

ENERO 2026

# DECÁLOGO REUTILIZADOR DE DATOS DEL SECTOR PÚBLICO

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

# CONTENIDO


0. Introducción ..... 4
1. Qué son los datos abiertos ..... 5
2. Datos reutilizables: licencias ..... 7
3. Dónde encontrar los datos ..... 9
4. La calidad de los datos ..... 11
5. Interoperabilidad y estándares ..... 13
6. Persistencia y volatilidad ..... 15
7. Herramientas para el tratamiento de los datos ..... 16
8. Formarse y experimentar con datos ..... 18
9. El valor de los datos ..... 20
10. Uso responsable y ético de los datos ..... 21
Del reutilizador al científico de datos ..... 23



## Contenido elaborado por Carlos Iglesias

### Open data Researcher y consultor, World Wide Web Foundation

Este documento ha sido elaborado en el marco de la Iniciativa Aporta (datos.gob.es), desarrollada por el Ministerio para la Transformación Digital y de la Función Pública a través de la Entidad Pública Empresarial Red.es.

Aviso legal: Esta obra está sujeta a una licencia Atribución 4.0 de Creative Commons (CC BY 4.0). Está permitida su reproducción, distribución, comunicación pública y transformación para generar una obra derivada, sin ninguna restricción, siempre que se cite al titular de los derechos (Ministerio para la Transformación Digital y de la Función Pública a través de la Entidad Pública Empresarial Red.es). La licencia completa se puede consultar en:
https://creativecommons.org/licenses/by/4.0

# 0. INTRODUCCIÓN

En el transcurso de su actividad diaria la Administración produce una gran cantidad y variedad de datos con un alto valor y un gran potencial. El objetivo principal de una iniciativa de datos abiertos es habilitar y facilitar, en la medida de lo posible, la reutilización de esa información que proviene del sector público, de forma que ese valor repercuta finalmente en la sociedad.

Esta guía nace con la intención de aclarar algunos aspectos básicos y facilitar los primeros pasos a todos aquellos que puedan estar interesados en reutilizar dicha información.

Una primera versión del **Decálogo del Reutilizador de Datos del Sector Público** se publicó en 2013 para ofrecer una **guía práctica a todas aquellas personas y organizaciones interesadas en aprovechar el valor de la información pública.** Desde entonces, el ecosistema de datos ha evolucionado de forma significativa: han surgido nuevas tecnologías, se han consolidado estándares internacionales y se han aprobado nuevos marcos regulatorios europeos que influyen directamente en cómo se accede, gestiona y reutiliza la información del sector público. Esta versión actualizada mantiene el espíritu original del Decálogo, pero adapta su contenido a los desafíos y oportunidades actuales.

# 1. QUÉ SON LOS DATOS ABIERTOS

Los datos se consideran un activo estratégico con un enorme potencial para transformar nuestra sociedad en ámbitos tan diversos como la **transparencia,** la **participación,** la **innovación,** el **crecimiento económico** o la generación de **beneficios sociales.** En el contexto actual de la **economía del dato,** los datos constituyen una fuente de riqueza en su sentido más amplio y, por tanto, será una responsabilidad de la Administración el compartir ese valor con la sociedad, siempre que sea posible y seguro.

Desde un punto de vista normativo, los datos abiertos se definirían según la Ley 37/2007 sobre reutilización de la información del sector público como "*aquellos 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.*" Esta definición identifica tres características esenciales:

- **Uso libre,** lo que implica que los datos pueden emplearse para cualquier finalidad, incluida la comercial.
- **Reutilización,** es decir, la posibilidad de transformar, combinar o enriquecer los datos para generar nuevo valor.
- **Redistribución,** que permite compartir los datos originales o derivados con terceros, garantizando así su circulación y reaprovechamiento.

Para la aplicación práctica de estos atributos, las iniciativas de datos abiertos – u *Open Data* en inglés – se apoyan en una serie de principios técnicos y organizativos orientados a garantizar que la información se publique en **formatos abiertos y estándar,** con **metadatos adecuados** y bajo **licencias de uso no restrictivas,** de forma que cualquiera pueda acceder a ella y **reutilizarla fácilmente.** Los metadatos son los atributos (origen, contenido, fecha, formato, responsable, licencia, etc.) que permite identificar, contextualizar, interpretar y reutilizar correctamente un conjunto de **datos.** Su papel es, además, muy relevante en la evolución hacia modelos de datos más avanzados como los **grafos de conocimiento** y las **ontologías**, que nos permiten relacionar información de manera más inteligente y con mayor carga semántica.

