

¿Qué es un conjunto de datos?

Una serie de datos vinculados entre sí y agrupados dentro de un mismo sistema de información para su potencial reutilización. Los conjuntos de datos o *datasets* son accesibles a través del portal de datos abiertos.

¿Qué son los datos abiertos?

Datos abiertos (open data en inglés) es una iniciativa global, ligada a las políticas de gobierno abierto, que persigue que los datos y la información, especialmente los que posee el sector público, se publiquen de forma abierta, regular y reutilizable para todo el mundo, sin restricciones de acceso, copyright-derechos de autor, patentes u otros mecanismos de control.

Los datos abiertos pueden ser reutilizados y redistribuidos libremente por cualquiera, con la única condición de atribución a su fuente o reconocimiento de su autoría.

Los datos abiertos gozan de la capacidad de ser interoperables, esto es, la capacidad de vincular -o entremezclar- diferentes conjuntos de datos de diversos sistemas y organizaciones, de manera que pueden ser localizados entre las diferentes fuentes de distintos organismos.

La interoperabilidad consiste en establecer una estructura de la información estándar para poder relacionar el contenido y los elementos de los diferentes conjuntos de datos, facilitando así la explotación de diferentes conjuntos de datos independientemente de su procedencia. Esta capacidad de componer y conectar componentes es esencial para construir sistemas grandes y complejos.

Consideraciones del conjunto de datos

En el conjunto de datos hay parámetros e indicadores:

- **Parámetros:**

Son elementos de un sistema que permite clasificarlo y poder evaluar algunas de sus características.

En este conjunto de datos los parámetros son: Operador, servicio, tipo de paquete, trimestre, etc...

- **Indicadores:**

Son las variables cuantitativas (discretas o continuas) que determinan el valor de una serie de parámetros asociados en el universo muestral.

Unos ejemplos de indicadores son: Líneas, empleados, paquetes, etc...

Los datos están tabulados, estructurados y con las siguientes características:

- Fila única de cabecera.
- Registro único por fila.
- Estructura de datos vertical.
- No contiene totales ni agrupaciones.
- Campo de fechas codificado siguiendo el estándar ISO-8601.
- Cada una de las columnas incluye un parámetro que caracteriza el indicador final.
- Si ese dato no dispone de algún parámetro, se presenta como un valor “N/A”. (ejemplo: Las inversiones, en la columna de “tipo de paquete” hay un valor “N/A”, ya que ese dato no dispone de ese corte).

Diccionario de datos

Columna	Descripción
trimestre	Trimestre
pais	País
tipo_de_mercado	Tipo de mercado (minorista o mayorista)
servicio	Servicio
concepto	Concepto (ingresos, paquetes, empleados)
operador	Operador
tipo_de_ingreso	Tipo de Ingreso (Banda ancha minorista, Telefonía móvil, ...)
tipo_de_paquete	Tipo de Paquete (Tel. fija y banda ancha fija, ...)
unidades	Unidades
ingresos	Ingresos €
ingresos_por_operador	Ingresos por Operador €
empleados_por_operador	Empleados por Operador (Núm)
paquetes	Paquetes Unidades

Estructura de datos

		Ingresos totales por mercado	Ingresos totales por operador	Empleados por operador	Paquetes
	Trimestre	X	X	X	X
	País	X	X	X	X
	Tipo de mercado	X	X		
	Servicio	Datos generales	Datos generales	Datos generales	Datos generales
	Concepto	Ingresos	Ingresos	Número de empleados	Paquetes
	Operador		X	X	
	Tipo de Ingreso	X			
	Tipo de Paquete				X
Indicadores	Ingresos (Millones €)	X			
	Ingresos por Operador (Millones €)		X		
	Empleados por Operador (Núm)			X	
	Paquetes (Núm)				X

Ejemplo de uso

En la tabla anterior se indica la estructura de la información y qué parámetros e indicadores se utilizan para cada variable representada en las estadísticas de datos generales trimestrales de Telecomunicaciones.

Por ejemplo, si quisiésemos extraer la información de las “Paquetes”, deberíamos seguir los siguientes pasos:

- Seleccionar el Concepto de “Paquetes”
- Disponemos información en los siguientes parámetros: Año, trimestre, país y tipo de paquete.
- Seleccionar el indicador de “Paquetes (Núm)” para el número total de paquetes.