

3ª edición de las Jornadas de Cultura libre de la Universidad Rey Juan Carlos



Datos abiertos: Movimiento en expansión

21 marzo 2024

Sonia Castro García- Muñoz
sonia.castro@red.es

¿Qué son los datos abiertos y qué nos pueden aportar?



GOBIERNO
DE ESPAÑA

MINISTERIO
PARA LA TRANSFORMACIÓN DIGITAL
Y DE LA FUNCIÓN PÚBLICA

SECRETARÍA DE ESTADO
DE DIGITALIZACIÓN
E INTELIGENCIA ARTIFICIAL

red.es **Iniciativa**
aporta

datos.gob.es
reutiliza la información pública

Agricultura >>



Salud >>



Actividad industrial >>



Turismo >>



...



Gobierno Abierto >>



¿Por qué los datos abiertos son clave en el mundo “open” ?

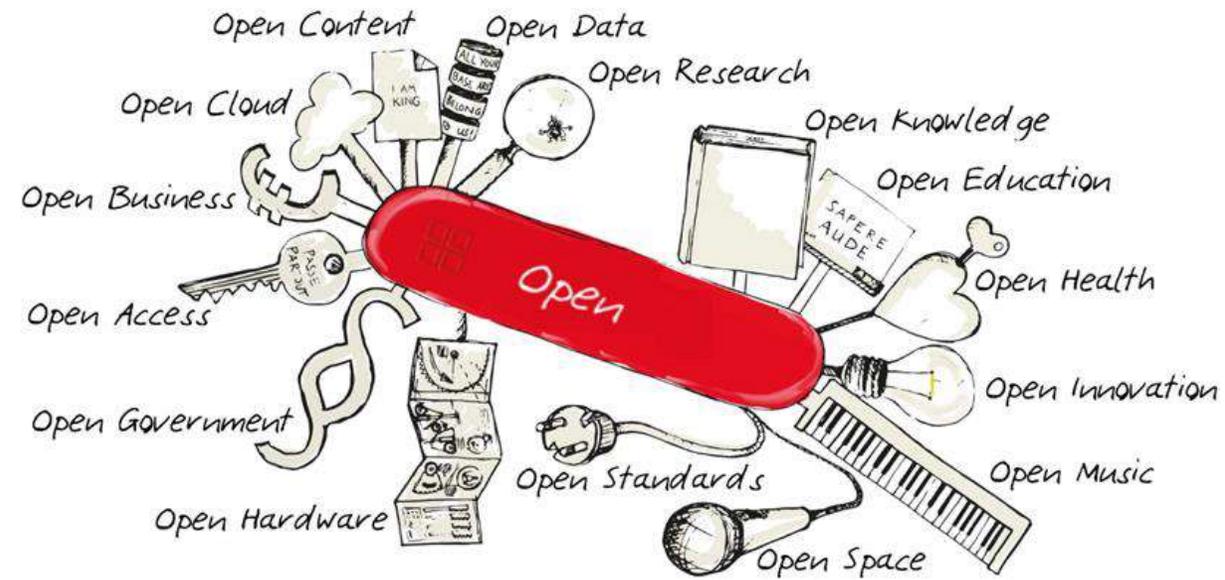
Los humanos aprenden a partir de la experiencia,
las máquinas a partir de los datos

↑ Capacidad de procesamiento ~ *open cloud*

Hiperconectividad ~ *open knowledge*

Tecnologías (sw) *open source*

Código y algoritmos ~ *open code*



Fuente: El open source como base de la competitividad europea, 2021. Festeros.es

DATOS ABIERTOS

¿Qué son los datos abiertos?

- ✓ Son aquellos datos que **cualquiera es libre de utilizar, reutilizar y redistribuir**, con el único límite, en su caso, del requisito de **atribución de su fuente o reconocimiento de su autoría**⁽¹⁾.
- ✓ Deben estar asociados a **condiciones de uso** que deben permitir **explícitamente la transformación, combinación y uso compartido** de los datos con otros, **con fines comerciales o no comerciales**.
- ✓ Deben ser **gratuitos**, aunque existen **excepciones**⁽²⁾.
- ✓ Deben ser **altamente reutilizables**



Fuente: Shutterstock. Créditos: Sachezi

Conjunto de datos → una **colección de datos**, publicada o "curada" (preparada) por un único agente disponible para acceder o descargar **en uno o más formatos (distribuciones)**

(2) Costes marginales en que se incurra para su reproducción, puesta a disposición, difusión y anonimización. De acuerdo con el nuevo RD 24/2021, de 2 de noviembre, organismos del sector público a los que se exige generar ingresos para sufragar sus competencias públicas, pueden cobrar tarifas superiores a los costes marginales por los datos.

(1) [Ley 18/2015, de 9 de julio, por la que se modifica la Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público.](#)

¿Qué características tienen los datos abiertos reutilizables?



Datos **completos**.



Datos **primarios**.



Datos proporcionados **a tiempo**.



Datos **accesibles**.



Datos **procesables por máquina**.



Datos en formatos **no propietarios**.



Datos **asociados a condiciones de uso no restrictivas**.



Datos acompañados de **elementos que faciliten su interpretabilidad** (metadatos, ...)



Fuente imagen: Shutterstock. Créditos: rovodread

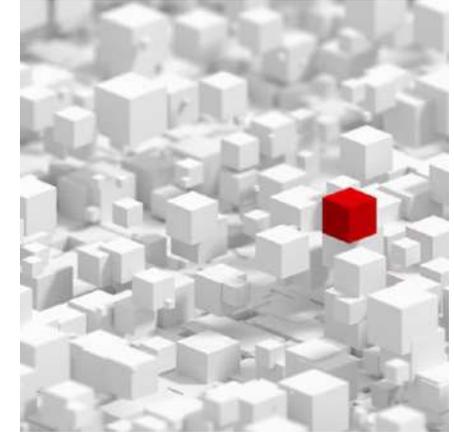
Fuente: [Declaración original de los principios de los datos abiertos](#), 2007

¿Qué son los datos de alto valor?

“Datos del sector público de alto valor (HVDS) son aquellos asociados a **considerables beneficios para la sociedad, el medio ambiente y la economía.**

Se exige que se ofrezcan como DATOS ABIERTOS REUTILIZABLES, cumpliendo estos requisitos:

- se suministren principalmente **de manera gratuita**
- se ofrezcan en formatos que sean **legibles de manera automatizada**
- se **suministren mediante descarga masiva, a través de APIs y, cuando proceda, en tiempo real**



© Imagen de Shutterstock. Créditos: tostphoto

Se corresponden a 6 categorías temáticas, que se traducen en más de 80 temas*.



geoespaciales



observación de la tierra y medio ambiente



meteorología



estadísticas



empresas y propiedad de empresas



movilidad

Posibles nuevas Categorías**

Pérdida climática

Energía

Finanzas

Gobierno y administración pública

Salud

Justicia y Asunto Jurídicos

Datos lingüísticos

*Establecidos por el Reglamento 2023/138, 21 diciembre 2022.