Por otro lado, la normativa europea ha introducido también el concepto de **conjuntos de datos de alto valor**, que son aquellos conjuntos de datos que se identifican por su elevado impacto económico y social, y que deben publicarse de forma gratuita, en formatos legibles por máquina y mediante interfaces automatizadas, es decir, en formatos estructurados que puedan ser procesados automáticamente por aplicaciones y servicios digitales.

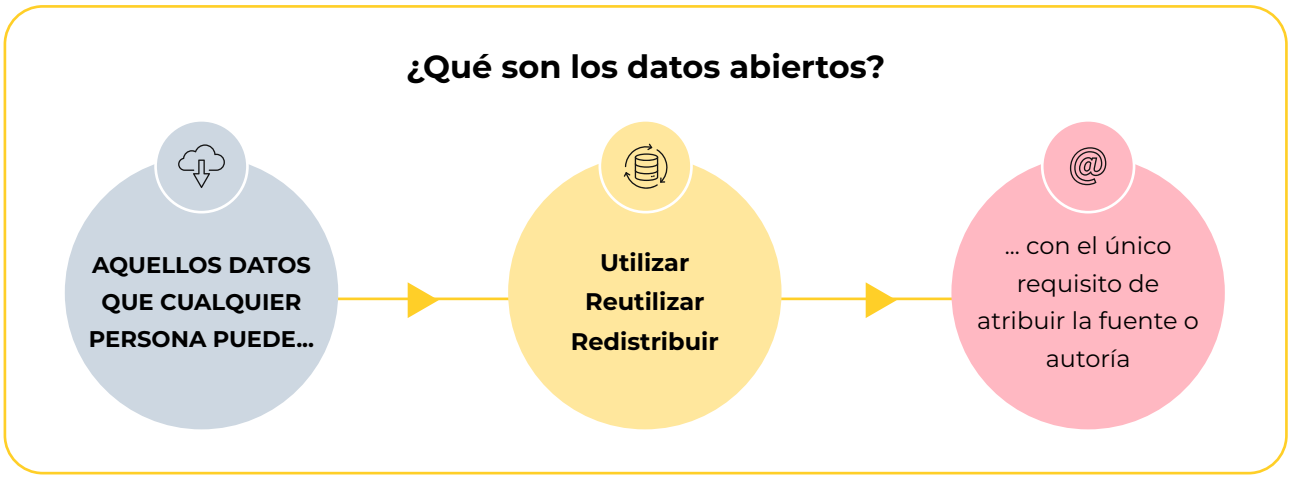



***Figura 1.*** *¿Qué son los datos abiertos?. Fuente: elaboración propia - datos.gob.es.*

## Más allá de los datos abiertos

No obstante, debemos ser conscientes de que una parte importante de los datos disponibles no cumplen plenamente todos los principios de los datos abiertos. Incluso en esos casos, seguirá siendo posible su reutilización. En este sentido, es también relevante la aparición de nuevas figuras como los **datos protegidos y los entornos seguros de datos**, como por ejemplo el Registro Europeo de Datos Protegidos del Sector Público (ERPD), que permiten el acceso controlado a información del sector público particularmente sensible garantizando su seguridad, trazabilidad y la supervisión de su uso, cumpliendo así con las obligaciones establecidas por la Ley de Gobernanza de Datos.

Por otro lado, en los últimos años ha cobrado especial protagonismo el concepto de espacios de datos, entendidos como **entornos de colaboración** donde distintos actores públicos y privados comparten datos de forma **segura, gobernada** e **interoperable**. Estos espacios permiten combinar datos abiertos con otros datos protegidos o privados, y están llamados a convertirse en uno de los principales motores de innovación de la economía del dato.

También es importante aclarar que los datos abiertos constituyen tan sólo una parte del ecosistema actual de datos, cuya principal característica es precisamente su capacidad de reutilización. Existen muchas otras denominaciones para los datos dependiendo de sus características, y así podremos oír hablar de *Big Data* cuando se gestionan grandes volúmenes de datos variados; datos dinámicos cuando están sujetos a altas frecuencias de actualización; datos enlazados o *Linked Data* cuando se crean conexiones entre los datos para representar las relaciones existentes; datos sintéticos, cuando son generados de forma artificial, etc. Los datos abiertos son compatibles con todas estas categorías, dando lugar a combinaciones del tipo *Big Open Data, Linked Open Data, Big Linked Open Data,* etc.

