

# Propuesta trabajo en Grupo

M3.02-Data Life Cycle

Fernando Aguilar

# Trabajo en Grupo

- **Informe + Presentación (Final de Enero)**
- Punto de partida: Open Data (al menos **dos** fuentes diferentes)
- Descripción general del proyecto: objetivos, resultados esperados, requisitos, requerimientos técnicos...
- Descripción del problema:
  - Interés
  - Cobertura (geográfica, temporal). ¿Han limitado esta elección los datos disponibles?
  - Objetivos
- Descripción de las fuentes de datos:
  - Orígenes
  - Formatos
- Realización del Data Management Plan
  - Asumiendo proyecto de 0 (Como si fueran datos generados)
  - Basado en la plantilla para proyectos de la Comisión Europea. Subconjunto en últimas slides.
  - Tener claro cuál sería el objetivo científico/técnico del proyecto

# Trabajo en Grupo

- Diagrama Gantt/Distribución temporal (Teniendo en cuenta el Ciclo de vida de los datos)
- Curar/Limpiar datos. Explicar técnicas utilizadas. Si el dataset está curado ya, plantear qué técnicas se han podido utilizar
- Crear metadatos basados en Dublin Core, pero extendido (estándar? DataCite? uno ad-hoc?).
  - Incluir lista de parámetros utilizados, detalles del formato de fichero (suficientes para automatizar el procesado).
  - Metadatos del contexto (proyecto, creador, etc.)
  - Términos Dublin Core posibles
- Plan de preservación
- Análisis de los datos (Un par de gráficas diferentes), conclusiones.
- Entregar: Informe con la descripción general, el diagrama temporal, detalles curación, plan de preservación y análisis. DMP. Documento metadatos.

# Temas anteriores

- Datos de localizaciones de casas de apuestas y renta media de las zonas
- Picos de consumo eléctrico y desempleo de un país
- Bienestar/Salud mental en función del clima y horas de luz solar
- Datos ambientales - cultivos.
- Vegetación del terreno con la calidad de aire
- Datos de enfermos por cáncer, con la presencia y actividad de fábricas en distintas zonas de España
- Impacto del turismo en el precio del suelo urbano
- Distancia recorrida por vehículos y contaminantes en Europa
- Relación entre PIB y casos de Cáncer en el mundo

# Posibles temas

- Calidad de vida urbana y contaminación (Eurostat, OpenAQ)
- Temperatura del mar y desastres naturales
- Movilidad sostenible y meteorología
- Accesibilidad sanitaria y problemas de salud
- Islas de calor urbana, zonas verdes y calidad de vida
- Obesidad y alimentación (productos, acceso a mercados, etc.)
- Clima y accidentes de tráfico
- Relación entre artículos y datasets en eventos extremos
- ...

## Resumen de datos / Descripción general del proyecto & datos

- Tipo de datos a generar o usar (crudos, procesados, modelos, resultados, metadatos, etc.).
- Volumen estimado / tamaño aproximado de los datos.
- Origen de los datos (nuevos, reutilizados, de fuentes externas) y finalidad.
- Potencial de reutilización fuera del proyecto.

## Principios FAIR y compartición / acceso a los datos

- Repositorio(s) donde se depositarán los datos, posibles identificadores persistentes (DOI, URIs), esquema o estándar de metadatos utilizado.
- Políticas de acceso: abierto por defecto, salvo excepciones (“as open as possible, as closed as necessary”).
- Licencias para la reutilización (por ejemplo, licencias tipo CC-BY, CC0, u otras según normativa del proyecto).

## Gestión de curación, preservación y almacenamiento

- Dónde se almacenarán los datos durante el proyecto (servidores, repositorios, backups, copias de seguridad) y su preservación a largo plazo.
- Persona(s) o entidad(s) responsables de la gestión de datos / curación de metadatos.
- Estimación de los costes y recursos necesarios para almacenamiento, mantenimiento, backups, gestión de datos.

## Documentación, metadatos y estándares (interoperabilidad)

- Estructura de metadatos: qué estándares, vocabularios o esquemas se usarán (formatos de archivo, metadatos describiendo contexto, métodos, variables, etc.).
- Convenciones de nomenclatura, versionado de datos, trazabilidad, registro de cambios y documentación del “workflow” de procesamiento.
- Plan de control de calidad — cómo se verificará validez, integridad y consistencia de los datos.

# DMP

## Aspectos éticos, legales y de privacidad (si aplica)

- Protección de datos personales o sensibles, cumplimiento de normativas (privacidad, GDPR, etc.).
- Derechos de propiedad intelectual, licencias, permisos de reutilización, restricciones de acceso si las hubiese.

## Política de acceso y difusión / reutilización

- Cuándo y cómo se hará público el acceso a los datos (inmediato, tras embargo, tras publicación, etc.).