**European Commission, Identification of data themes for the extensions of public sector High-Value Datasets, 2023

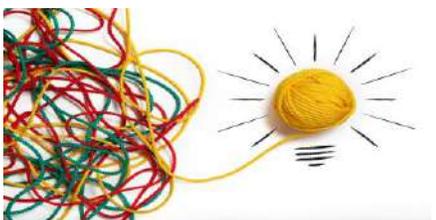
¿Qué logros son posibles gracias al uso de los datos abiertos?



Una nueva **forma de relación** de las administraciones con los ciudadanos fundamentada en la **participación**, la **colaboración** y **transparencia**



La **mejora** y al aumento de la **calidad de los servicios públicos** gracias al uso de los datos que pueden realizar las propias administraciones que producen los datos



La **creación de nuevos productos** y el **ajuste de servicios ya existentes** por parte de las empresas, los investigadores o cualquier colectivo con espíritu innovador

¿Cuál es la diferencia entre gobierno abierto y RISP?



Gobierno abierto: tiene como objetivo que los ciudadanos colaboren en la creación y la mejora de los servicios públicos y en el robustecimiento de la transparencia y la rendición de cuentas.

Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno.

- Organización y Empleo Público
- Altos Cargos
- Planificación y Estadística
- Normativa y Otras Disposiciones
- Contratos, Convenios y Subvenciones
- Información Económico-Presupuestaria



Reutilización de la información pública (RISP): uso de datos que obran en poder de las AA.PP., por personas físicas o jurídicas, con fines comerciales o no comerciales.



Fomenta la **creación de nuevas oportunidades de negocio** y el **ajuste de los servicios existentes**

Ley 37/2007 de 16 de noviembre, sobre reutilización de la información del sector público, a la luz del RD Ley 24/2021, de 2 de noviembre.



Contribuye a la **innovación**, al **ahorro** y al aumento de la **calidad de los servicios públicos**, y a su **revalorización**

¿Cuáles son las obligaciones de las AA.PP. en datos abiertos?

PRINCIPALES OBLIGACIONES DE LA LEY 37/2007 DE 16 DE NOVIEMBRE

Contenidos clave de la Ley 37/2007 sobre reutilización de la información del sector público, tras la incorporación de las previsiones de la Directiva (UE) 2019/1024 relativa a los datos abiertos y la reutilización de la información del sector público.



OBJETIVO:

Autorización generalizada de reutilización de los datos del sector público aplicando el principio de datos abiertos desde el diseño y por defecto.

ÁMBITO DE APLICACIÓN

- Datos producidos por la Administración General del Estado, las Comunidades Autónomas y entidades que integran la Administración Local.
- Datos generados por organismos y entidades del sector público institucional, incluidas bibliotecas, museos y archivos.
- Datos y servicios referidos en la **Directiva INSPIRE**.
- Datos de investigación financiada con fondos públicos.
- Datos de determinadas sociedades mercantiles.



01 >>> OFRECER LOS DATOS:



Formatos abiertos y legibles por máquinas



Con el máximo nivel de precisión y desagregación posible



Accesibles



Fáciles de localizar y reutilizables



Accesibles mediante descarga masiva



A través de interfaces de programación de aplicaciones (API)



En tiempo real para los datos dinámicos



Publicados conjuntamente con sus metadatos

02 >>>

IMPULSAR LA APERTURA DE DATOS DE ALTO VALOR

Focalizar esfuerzos en abrir datos de alto valor establecidos a nivel europeo y a nivel nacional.

¿Qué son los datos de alto valor (HVDs)?

Son aquellos cuyo uso está asociado a considerables beneficios para la sociedad, el medio ambiente y la economía, en particular debido a su idoneidad para la creación de servicios de valor añadido.

CATEGORÍAS



Requisitos legales y técnicos obligatorios de los HVDs



- Uso de licencia de datos abiertos (CC BY 4.0 o equivalente)
- Disponibilidad de documentación pública sobre los datos
- Formatos legibles por máquinas
- Disponibilidad vía APIs y descargables de forma masiva
- Gratuidad
- Con una descripción exhaustiva mediante metadatos

03 >>> DISPONER DE UN CATÁLOGO DE DATOS

Todas las AA.PP. dispondrán de catálogos propios que interoperarán con el Catálogo Nacional (datos.gob.es) siguiendo la NTI-RISP.

04 >>> CONTRIBUIR AL CATÁLOGO NACIONAL

Las AA.PP. deben contribuir a su enriquecimiento mediante procesos de federación con datos.gob.es

05 >>> OFRECER UN ACCESO GRATUITO

Excepcionalmente pueden aplicarse costes marginales de reproducción, puesta a disposición, difusión o anonimización.

06 >>> FOMENTAR LA REUTILIZACIÓN

Esta no estará sujeta a condiciones a menos que estas sean objetivas, proporcionadas, no discriminatorias y estén justificadas por un objetivo de interés público. Se pueden fijar utilizando licencias-tipo.

07 >>> PONER EN MARCHA LA UNIDAD RESPONSABLE DE INFORMACIÓN:

Todas las AAPP deben disponer de una unidad encargada de coordinar las actividades, resolver solicitudes, promover que los datos se ofrezcan con la calidad precisa y llevar a cabo las actuaciones que fomenten su uso.

Fuente: Las claves de la Ley sobre reutilización de la información del sector público en España



GOBIERNO DE ESPAÑA

MINISTERIO PARA LA TRANSFORMACIÓN DIGITAL Y DE LA FUNCIÓN PÚBLICA

SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

red.es **iniciativa** **aporta**

datos.gob.es
reutiliza la información pública

¿Cuál es el impacto real de los datos abiertos?

«Los datos abiertos suponen el 12% del valor de la economía del dato en Europa⁽¹⁾»



IMPACTO SOCIAL

- **54 - 202 mil vidas salvadas** por una respuesta de emergencia más rápida.
- **27 millones de horas ahorradas** en el transporte público.
- **5,8 Mtep* ahorrados** por la reducción del consumo de energía en los hogares.
- Mejora de los **servicios lingüísticos** con datos abiertos, aumentando la **traducción automática**.
*millones de toneladas equivalentes de petróleo

AHORRO DE COSTES

- **312 - 400 mil euros ahorrados en SALUD** debido a la rapidez de los primeros auxilios por parte de los transeúntes.
- **13.700 - 20.000 millones de euros ahorrados en COSTES LABORALES** al reducir el tiempo dedicado al transporte.
- **79.600 millones de euros ahorrados** debido a una mayor producción de **ENERGÍA SOLAR**.
- **1.100 millones de euros ahorrados** por la reducción de los costos de **traducción en el SECTOR PÚBLICO**.

«Los conjuntos de datos abiertos son vitales para acelerar la innovación en *machine learning*. El 78,3% de la investigación en este sector se basa en open data⁽²⁾»

(1) El impacto económico de los datos abiertos: oportunidades de creación de valor en Europa, 2020

(2) Data engineering for everyone, 2021

¿Qué aplicaciones útiles podemos encontrar? (1/5)

Soy una persona con una discapacidad y, a menudo, me cuesta encontrar **una plaza de parking accesibles cercana al destino al que me dirijo**



Nacho

¡Solución!