Estas definiciones, junto con otros conceptos clave relacionados con el ecosistema del dato, pueden consultarse y ampliarse en el glosario de términos relacionados con las nuevas tecnologías y los datos.

# 2. DATOS REUTILIZABLES: LICENCIAS

Las licencias de uso de los datos son un elemento de gran importancia en el que tendremos que fijarnos a la hora de plantearnos la explotación de cualquier dato disponible, ya que son el primer habilitador de la reutilización.

El escenario que nos podemos encontrar puede ser muy variable. Algunas entidades pueden optar por habilitar la **reutilización libre** de sus datos sin condiciones concretas y para cualquier fin, mientras que otras pueden optar por aplicar algún tipo de **licencia,** ya sea **una licencia predefinida o una personalizada,** donde se establezcan una serie de **condiciones de reutilización o restricciones específicas,** entre las que habitualmente se encuentran:

- No **alterar** o **desnaturalizar** el sentido original de la información.
- Citar la **fuente original** de los datos.
- Mencionar y mantener la **fecha de actualización** de los datos, siempre y cuando sea conocida.
- No sugerir de ningún modo que el titular de los datos participa en la reutilización que se lleve a cabo.
- **Conservar y aplicar las mismas condiciones** de reutilización para cualquier uso posterior que se haga de los datos.

Algunas de las licencias predefinidas y compatibles con los principios de los datos abiertos que nos podemos encontrar con más frecuencia son:

