



Anexo_3. Orientaciones para la identificación de conjunto de datos por área de conocimiento según OCDE

Introducción

En el contexto de la ciencia abierta y la creciente importancia de los datos como productos de investigación, la generación, documentación y publicación de datasets se ha convertido en una actividad científica relevante y reconocida. Esta guía tiene como objetivo orientar a los investigadores de todas las disciplinas en la identificación de posibles datasets derivados de su trabajo, utilizando como referencia las cinco grandes áreas de conocimiento de la OCDE (clasificación Frascati).

Cada sección incluye:

- - Una breve descripción del área.
- - Tipos de datos comunes en esa área.
- - Ejemplos concretos de datasets.
- - Potenciales variables y dimensiones.
- - Casos de uso o reutilización en otras disciplinas.

Ciencias Naturales

Comprende disciplinas como física, química, ciencias de la Tierra, astronomía, matemáticas y ciencias del medio ambiente.

Tipos comunes de datasets

Datos experimentales de laboratorio, mediciones físicas o químicas, datos satelitales o geoespaciales, series temporales ambientales, modelos numéricos o simulaciones.

Ejemplos de datasets

- - Base de datos de calidad del aire
- - Series temporales sísmicas
- - Espectros químicos
- - Simulaciones climáticas
- - Catálogo de observaciones astronómicas

Posibles variables

Tiempo, ubicación geográfica, unidad de medida, sensor, método de medición, condiciones de entorno, etc.

Ingeniería y Tecnología

Incluye las ingenierías aplicadas, tecnologías de la información, nanotecnología, automatización, telecomunicaciones, y más.



Tipos comunes de datasets

Registros de sensores, datos de desempeño de prototipos, logs de sistemas de software, redes eléctricas/logísticas/telecomunicaciones, modelos CAD.

Ejemplos de datasets

- - Datos de sensores IoT
- - Desempeño de algoritmos
- - Mantenimiento predictivo
- - Tráfico de red
- - Diseños CAD

Posibles variables

Tiempos de operación, materiales, tipo de sensor, protocolos, especificaciones técnicas, topología, etc.

Ciencias Médicas y de la Salud

Incluye medicina clínica, salud pública, odontología, farmacología, enfermería, psicología clínica, y biotecnología médica.

Tipos comunes de datasets

Datos de pacientes anonimizados, encuestas de salud, imágenes médicas, registros electrónicos de salud, ensayos clínicos.

Ejemplos de datasets

- - Encuesta de hábitos de salud
- - Resonancias magnéticas
- - Historiales de vacunación
- - Registros clínicos anonimizados
- - Estudios genómicos

Posibles variables

Edad, sexo, región, diagnóstico, tiempo de evolución, marcadores, tratamientos, frecuencia de síntomas, etc.

Ciencias Agrícolas

Agronomía, veterinaria, ciencias del suelo, silvicultura, pesca, producción agropecuaria, agroindustria y biotecnología agrícola.

Tipos comunes de datasets

Registros de cultivos, estudios de suelos, datos meteorológicos para agricultura, productividad animal o vegetal, imágenes de drones/satélites.

Ejemplos de datasets

- - Monitoreo de parcelas
- - Muestreo de suelo



- - Control lechero
- - Fenología de cultivos
- - Eficiencia de prácticas agrícolas

Posibles variables

Tipo de cultivo, ubicación, altitud, prácticas agrícolas, enfermedades, rendimiento, condiciones climáticas, etc.

Ciencias Sociales y Humanidades

Incluye sociología, antropología, historia, derecho, economía, ciencias políticas, educación, lingüística, filosofía, arte, literatura.

Tipos comunes de datasets

Encuestas sociales, transcripciones o textos anotados, datos económicos y censales, corpus lingüísticos, bases de datos históricas o jurídicas.

Ejemplos de datasets

- - Encuesta de percepción ciudadana
- - Corpus histórico
- - Datos educativos
- - Indicadores económicos
- - Estudios jurídicos

Posibles variables

Edad, región, nivel educativo, género, ocupación, tipo de documento, temporalidad, categoría gramatical, etc.

Recomendaciones Generales

- - Identifique el potencial de reutilización: otros investigadores pueden usar sus datos para nuevas hipótesis.
- - Documente bien los metadatos: incluya metodología, diccionario de variables, licencia de uso, limitaciones.
- - Considere aspectos éticos y legales: especialmente en ciencias sociales, médicas y agrícolas.
- - Use estándares y formatos abiertos: CSV, JSON, NetCDF, GeoTIFF, RDF, entre otros.
- - Alinee sus datos con los principios FAIR: que sean Findable, Accessible, Interoperable y Reusable.

Conclusión

Esta guía busca servir como una orientación para que investigadores de cualquier área reconozcan el valor de los datos que producen y se animen a organizarlos y publicarlos como contribuciones científicas. La publicación de datasets de calidad fortalece la colaboración, la transparencia y la reproducibilidad en la ciencia.