La empresa Park4Dis ha desarrollado una aplicación a partir de datos abiertos ofrecidos por más de 200 ayuntamientos, y enriquecidos gracias a la colaboración ciudadana que permite al usuario encontrar plazas reservadas para u otros estacionamientos permitidos

¿Qué aplicaciones útiles podemos encontrar? (2/5)

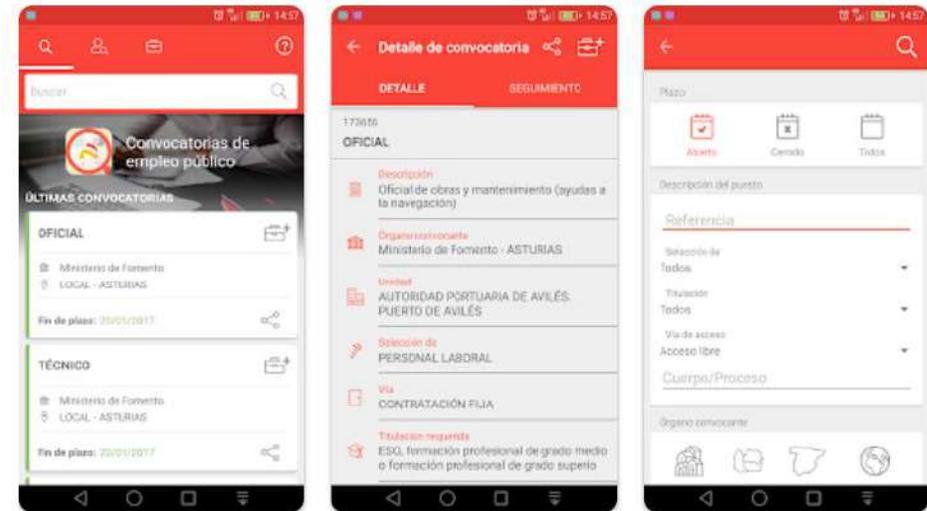
Acabo de terminar mis estudios y estoy interesada en conocer oferta de empleo público existente en España



Sandra

¡Solución!

A partir de la información de los boletines oficiales y de las distintas webs de organismos públicos (datos abiertos), el Ministerio para la Transformación Digital y de la Función Pública ha desarrollado una aplicación que funciona a modo de base de datos única de las convocatorias de empleo de las diferentes Administraciones Públicas



¿Qué aplicaciones útiles podemos encontrar? (3/5)

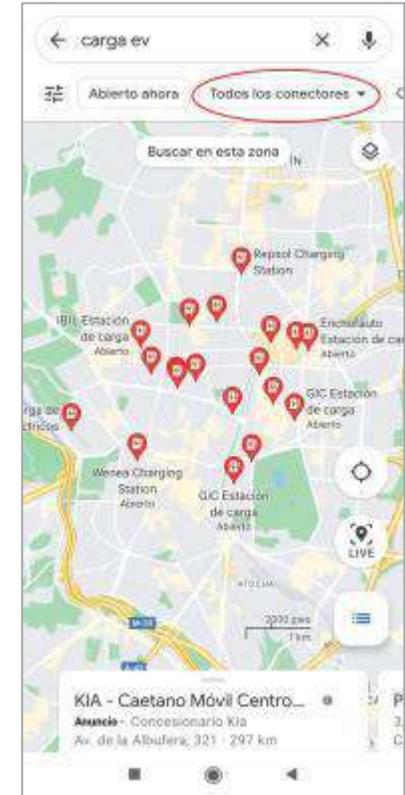


Berta

Me decanté por un coche eléctrico. Y ahora, ¿dónde encuentro puntos de recargar?

¡Solución!

A partir de datos abiertos ofrecidos por la DGT sobre Puntos de recarga eléctrica para vehículos con actualización diaria, Google Maps ha desarrollado una capa en la que se incorpora esta información



¿Qué aplicaciones útiles podemos encontrar? (4/5)



Irene

Estoy pensando en plantear a mi familia la posibilidad de montar placas solares en mi casa de la sierra y aprovechar el excedente en calentar el agua de la piscina

¡Solución!



A partir de datos abiertos ofrecidos por el Catastro, de datos LiDAR del IGN y de datos de radiación solar de la Comisión Europea, el CNIG ha elaborado un visualizador del potencial solar de los edificios en España

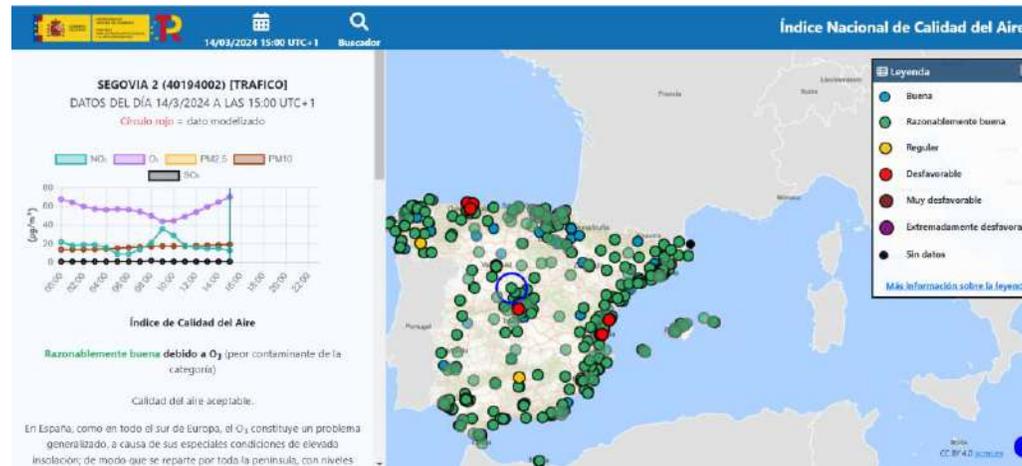
¿Qué aplicaciones útiles podemos encontrar? (5/5)

Tengo problemas severos de asma y soy propenso a padecer enfermedades respiratorias. La calidad del aire es para mí una prioridad a la hora de elegir mi destino vacacional



Felipe

¡Solución!



A partir de datos abiertos ofrecidos por el Ministerio de Transición ecológica ligados a la calidad del aire: dióxido de azufre, dióxido de carbono, tenemos de una aplicación que nos permite conocer en tiempo real la calidad del aire de diferentes ubicaciones del territorio nacional

¿Quiénes son los principales productores de datos abiertos?

Sector público



80.702 conjuntos de datos
(490.364 ficheros de datos)

250 organismos públicos



1.699.470 conjuntos de datos

35 países

Organizaciones internacionales



NASA Open Data Portal



UNITED NATIONS
INDUSTRIAL DEVELOPMENT ORGANIZATION

UNIDO Statistics Data Portal



re3data



Copernicus



World Bank Open Data



OCDE Data



Climate Data Online

¿Cómo podemos contribuir a este movimiento?

La ciencia ciudadana invita al **público general a participar en actividades científicas** y busca **fomentar la contribución activa de los ciudadanos** a la investigación a través de su **esfuerzo intelectual**, sus **conocimientos** y su **tiempo**.