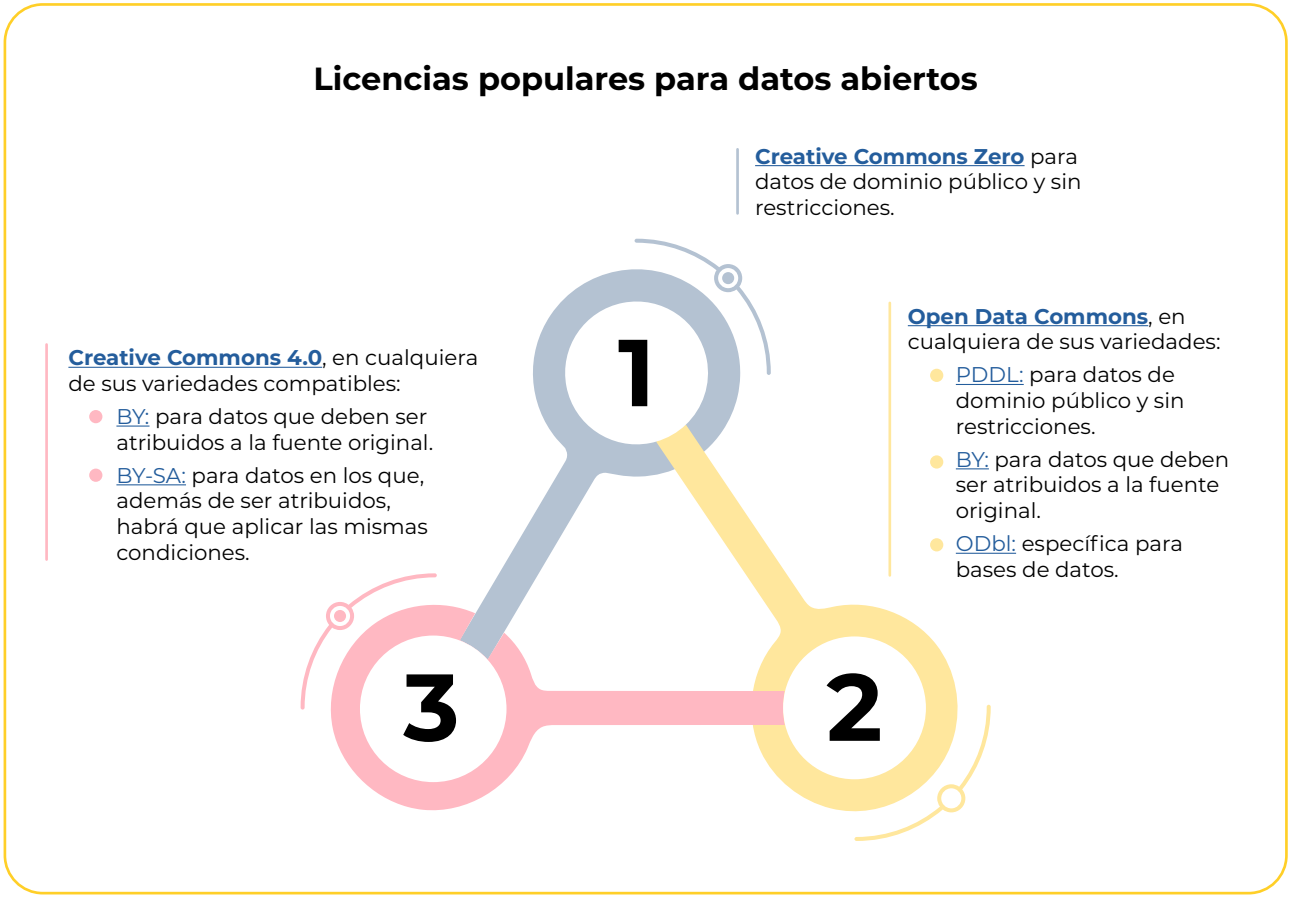



***Figura 2.*** *Licencias predefinidas más populares para datos abiertos. Fuente: elaboración propia - datos.gob.es.*

No obstante, existen también múltiples casos de **licencias personalizadas** con sus propias condiciones de uso aplicables. Por lo tanto, en caso de duda sobre la licencia aplicable o los términos de uso específicos para un caso concreto, lo más recomendable será consultar primero con los responsables de los datos, puesto que la reutilización de cualquier dato para cuyo uso no estemos autorizados podría conllevar consecuencias legales a posteriori.

Es importante no asumir que la ausencia de información sobre la licencia aplicable o las condiciones de reutilización implica que los datos se pueden reutilizar libremente. Los derechos del autor pueden ser aplicables igualmente, aunque no esté indicado de forma explícita.

## Nuevos retos por el uso masivo de datos

La aparición de nuevas tecnologías basadas en **inteligencia artificial** (IA) ha puesto en relieve la necesidad tanto de reforzar la **claridad de las licencias** como la **calidad de los metadatos** que las acompañan. Muchos sistemas de IA utilizan grandes volúmenes de datos para su entrenamiento, y durante esos procesos puede resultar especialmente difícil trazar el origen de los datos o aplicar correctamente condiciones como la atribución o la restricción de usos derivados. Este reto ha dado lugar a un nuevo debate sobre cómo asegurar el respeto a los derechos de los titulares de los datos en estos procesos automatizados a gran escala. Algunas estrategias que se están explorando son:

- La inclusión explícita en las licencias de **cláusulas específicas** sobre el uso de datos por estos sistemas de IA, cuando se desee limitar, condicionar o facilitar estos procesos como parte de un nuevo contrato social para la era de la IA.
- El fortalecimiento de los **metadatos de atribución y procedencia** de los datos, facilitando su lectura y procesamiento automatizados, como los proporcionados por la propia Creative Commons.
- El desarrollo de mejores sistemas de **trazabilidad** que permitan identificar y atribuir correctamente los datos incluso en contextos complejos o automatizados, como los metadatos incluidos en el perfil de aplicación de DCAT para los portales de datos Europeos (DCAT-AP).




***Figura 3.*** *Cómo asegurar el respeto a los derechos de los titulares de los datos. Fuente: elaboración propia - datos.gob.es.*

# 3. DÓNDE ENCONTRAR LOS DATOS

Hoy en día podemos encontrar fácilmente múltiples fuentes de datos abiertos que han sido específicamente diseñadas para facilitar su reutilización a través de lo que se denomina **catálogos de datos.** Estos catálogos no son más que repositorios electrónicos *online* – generalmente en la forma de páginas web – donde se almacenan y administran tanto los **datos** como algunos **metadatos** que los describen.

Podemos encontrar catálogos con datos abiertos del sector público en todos los niveles administrativos, ya sea nacional, regional o local. Sin embargo, dado que el número de portales y repositorios continúa creciendo día a día, la tendencia actual es facilitar el uso a los potenciales reutilizadores mediante la consolidación a través de **meta-catálogos y agregadores,** que ofrecen una visión consolidada y estable de los distintos conjuntos de datos a través de un único punto de acceso común.

