder a los datos en diferentes formatos:

2.2. Responsabilidades

La responsabilidad de la gestión continuada de los datos, incluyendo el acceso a los mismos, es del Investigador Principal del proyecto **MuSTARD**, [José Camacho](#).

2.3. Datos FAIR

2.3.1. Datos localizables y Accesibilidad

Tanto repositorio Github como servicio DAaaS son localizables desde la web del proyecto¹.

2.3.2. Interoperabilidad

Todos los formatos de datos son abiertos y legibles por herramientas de software libre. Se utiliza el conjunto de metadatos nativo de Github. En la web del proyecto se incluye título, link y referencias.

2.3.3. Reutilización

Los datos son reutilizables bajo licencia GPL-3.0. Los datos permanecerán reutilizables tras el fin del proyecto, sin limitación ni restricción de acceso.

¹<https://codas.ugr.es/animalicos>

2.4. Seguridad de los Datos

Todos los datos se basan en versiones anonimizadas de la traza original [?] o bien han sido completamente anonimizados al pasar a características.

La seguridad de los datos en formato F.3 se implementan por el propio repositorio Github.

Con respecto al resto, el **Data Analysis as a Service (DAaaS)** incluye un sistema complejo de copias de seguridad mediante *crontab*. Lo primero que se hace son las copias de seguridad del servicio DAaaS, una cada hora, de los 10 últimos días (añadiendo un *tag* con la fecha), para permitir guardar el estados de los usuarios. Adicionalmente, se hacen copias externas diarias del conedor que alberga el servicio a Docker Hub. Finalmente, se mantienen copias de seguridad de los datos en sendos HDD en el servidor y en un disco externo.

3. Base de datos sismo-volcánicos

3.1. Resumen de los datos

La base de datos está formada por ficheros de registros obtenidos por las estaciones sísmicas administradas por el Instituto Volcanológico de Canarias (INVOLCAN). Cada fichero, en formato miniSEED, corresponde a una señal medida por un canal de una estación a lo largo de un día.

3.2. Responsabilidades

La responsabilidad de la gestión continuada de los datos, incluyendo el acceso a los mismos, es del Investigador Principal del proyecto **MuSTARD**, [José Camacho](#).

3.3. Datos FAIR

3.3.1. Datos localizables y Accesibilidad

Los datos se encuentran alojados en un servidor administrado por INVOLCAN, accesible mediante conexión sftp, y pueden ser descargados libremente por los miembros del proyecto. Puesto que se trata de datos asociados a otro proyecto (Proyecto PLEC2022-009271 financiado por MICIU/AEI /10.13039/501100011033 y por la Unión Europea NextGenerationEU/ PRTR),

su accesibilidad libre está sujeta a criterios de accesibilidad de este proyecto.

3.3.2. Interoperabilidad

Todos los formatos de datos son abiertos y legibles por herramientas de software libre.

3.3.3. Reutilización

Los datos son reutilizables bajo licencia GPL-3.0. Los datos permanecerán reutilizables tras el fin del proyecto, sin limitación ni restricción de acceso.

3.4. Seguridad de los Datos

Todos los datos se basan en versiones anonimizadas de la traza original [?] y han sido completamente anonimizados al pasar a características. La seguridad de los datos se implementan por el propio repositorio Github.

4. Base de datos de monitorización del tráfico urbano

4.1. Resumen de los datos

La base de datos está formada por ficheros de registros obtenidos con la tecnología de sensorización acústica distribuida, en los que se monitorizan vibraciones mecánicas en dos fibras ópticas de la ciudad de Granada, que son debidas a eventos de movilidad urbana (tráfico de autobuses, peatones, coches, etc) así como eventos sísmicos o de cualquier otra índole que generen vibraciones mecánicas en el suelo solidario a la fibra. Se puede acceder a los datos en diferentes formatos:

- F1. Ficheros de variación de strain de duración de un minuto.
- F2. Ficheros de *waterfall de energía* de duración 1 hora.
- F3. Matrices de parametrizaciones diversas de la señal de variación de strain (formato F1).
- F4. Matrices de etiquetas manuales asociadas a los datos en formatos F1, F2, y F3.

Para la correcta lectura y preprocesado de todos los formatos de datos, se se ofrece una librería de lectura y preprocesado accesible en el github <https://github.com/luzgm/Urban-DAS-Monitoring>. Este repositorio contiene scripts para (i) lectura de los datos, (ii) pre-procesado, (iv) extracción de ruido en los mismos, (v) extracción de características paramétricas para generación de ficheros en formato F3 referido anteriormente.

4.2. Responsabilidades

La responsabilidad de la gestión continuada de los datos, incluyendo el acceso a los mismos, es del Investigadora Principal del proyecto **MuSTARD**, [Luz García](#).

4.3. Datos FAIR

4.3.1. Datos localizables y accesibilidad

La base de datos está disponible en un servidor NAS accesible mediante conexión sftp y acceso identificado, desde el que pueden ser descargados libremente por los miembros del proyecto. Como es resultado de la ejecución de 2 proyectos anteriores (Proyecto TED2021-132178B-I00/MICIU/AEI/ y Proyecto Junta de Andalucía/Consejería de Universidad, Investigación e Innovación/Proyecto P21₀0051.), su accesibilidad libre está sujeta a criterios de accesibilidad de estos proyectos originarios.

4.3.2. Interoperabilidad

Todos los formatos de datos son abiertos descritos anteriormente son legibles por herramientas de software libre. Se utiliza el conjunto de metadatos nativo de Github. En la web del proyecto se incluye título, link y referencias. ([Esto sin completar aún en la web](#))

4.3.3. Reutilización

Los datos son reutilizables bajo licencia GPL-3.0. Los datos permanecerán reutilizables tras el fin del proyecto, sin limitación ni restricción de acceso.

4.4. Seguridad de los Datos

La seguridad de los datos se implementan por el propio repositorio Github.

5. Base de datos de imágenes satelitales para el estudio del cambio climático

5.1. Resumen de los datos

5.2. Responsabilidades

La responsabilidad de la gestión continuada de los datos, incluyendo el acceso a los mismos, es del Investigador Principal del proyecto **MuSTARD**, [José Camacho](#).

5.3. Datos FAIR

5.3.1. Datos localizables y accesibilidad

5.3.2. Interoperabilidad

5.3.3. Reutilización

5.4. Seguridad de los Datos

Referencias

- [1] ANIMaLiCoS. Advanced networkmetrics: Interpretable machine learning for intelligent communication systems. <https://www.codas.ugr.es/animalicos/en>.