Mosquito Alert

Controlar expansión del mosquito que transmite el Dengue



Análisis de imágenes >> Monitorización de zonas y **banco de datos**

Plant-tes

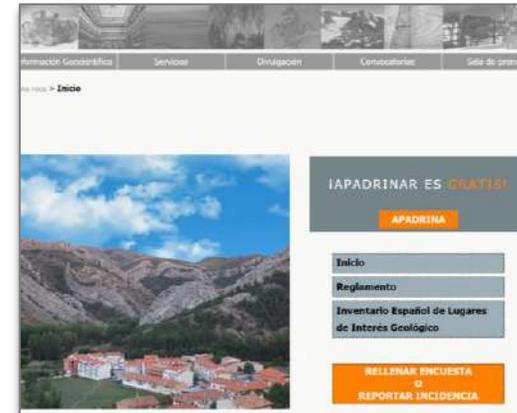
Mejorar la calidad de vida de las personas con alergia



Recolección datos >> **mapa de estado fenológico** de las plantas del entorno

Apadrina una roca

Conservación del patrimonio geológico español



Visitar y documentar un lugar de interés geológico >> **Inventario Español de Lugares de Interés Geológico**

Mercè

Mejorar la habitabilidad de las ciudades



Elección entre imágenes - **Banco de conocimiento**+IA=Toma decisiones

“Si quieres viajar rápido, hazlo solo. Si quieres viajar lejos, hazlo acompañado”

Proverbio africano

La Iniciativa Aporta: la suma de todos



GOBIERNO
DE ESPAÑA

MINISTERIO
PARA LA TRANSFORMACIÓN DIGITAL
Y DE LA FUNCIÓN PÚBLICA

SECRETARÍA DE ESTADO
DE DIGITALIZACIÓN
E INTELIGENCIA ARTIFICIAL

red.es **Iniciativa
aporta**

datos.gob.es
reutiliza la información pública

¿Qué es la Iniciativa Aporta?

Aporta es iniciativa nacional de **datos abiertos**

Se lanza en 2009 **con la misión el impulsar la apertura de datos** por parte del sector público y la **promoción del desarrollo de servicios avanzados** basados en ellos.

Se desarrolla por parte de la **Entidad pública Red.es**, en el marco de un convenio de colaboración con la Secretaria General de Administración Digital



¿Qué es datos.gob.es?

Datos.gob.es es el espacio virtual de la Iniciativa Aporta

...**punto de encuentro** entre las **administraciones**, las **empresas** y los **ciudadanos** interesados en los datos abiertos

... punto de acceso al **Catálogo Nacional** de datos abiertos
(+**80.000** conjuntos de datos y +**470.000** ficheros de datos (distribuciones))

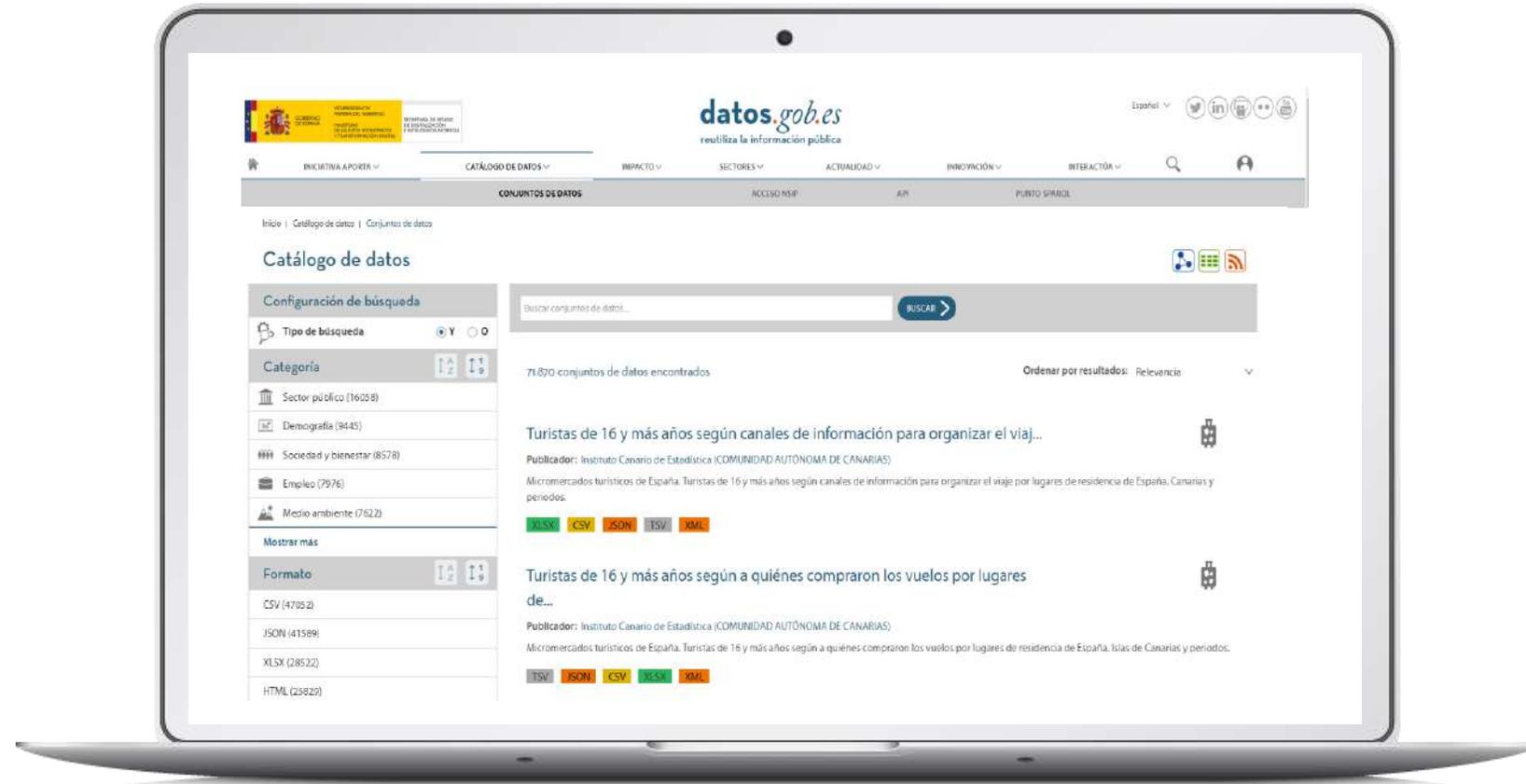
... espacio a través del cual **informar y estar informado de novedades** ligadas a datos

... lugar donde encontrar **ejemplos relevantes** de reutilización



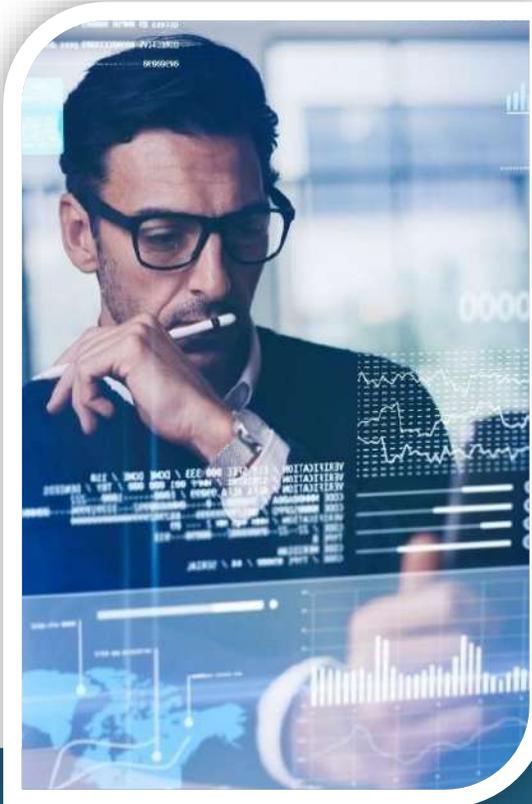
¿Qué es el Catálogo Nacional de datos abiertos?