Es importante también señalar que si buscamos datos sobre una temática concreta –como por ejemplo educación o sanidad– éstos podrían encontrarse dispersos a través de los distintos niveles administrativos, debido a la forma en la que se dividen las competencias y responsabilidades entre dichas administraciones. Para tratar de paliar el problema que esto puede ocasionar a la hora de localizar la información que nos interesa, la tendencia actual consiste en utilizar mecanismos de **federación** entre los distintos catálogos de datos. De esta forma podremos contar con un único punto de acceso final a los datos, independientemente de cuál sea su origen administrativo.

Además, existen también otros **repositorios temáticos** que concentran datos de una misma materia y que pueden resultar especialmente útiles para los reutilizadores más especializados, como, por ejemplo:

- Repositorios del sector de la economía.
- Repositorios de salud y bienestar.
- Repositorios de ciencias naturales y medio ambiente.
- Repositorios del ámbito del turismo.




***Figura 4.*** *Ejemplos de repositorios temáticos. Fuente: elaboración propia - datos.gob.es.*

## Catálogo Nacional de Datos

A nivel nacional debemos destacar en primer lugar el propio **datos.gob.es**, la iniciativa que organiza y gestiona el **Catálogo Nacional de Datos del sector público** y que actualmente aglutina aproximadamente cien mil conjuntos de datos provenientes de 300 entidades distintas en todos los niveles de la administración. Además, también cada vez son más los organismos de la Administración General del Estado (AGE), autonómicos, locales y otras entidades (como por ejemplo las Universidades) que publican sus propios catálogos de datos para facilitar la reutilización. Sin embargo, entre ellos son amplia mayoría los que deciden integrarse también dentro del catálogo nacional como parte de la estrategia de federación comentada anteriormente.

A nivel europeo el portal **data.europa.eu** cumple esa misma función como agregador de los países de la Unión, federando los catálogos nacionales de los Estados miembros y facilitando la **interoperabilidad** y la localización de datos en iniciativas transfronterizas, incluidos los futuros espacios comunes europeos de datos.

Por último, cabe también resaltar que en la actualidad existe además un volumen significativo de conjunto de datos que se publican sin formar parte de ningún catálogo en concreto, lo que puede dificultar en parte su localización y posterior uso.

# 4. LA CALIDAD DE LOS DATOS

Los problemas con la calidad no son una novedad de los datos abiertos, sino un reto histórico en la gestión de los datos que nos acompaña desde los inicios de las Tecnologías de la Información y la Comunicación. La fiabilidad absoluta de los datos es, en cierto modo, una quimera que difícilmente podrá alcanzarse en la práctica, especialmente cuanto mayor es el volumen de datos gestionado y más dinámico sea su ciclo de vida.

En la actualidad, y en parte gracias a la proliferación de iniciativas de datos abiertos, **los datos se encuentran cada vez más expuestos** y son también **puestos a prueba** en nuevos escenarios en los que se analizan, visualizan y combinan para generar servicios y aplicaciones innovadoras. Esta exposición permanente hace aflorar problemas antes no detectados, lo que aumenta al mismo tiempo el nivel de exigencia en cuanto a su **calidad, integridad y consistencia.**

Por otro lado, esta mayor visibilidad de los datos permite también aprovechar nuevas herramientas de **colaboración y participación ciudadana que nos ayudan a mejorar la calidad.** Muchas iniciativas de apertura de datos incorporan canales de retroalimentación mediante los cuales los usuarios pueden comunicar errores o cualquier otro problema. Estos ciclos de mejora continua contribuyen también de forma significativa a **la mejora de la calidad final de los datos.**

## La inteligencia artificial como aliada para mejorar la calidad

A estos mecanismos se suma la **inteligencia artificial,** que se está convirtiendo en una herramienta muy valiosa a la hora de automatizar tareas de validación, limpieza, detección de duplicados, estandarización de formatos o identificación de valores atípicos. La IA puede ayudar a detectar patrones de error difíciles de ver manualmente, facilitar la normalización de los metadatos y acelerar significativamente los procesos de preparación de datos.

No obstante, es fundamental utilizar estas herramientas con cierta cautela, ya que la IA puede también introducir o amplificar **sesgos o interpretaciones incorrectas** si no se supervisa adecuadamente. En este contexto, cobra especial relevancia la **trazabilidad** o **linaje del dato,** que nos permite conocer el origen, las transformaciones, los responsables y las versiones de los datos a lo largo de su ciclo de vida. La trazabilidad nos ayudará a detectar el origen de los errores, asegurar la reproducibilidad, mejorar la confianza en los datos y mitigar riesgos en usos avanzados, especialmente cuando usemos análisis automatizado.