El Catálogo Nacional es un **espacio unificado de acceso a los datos abiertos** que las AA.PP. españolas ponen a disposición para su reutilización.



El Catálogo Nacional alberga los **metadatos** que describen los **datos** que quieren dar a conocer las **AA.PP.**

¿A quién nos dirigimos?



INSTITUCIONES que OFRECEN DATOS para su reutilización Y quieren ESTRECHAR LAZOS con los reutilizadores de esos datos



REUTILIZADORES que BUSCAN DATOS, y que nos cuentan qué han hecho con ellos



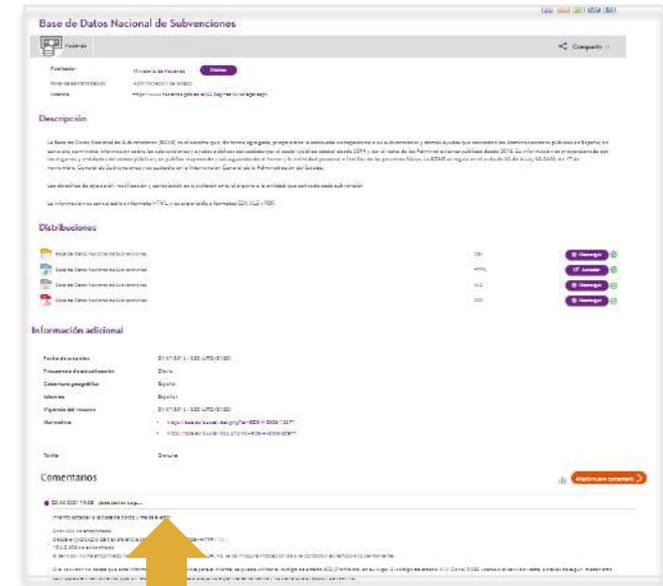
TODOS aquellos que CREEN Y CONTRIBUYEN al movimiento de los datos abiertos

¿Qué te ofrecemos?



¿Quieres localizar datos abiertos de forma sencilla?

Encuentra datos de diversa naturaleza a través de una “ventanilla única”, el Catálogo Nacional



Texto libre:

Buscar conjuntos de datos...

Mecanismos avanzados



<https://datos.gob.es/es/apidata>

<https://datos.gob.es/es/sparql>

Categoría	Nivel de administración	Formato	Frecuencia de actualización	Publicador	Últimos conjuntos de datos publicados	Conjuntos de datos en CSV	Conjuntos de datos de APIs y Servicios web
Sector público (590)	Administración Autónoma (12030)	CSV (14903)	Anual (3696)	Comunidad Autónoma de País Vasco (3842)	2024-09-20 Comercio exterior 2019	2024-09-20 Registros de subvenciones de "Medios de transporte"	2024-09-20 Servicios de apoyo técnico y asistencia al cliente
Medio ambiente (444)	Administración del Estado (780)	XLS (10050)	Mensual (1122)	Gobierno de Aragón (2761)	2024-09-20 Pérdidas comerciales 2020	2024-09-20 Hacia la Ley de Protección Ambiental de los Recursos Acuáticos (LPA)	2024-09-20 Servicios de apoyo técnico y asistencia al cliente
BNI - Sociedad y bienestar (4030)	Administración Local (1943)	JSON (8689)	Diario (932)	Instituto Geológico y Minero de España (2254)	2024-09-20 Pérdidas comerciales 2021	2024-09-20 Pérdidas comerciales 2021	2024-09-20 Servicios de apoyo técnico y asistencia al cliente
Economía (855)	Universidades (223)	HTML (6029)	Trimestral (678)	Centro de Investigaciones Sociológicas (2218)	2024-09-20 Pérdidas comerciales 2022	2024-09-20 Pérdidas comerciales 2022	2024-09-20 Servicios de apoyo técnico y asistencia al cliente
Demografía (5276)		PDF (4227)	Bimensual (535)	Instituto Nacional de Estadística (1881)			
Mostrar más		Mostrar más	Mostrar más	Mostrar más	Ver más	Ver más	Ver más

¿Te cuesta encontrar los datos que buscas?



Canal de solicitud de datos

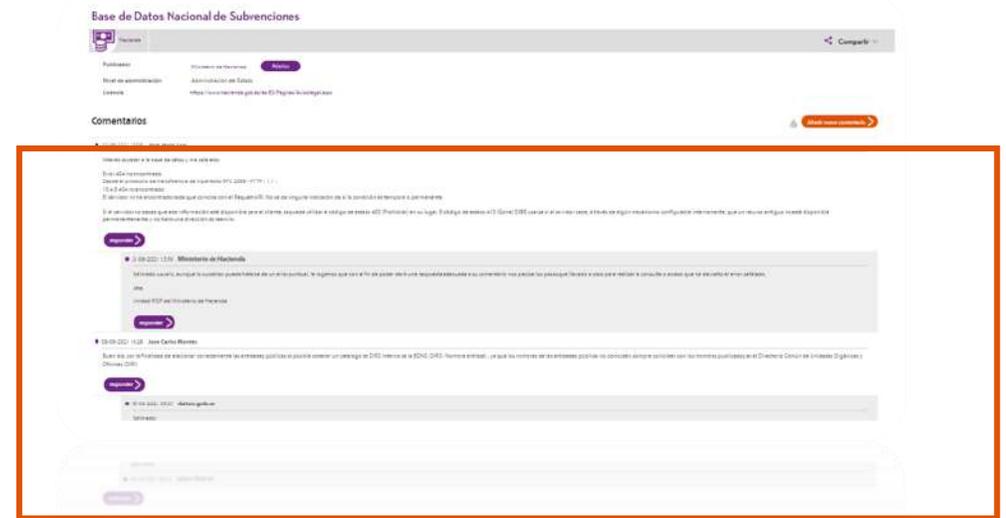
Disponibilidad de datos

En esta sección encontraras consultas que han realizado los usuarios sobre datos que aún no están disponibles en el [Catálogo de datos](#). Las consultas recibidas son asignadas al organismo competente en la materia y el organismo será el responsable de dar respuesta a la consulta e indicar el estado de tramitación en el que se encuentra. Para adherirse a cualquiera de las solicitudes, es necesario acceder al detalle de la solicitud.

Importante: Este canal tiene por objetivo dar a conocer las solicitudes de datos de los ciudadanos a la Administración para ayudar a avanzar en las iniciativas de datos abiertos puestas en marcha. En ningún caso sustituirá el procedimiento de tramitación de solicitudes de reutilización previsto en el artículo 10 de la Ley 37/2007, que estará a disposición de los usuarios a través de la sede electrónica de cada órgano administrativo.

A screenshot of the 'Disponibilidad de datos' web interface. It features a search bar with a 'BUSCAR' button, a category list on the left (Sector público, Hacienda, Economía, Medio ambiente, Cultura y ocio, Transporte), and a flowchart showing the status of requests: Recibida, Asignada (238), En Estudio (10), Programada (2), Publicada (76), No Viable (22), and Parcialmente publicada (4). Below the flowchart, it shows '352 disponibilidades de datos encontrados' and a sorting option 'Ordenar por: Fecha descendente'. The first result is 'DIRECTORIO MUNICIPIOS ESPAÑOLES'.

Área de comentarios

A screenshot of the 'Base de Datos Nacional de Subrecciones' web interface. It shows a search bar, a 'Comentarios' section with a 'Añadir nuevo comentario' button, and a list of comments. A red box highlights a comment from 'Ministerio de Hacienda' dated 20-09-2021, which discusses the availability of data for the 'Directorio Municipios Españoles'.

¿Quieres aprender y estar al día?

Noticias

Blog de innovación

Entrevistas

Agenda de eventos

Boletín

Twitter

LinkedIn

28-06-2022
Realidad aumentada para la visualización de datos

23-06-2022
Probamos spaCy: mucho más que una librería para crear proyectos reales de procesamiento del lenguaje natural

21-06-2022
La reutilización en la futura regulación del Espacio Europeo de datos de Salud

25-05-2022
10 librerías populares de análisis de datos y Machine Learning

24-05-2022
Normas Técnicas para alcanzar la Calidad del Dato

11-05-2022
Recomendaciones para facilitar el uso de datos privados para el interés público en la Unión Europea

24-06-2022
El nuevo Portal de Datos Abiertos de la Diputación Provincial de Castellón

17-06-2022
10 cursos para ampliar tu conocimiento sobre datos este verano

15-06-2022
Novedades del ecosistema de datos abiertos (primavera 22)

datos.gob.es - Oficina del Dato
12 mil Tweets

datos.gob.es
@datosgob Te sigue

Novedades y tendencias relacionadas con la innovación basada en datos y en open data, ofrecidas por la Oficina del Dato (bit.ly/3GvrtXGn) y datos.gob.es

España Se unió en mayo de 2009

2,953 Siguiendo 16,6 mil Seguidores

Iniciativa Open Data, Red Transparencia y Participación - RED FEMP y 150 más de las cuentas que sigues siguen a este usuario

datos.gob.es - Oficina del Dato (seleccionar)

datos.gob.es - Oficina del Dato
Novedades y tendencias relacionadas con la innovación basada en datos y en open data
Government Administration - Madrid, Spain - 4,888 Members

Analítica
datos.gob.es

107 Unique visitors

10,200 seguidores

¿Quieres seguir a datos.gob.es? Sigue a datos.gob.es and read by inviting connections to follow your Page



@datosgob (+20.800 seguidores)



+10.200 seguidores

¿Quieres practicar en ciencia de datos?

SOLUCIONES DESARROLLADAS EN BASE A MACHINE LEARNING

Digital Surface Model - Modelo de inundación

```

class
  band : 1 (of : 4 bands)
  dimensions : 9000, 12500, 118500000 (rows, cols, cells)
  resolution : 1.102401, 1.102429 (x, y)
  extent : 282726, 296586, 4815427, 4829978 (xmin, xmax, ymin, ymax)
  crs : epsg:utm_zone=38_wellps=SR586_units= m_defn
  source : mapserver25_epsg586_0014-4_C06.tif
  names : mapserver25_epsg586_wm6_4_C06
    
```

VISUALIZACIONES PASO A PASO

1. Introducción
2. Objetivos
3. Recursos
 - 3.1. Conjunto de datos
 - 3.2. Herramientas
4. Tratamiento de datos
5. Creación de la visualización
6. Conclusiones

¿Cuántos accidentes ocurren en la ciudad de Madrid?

```

class
  band : 1 (of : 4 bands)
  dimensions : 9000, 12500, 118500000 (rows, cols, cells)
  resolution : 1.102401, 1.102429 (x, y)
  extent : 282726, 296586, 4815427, 4829978 (xmin, xmax, ymin, ymax)
  crs : epsg:utm_zone=38_wellps=SR586_units= m_defn
  source : mapserver25_epsg586_0014-4_C06.tif
  names : mapserver25_epsg586_wm6_4_C06
  values : 0, 355 (min, max)
    
```

```

class
  band : 1 (of : 1 bands)
  dimensions : 2863, 4472, 9221758 (rows, cols, cells)
  resolution : 1, 1 (x, y)
  extent : 283068, 287688, 4822814, 4824877 (xmin, xmax, ymin, ymax)
  crs : epsg:utm_zone=38_wellps=SR586_units= m_defn
  source : S5018_AST_S7_1_Q10_CALAD06_1C.tif
  names : S5018_AST_S7_1_Q10_CALAD06_1C
    
```

Este uso que tenemos las tablas con los tres campos temáticos (nombre de departamento de empresa, gasto por departamento y número de accidentes de empresa), creamos una nueva tabla que incluya los valores que interesan de cada una de ellas.
 En primer lugar, registramos por separado los campos de departamentos de empresa ("departamento") y accidentes, con sus respectivos departamentos de empresa ("departamento") para facilitar la visualización por provincia de la tabla de gastos en accidentes por departamento ("gasto_departamento"). En este paso, seleccionamos de la tabla los valores que interesan de los tres campos de datos.



¿Quieres poner a prueba tus conocimientos?

Los **Desafíos Aporta** son **concursos de idea+prototipo** que, utilizando datos abiertos, puedan suponer **mejoras en un sector de actividad**



I Desafío Aporta
El valor del dato para la administración

El valor del dato para la administración

Concurso de ideas

45 propuestas recibidas

10 propuestas seleccionadas

Finalistas Fase I Desafío Aporta

El jurado ha seleccionado

II Desafío Aporta
El valor del dato para el sector agroalimentario, forestal y rural

El valor del dato para el sector agroalimentario, forestal y rural

CONCURSO DE IDEAS

10 PROYECTOS RECIBIDOS

FINALISTAS FASE I

DESAFÍO APORTA 2019

40 ideas presentadas para mejorar la eficiencia en los sectores agrícola, forestal y del medio rural.

3 propuestas ganadoras: Optimiza, Chopperapp y...