Por todo lo anterior, cualquier potencial reutilizador debería seguir algunas normas básicas para aprender a convivir con cierto grado de errores en los datos que vendrán ocasionados en su mayoría por su propio ciclo de vida (inserciones, actualizaciones, migraciones, etc.) y que en la gran mayoría de ocasiones son inevitables:

- Acostumbrarse a **trabajar con cierto grado de incertidumbre** en los datos como parte natural e inevitable de su tratamiento.
- **Planificar** sus proyectos de reutilización adecuadamente incluyendo el esfuerzo y los procedimientos necesarios para la validación y el tratamiento de errores[1] .
- Mantener una actitud crítica ante los resultados generados por herramientas automáticas, comprobando posibles **sesgos o errores sistemáticos.**
- Utilizar y verificar la **trazabilidad disponible** para comprender el origen y las transformaciones del dato.
- **Colaborar en el ciclo de mejora continua de los datos** reportando cualquier problema que se detecte y que pueda afectar a la calidad de los mismos.

[1] En la sección dedicada a las herramientas se incluyen varias que pueden ser útiles en este aspecto.




***Figura 5.*** *Recomendaciones para trabajar con datos. Fuente: elaboración propia - datos.gob.es.*

# 5. INTEROPERABILIDAD Y ESTÁNDARES

Otro de los grandes retos a la hora de reutilizar los datos es su nivel de **interoperabilidad,** un aspecto especialmente crítico cuando intentemos combinar datos procedentes de diversas fuentes, que es precisamente cuando el valor de los datos abiertos es mayor. Generalmente podemos decir que existen tres vertientes complementarias:

- **Interoperabilidad técnica:** condicionada por las **infraestructuras,** los **protocolos** y las **tecnologías** que se utilicen a la hora de compartir datos para que los sistemas puedan comunicarse entre sí.
- **Interoperabilidad semántica:** cuyo objetivo es asegurar que los sistemas no solo se comuniquen, sino también que se entiendan entre ellos, seleccionando los estándares adecuados para la representación de los datos (vocabularios, taxonomías y modelos de datos compartidos).
- **Interoperabilidad organizativa:** basada en **acuerdos, convenios, principios y procedimientos** comunes que garanticen coherencia en la publicación y gestión de los datos.

Para poder garantizar la interoperabilidad entre las distintas iniciativas de datos abiertos a nivel nacional se ha desarrollado la **Norma Técnica de Interoperabilidad de Reutilización de Recursos de Información (NTI-RISP).** A través de esta norma se definen las reglas para que las iniciativas puedan identificar, seleccionar y publicar la información de forma interoperable, orientando a la Administración en cuestiones de gran utilidad para los reutilizadores como:

- Los **formatos** que se deben utilizar.
- El esquema de **identificadores** para los datos.
- Los **metadatos** adecuados para describir los conjuntos de datos.
- Las **taxonomías** a utilizar para la clasificación de los datos.
- Los **términos y condiciones de uso** aplicables.

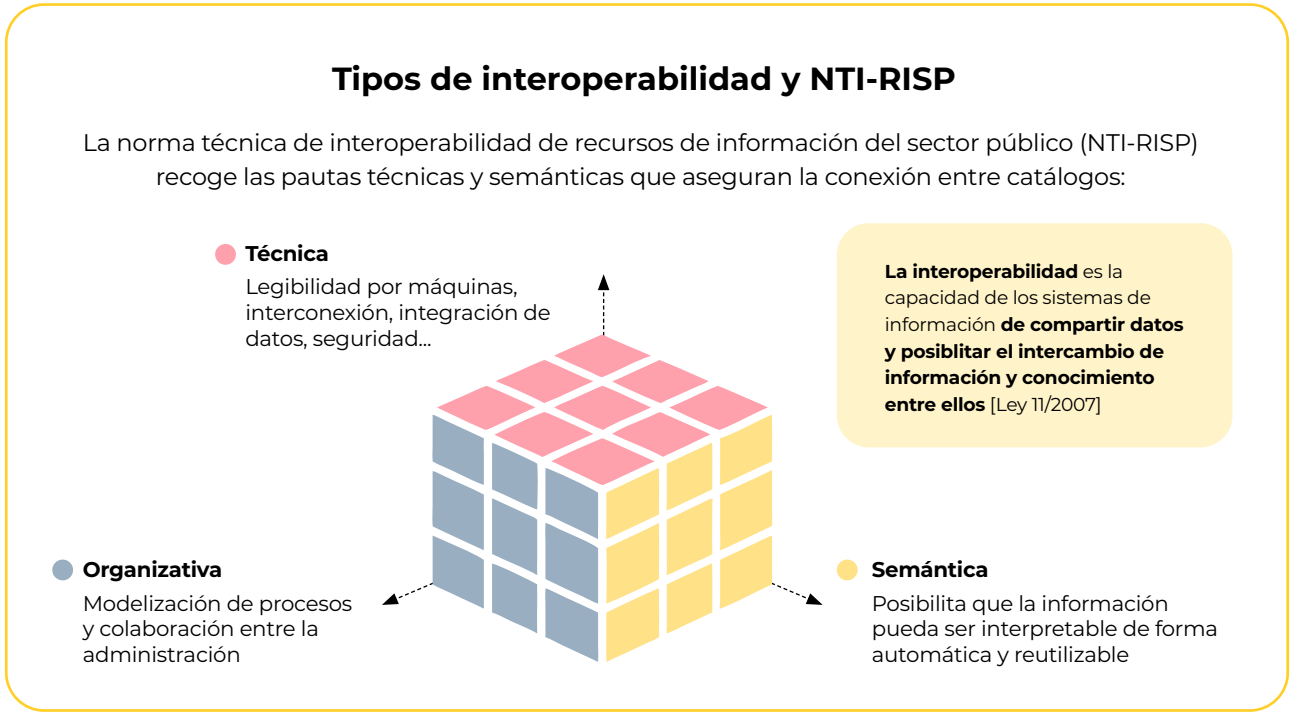



***Figura 6.*** *Tipos de interoperabilidad y NTI-RISP. Fuente: elaboración propia - datos.gob.es.*

Más recientemente, la NTI-RISP se ha **enriquecido y complementado** con DCAT-AP-ES, el perfil español del estándar europeo DCAT-AP, que proporciona un modelo de metadatos común para describir catálogos y conjuntos de datos y permite la interoperabilidad de los catálogos nacionales, regionales y locales, así como la federación automática entre ellos.

Además, la Asociación Española de Normalización y Certificación (AENOR) ha publicado un nuevo conjunto de especificaciones UNE orientadas a reforzar la gobernanza, la calidad y la gestión del dato, proporcionando a las organizaciones un marco común para evaluar procesos, asegurar la calidad y mejorar la coherencia en el ciclo de vida de los datos. A este conjunto se suma la incorporación de otra norma UNE pionera dedicada a los **espacios de datos**, que define su estructura, roles y requisitos fundamentales.

## Marco Europeo

En los últimos años, la Unión Europea ha reforzado significativamente sus esfuerzos en materia de interoperabilidad a través de la aprobación de un renovado **Marco Europeo de Interoperabilidad** (EIF), así como la nueva **Ley Europea de Interoperabilidad**, que establecen **un enfoque común y obligatorio** para compartir datos, procedimientos y servicios entre administraciones públicas europeas. Todo ello con el objetivo de conseguir conjuntos de datos publicados siguiendo prácticas lo más **homogéneas** e **interoperables** que sea posible.

Además, el programa SEMIC de la Comisión Europea mantiene y actualiza una serie de vocabularios de referencia ampliamente utilizados en datos abiertos y en los espacios de datos. Estos vocabularios facilitan que información publicada por distintas administraciones pueda combinarse y reutilizarse sin pérdida de significado.

# 6. PERSISTENCIA Y VOLATILIDAD

Una barrera con la que todavía pueden encontrarse los reutilizadores de la información con cierta frecuencia es la volatilidad de los datos, es decir, la falta de continuidad o estabilidad en su publicación. Este problema es **cada vez menos frecuente, gracias en parte a las obligaciones normativas** en cuanto a transparencia, reutilización y publicación activa de los datos. No obstante, sigue siendo un reto relevante, especialmente en iniciativas sin una estrategia clara de gobernanza del dato o sin recursos suficientes para mantener su continuidad.