III Desafío Aporta
El valor del dato en la educación digital

El valor del dato en la educación digital

CONCURSO DE IDEAS

10 PROYECTOS RECIBIDOS

FINALISTAS FASE I

DESAFÍO APORTA 2019

El jurado ha seleccionado

IV Desafío Aporta
El valor del dato para la salud y el bienestar de los ciudadanos

El valor del dato para la salud y el bienestar de los ciudadanos

CONCURSO DE IDEAS

10 PROYECTOS RECIBIDOS

FINALISTAS FASE I

DESAFÍO APORTA 2019

El jurado ha seleccionado

**Te presentamos una
selección de
contenidos**



GOBIERNO
DE ESPAÑA

MINISTERIO
PARA LA TRANSFORMACIÓN DIGITAL
Y DE LA FUNCIÓN PÚBLICA

SECRETARÍA DE ESTADO
DE DIGITALIZACIÓN
E INTELIGENCIA ARTIFICIAL

red.es *Iniciativa*
aporta

datos.gob.es
reutiliza la información pública



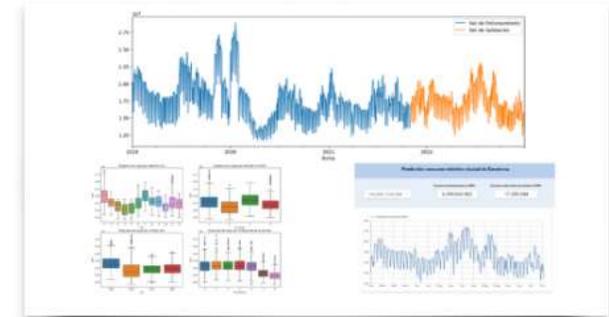
El principal objetivo de este ejercicio es mostrar cómo realizar, de una forma didáctica, un **análisis predictivo de series temporales partiendo de datos abiertos** sobre el consumo de electricidad en la ciudad de Barcelona. Para ello realizaremos un análisis exploratorio de los datos, definiremos y validaremos el modelo predictivo, para por último generar las predicciones junto a sus gráficas y visualizaciones correspondientes.

Los análisis predictivos de series temporales son técnicas estadísticas y de aprendizaje automático que se utilizan para prever valores futuros en conjuntos de datos que se recopilan a lo largo del tiempo. Estas predicciones se basan en patrones y tendencias históricas identificadas en la serie temporal, siendo su objetivo principal anticipar cambios y eventos en función de datos pasados.

El conjunto de datos abiertos inicial consta de registros **desde el año 2019 hasta el año 2022 ambos inclusive**, por otra parte, las **predicciones las realizaremos para el año 2023**, del cual no tenemos datos reales.

Una vez realizado el análisis, podremos contestar a preguntas como las que se plantean a continuación:

- ¿Cuál es la predicción futura de consumo eléctrico?
- ¿Cómo de preciso ha sido el modelo con la predicción de datos ya conocidos?
- ¿Qué días tendrán un consumo máximo y mínimo según las predicciones futuras?
- ¿Qué meses tendrán un consumo medio máximo y mínimo según las predicciones futuras?



Laboratorio-de-Datos / Visualizaciones / Análisis_predictivo_consumo_BCN /

Administratosgobes Update readme.md (1100019 - 5 months ago) History

This branch is 258 commits ahead of, 147 commits behind Administratosgobes/Laboratorio-de-Datos:main

Name	Last commit message	Last commit date
..		
Codigo	Add files via upload	5 months ago
Datos_origen	Update readme.md	5 months ago
Datos_preprocesados	Update readme.md	5 months ago
Imagenes	Add files via upload	5 months ago
readme.md	Update readme.md	5 months ago

```
Analisis_predictivo_consumo_BCN.ipynb
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Los cambios no se guardarán
+ Código + Texto Copiar en Drive Conectar
El principal objetivo de este ejercicio es mostrar cómo realizar, de una forma didáctica, un análisis predictivo de series temporales partiendo de datos abiertos sobre el consumo de electricidad en la ciudad de Barcelona. Para ello realizaremos un análisis exploratorio de los datos, definiremos y validaremos el modelo predictivo para, por último, generar las predicciones junto a sus gráficas y visualizaciones correspondientes.
1. Preparación de los datos
Lo primero que debemos hacer es importar y cargar las librerías que utilizaremos en el ejercicio.
• pandas es una librería de Python especializada en el manejo y análisis de estructuras de datos.
• matplotlib es una librería de Python especializada en la creación de visualizaciones
• seaborn es una librería de Python basada en matplotlib especializada en la creación de visualizaciones.
• git es una librería de Python que permite interactuar con el sistema de control de versiones Github.
• xgboost es una librería de Python para la implementación de algoritmos de machine learning en el marco de "Gradient Boosting"
[ ] # Importación de las librerías a utilizar en el ejercicio
!pip install gitpython

import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import xgboost as xgb
import git
import warnings
warnings.filterwarnings('ignore')

Collecting gitpython
  Downloading GitPython-3.1.40-py3-none-any.whl (190 kB)
  190.6/190.6 kB 3.0 MB/s eta 0:00:00
Collecting gitdb<5, >=4.0.1 (from gitpython)
  Downloading gitdb-4.0.11-py3-none-any.whl (62 kB)
  62.7/62.7 kB 7.5 MB/s eta 0:00:00
```

REGLAMENTO EUROPEO DE GOBERNANZA DE DATOS (DATA GOVERNANCE ACT O DGA)



¿QUÉ ES?

Es un instrumento horizontal, es decir aplicable a los diferentes ámbitos sectoriales, para **regular la reutilización de datos e impulsar su intercambio** bajo los principios y valores de la Unión Europea (UE).

Reglamento (UE) 2022/868 relativo a la gobernanza europea de datos (DGA)

Regula tres aspectos concretos de la compartición de datos:

1. El intercambio seguro y voluntario de conjuntos de **datos que están bajo el poder de organismos públicos sobre los que concurren derechos de terceros** que dificultan su reutilización (*protección de los datos de carácter personal, propiedad intelectual, confidencialidad comercial, etc.*).
2. Los **servicios de intermediación de datos** entre los titulares de datos y sus potenciales usuarios.
3. La **cesión altruista** de datos por parte de empresas, ciudadanos, etc.

Se trata de un pilar clave de la [estrategia europea en materia de datos](#), que complementa a otras normativas:

Reglamento (UE) 2023/2854 sobre normas armonizadas para un acceso justo a los datos y su utilización (Data Act)

Impulsa **reglas armonizadas relativas al acceso y uso equitativo de los datos** en el marco de la Estrategia Europea.

Directiva (UE) 2019/1024 relativa a los datos abiertos y la reutilización de la información del sector público

Establece el marco jurídico para la reutilización de la **información del sector público**, puesta a disposición del público como **datos abiertos**.

Los datos compartidos siempre deberán cumplir el [Reglamento General de Protección de Datos \(RGPD\)](#) - Reglamento (UE) 2016/679.



PRINCIPALES OBJETIVOS DEL REGLAMENTO EUROPEO DE GOBERNANZA DE DATOS

1. Promover un **mercado único y armonizado de los datos**, facilitando su puesta en común entre sectores y países de la UE.
2. Apoyar la **creación y el desarrollo de espacios comunes europeos** de datos en ámbitos estratégicos, con la participación de agentes públicos y privados.
3. Reforzar la **confianza** de ciudadanos y empresas en que sus datos van a ser reutilizados bajo su control, conforme a unos mínimos estándares jurídicos.
4. **Superar los obstáculos técnicos a la reutilización de datos**, favoreciendo la agilidad en su manejo.



LAS CLAVES DE LA DGA

Se instituyen una serie de mecanismos idénticos en toda la Unión con los que impulsar el intercambio de datos:

- Los Estados miembros están obligados a establecer un **punto único de información** sujeta a derechos de terceros donde presentar las solicitudes a resolver en el plazo de dos meses.
- Se establece un **régimen de notificación de los proveedores de servicios de intercambio de datos**, que serán neutrales, transparentes y no discriminatorios.
- Se crean mecanismos para impulsar el **altruismo y facilitar voluntariamente el uso de los datos para el bien común**.
- Se crea un **comité de expertos a nivel europeo** que asesora a la Comisión (*Comité Europeo de Innovación en materia de datos*).

Se adoptan medidas específicas para el sector público, bajo un modelo de gobernanza que refuerza las garantías:

- Los organismos públicos deben **impulsar las medidas técnicas, organizativas y jurídicas** necesarias para garantizar la protección de los datos compartidos.
- Se impone la obligación de que solo se puedan reutilizar aquellos datos sometidos a derechos de terceros que cuenten con un **"pretratamiento"** (*anonimizados, seudonimizados o sin información confidencial*).
- Se contempla que la **reutilización sólo se admita en entornos directamente controlados por el organismo público** si no existiera otra alternativa.
- Se establecen **condiciones y garantías efectivas** para los supuestos en que el tratamiento de los datos sea fuera de la UE.



¿A QUIÉN AFECTA?

Al tratarse de un Reglamento, su **aplicación directa es obligatoria en toda la Unión Europea**. No obstante, los **Estados miembros tienen competencia** para:

- Adoptar las medidas organizativas necesarias.
- Legislar sobre el de acceso a la información del sector público (*el Reglamento no afectará a las normas estatales vigentes en este campo*).



¿CUÁNDO ENTRÓ EN VIGOR?

Entró en vigor el **23 de junio de 2022**.

Tras un periodo de gracia de 15 meses, es **aplicable desde septiembre de 2023**.



¿QUIERES SABER MÁS SOBRE EL REGLAMENTO DE GOBERNANZA DE DATOS?

Haz clic sobre los siguientes enlaces:

- [European Data Governance Act | Shaping Europe's digital future \(europa.eu\)](#)
- [El Reglamento de Gobernanza de Datos explicado](#)
- [La propuesta europea sobre gobernanza de los datos, ¿un paso adelante?](#)
- [La regulación europea en materia de datos frente al resto de una aplicación armonizada que impulsa la compartición de datos](#)
- [La protección de datos personales en el Reglamento de Gobernanza de los Datos \(Data Governance Act\)](#)



GOBIERNO DE ESPAÑA

MINISTERIO PARA LA TRANSFORMACIÓN DIGITAL Y DE LA FUNCIÓN PÚBLICA

SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

red.es **iniciativa** **aporta**

datos.gob.es
reutiliza la información pública

ANÁLISIS EXPLORATORIO DE DATOS (EDA)

Guía Práctica de Introducción al Análisis Exploratorio de Datos

1 ¿QUÉ ES?

Conjunto de técnicas estadísticas dirigidas a explorar, describir y resumir la información que contienen los datos, maximizando su comprensión.

Gracias a ello puedes:

Realizar un análisis descriptivo

Identificar posibles errores

Revelar la presencia de datos atípicos

Comprobar la relación entre las variables



¿POR QUÉ ES IMPORTANTE?

- Las técnicas estadísticas de análisis de datos y el machine learning presuponen el cumplimiento de unas condiciones previas para garantizar la **objetividad e interoperabilidad** de los datos.
- El **EDA** es esencial para garantizar que los resultados de cualquier análisis estadístico sean **consistentes y veraces**.



2 ¿CUÁLES SON LOS PASOS A SEGUIR?

Fuente: Se ha tomado como referencia el libro **R for Data Science de Wickham y Golemund (2017)**



1 ANÁLISIS DESCRIPTIVO

- ¿Qué es?** Síntesis de la información que proporciona el conjunto de datos, extrayendo sus características más representativas.
- ¿Por qué es necesario?** Permite conocer los tipos de datos, descubrir patrones y preparar los datos para futuros análisis.
- Tratamiento:** Aplicar funciones de estadística descriptiva para explorar la estructura del conjunto de datos, examinar los datos y las variables que presenta.

3 DETECCIÓN Y TRATAMIENTO DE DATOS AUSENTES

- ¿Qué es?** Identificar la falta de algunos de los datos en la variable.
- ¿Por qué es necesario?** Los datos ausentes pueden generar problemas a la hora de aplicar técnicas de machine learning, elaborar modelos predictivos, realizar análisis estadísticos o generar representaciones gráficas.
- Tratamiento:** Existen varias maneras de tratar los valores ausentes, como por ejemplo sustituirlos por la media o la mediana, o completar los valores faltantes con el valor anterior o posterior de la columna.

5 ANÁLISIS DE CORRELACIÓN DE VARIABLES

- ¿Qué es?** Analizar la relación entre dos o más variables.
- ¿Por qué es necesario?** Entre otras razones, para descartar posibles variables que aporten información redundante en el conjunto de datos, ocasionando ruido en los análisis.
- Tratamiento:** Calcular los coeficientes de correlación para las variables para detectar coeficientes cercanos a 1 o -1.

2 RE-AJUSTE DE LOS TIPOS DE VARIABLES

- ¿Qué es?** Verificar que las variables se han almacenado con el tipo de valor correspondiente.
- ¿Por qué es necesario?** Una mala codificación de las variables puede influir negativamente en la agrupación de los datos o los resultados de los análisis.
- Tratamiento:** Aplicar la codificación apropiada para cada una de las variables.

4 DETECCIÓN Y TRATAMIENTO DE DATOS ATÍPICOS

- ¿Qué es?** Identificar datos con valores significativamente distintos a los que presenta la variable.
- ¿Por qué es necesario su tratamiento?** Pueden modificar los resultados y restar potencia a los análisis estadísticos o técnicas de machine learning aplicadas.
- Tratamiento:** Disminuir su influencia en análisis posteriores o, en casos muy extremos, eliminarlos del conjunto de datos.

3 ¿QUIERES SABER MÁS?

Descubre la **"Guía Práctica de Introducción al Análisis Exploratorio de Datos"**

Aprende las técnicas anteriores mediante el desarrollo de un **caso práctico**.

Metodología para el ejemplo práctico:

- Conjunto de datos utilizado:** [registro de la calidad del aire en Castilla y León](#).

- Herramientas utilizadas:** código R en el entorno de programación RStudio

datos.gob.es

Cómo afrontar un proyecto de análisis...

Ver más ta... Compartir

FASE I: Importación y limpieza



Ver en YouTube

Guía práctica de introducción al Análisis Exploratorio de Datos



red.es

Iniciativa **aporta**



GOBIERNO DE ESPAÑA

MINISTERIO PARA LA TRANSFORMACIÓN DIGITAL Y DE LA FUNCIÓN PÚBLICA

SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

red.es **Iniciativa aporta**

datos.gob.es
reutiliza la información pública



¿Te anima a participar activamente?

GRACIAS.

¿Quieres conocer más acerca de la Iniciativa Aporta?

Visita datos.gob.es

Twitter: [@datosgob](https://twitter.com/datosgob)

Linkedin: [datos.gob.es](https://www.linkedin.com/company/datos-gob-es)

Suscíbete al [boletín quincenal](#)

Escribe a contacto@datos.gob.es