Dado que los acuerdos de nivel de servicio (SLA) y el control de versiones de datos (CVD) no son algo que esté instaurado todavía en las iniciativas de datos abiertos, algunos ejemplos de posibles problemas que pueden encontrarse los reutilizadores son:

- **Ausencia de información histórica** publicada por haberse iniciado la iniciativa en cierto tiempo sin afectar al histórico previo.
- **Borrado de información histórica** al realizar una actualización de los datos publicados, conservando únicamente la información más reciente.
- **Falta de continuidad de los datos,** bien sea por una falta temporal de recursos o incluso porque la iniciativa llegue a desaparecer por completo.

Ante estos posibles escenarios, la estrategia más prudente que podemos adoptar a la hora de actuar como reutilizadores cuando la disponibilidad de los datos no esté garantizada, es **replicar la información que nos interese de forma local**[2] y mantenerla después siempre actualizada respecto a la fuente original.

En caso de que finalmente se produjese una discontinuidad en la publicación de los datos, es recomendable contactar también con el proveedor original de los datos para tratar de obtener un medio alternativo de acceso (por ejemplo, a través de repositorios institucionales). Cuando esto no sea posible, podemos intentar también buscar una fuente alternativa para la misma información, por ejemplo, en caso de que puedan existir meta-catálogos, agregadores, o repositorios temáticos que preserven versiones históricas de los datos.

## Gestión y gobernanza de los datos

La persistencia de los datos depende en gran medida de la existencia de una **gestión y gobernanza adecuadas,** que incluyan responsabilidades claras, procesos estables y criterios de conservación. Marcos como las recientes especificaciones **UNE** sobre el gobierno del dato ya anteriormente mencionadas, o **DAMA Internacional** y su *Data Management Body of Knowledge* (DMBoK), proporcionan orientaciones sobre cómo asegurar el ciclo de vida, la disponibilidad y la trazabilidad de la información. Sin embargo, no todas las organizaciones públicas disponen de estas capacidades plenamente implantadas.

En definitiva, la persistencia y continuidad de los datos no solo dependen de su publicación inicial, sino de contar con **procesos robustos de gobernanza, calidad y conservación,** alineados con los estándares nacionales e internacionales. A medida que estas prácticas se vayan consolidando, la volatilidad disminuirá y la reutilización será más fiable y sostenible en el tiempo.

[2] Siempre verificando que sea posible de acuerdo a la licencia y los términos de uso de los datos.

# 7. HERRAMIENTAS PARA EL TRATAMIENTO DE LOS DATOS

Hoy en día existe un amplio abanico de herramientas que nos serán de gran utilidad a la hora de trabajar con los datos. Entre los distintos tipos de herramientas disponibles podríamos destacar las siguientes categorías:

- Herramientas para el **tratamiento y limpieza de los datos,** cuya utilidad reside en la ayuda que ofrecen a la hora de filtrar, mejorar y clasificar la información, principal tarea para la que han sido diseñadas. Ejemplos de este tipo son herramientas como Open Refine, junto con librerías modernas de ciencia de datos como pandas o tidyverse, y entornos interactivos como Jupyter Notebook.
- Herramientas de **análisis de datos,** que serán de gran valor a la hora de sacar el máximo partido de los datos realizando análisis complejos. Algunas de los más comunes son los entornos de trabajo RStudio (para R), Spyder (para Python) o Positron (multi-lenguaje).
- Herramientas de **visualización y cuadros de mando,** que nos permiten crear visualizaciones de los datos más amigables a la hora de poder interpretarlos. Por ejemplo Tableau o PowerBI.
- Herramientas para la creación de **mapas y geolocalización,** que pueden considerarse en cierta forma un subconjunto de las anteriores, y cuya diferenciación consiste en utilizar los mapas como medio principal para la comunicación de la información. Ejemplos en este caso son Carto o ArcGIS.
- Herramientas para la generación de **redes** de datos conectados, que nos facilitan la labor de descubrir relaciones y patrones existentes. En esta área tenemos los ejemplos de Gephi y NetworkX.
- **Librerías e Interfaces de programación,** imprescindibles a la hora de realizar tareas más complejas y personalizadas para la transformación de los datos y su explotación a través de servicios y aplicaciones. Aquí contamos con ejemplos como D3.js o Apache Arrow.
- Las nuevas herramientas de **IA y aprendizaje automático** que facilitan el trabajo con grandes volúmenes de datos y nos ayudan a automatizar tareas como el análisis exploratorio, la clasificación o la predicción, como por ejemplo TensorFlow y PyTorch.



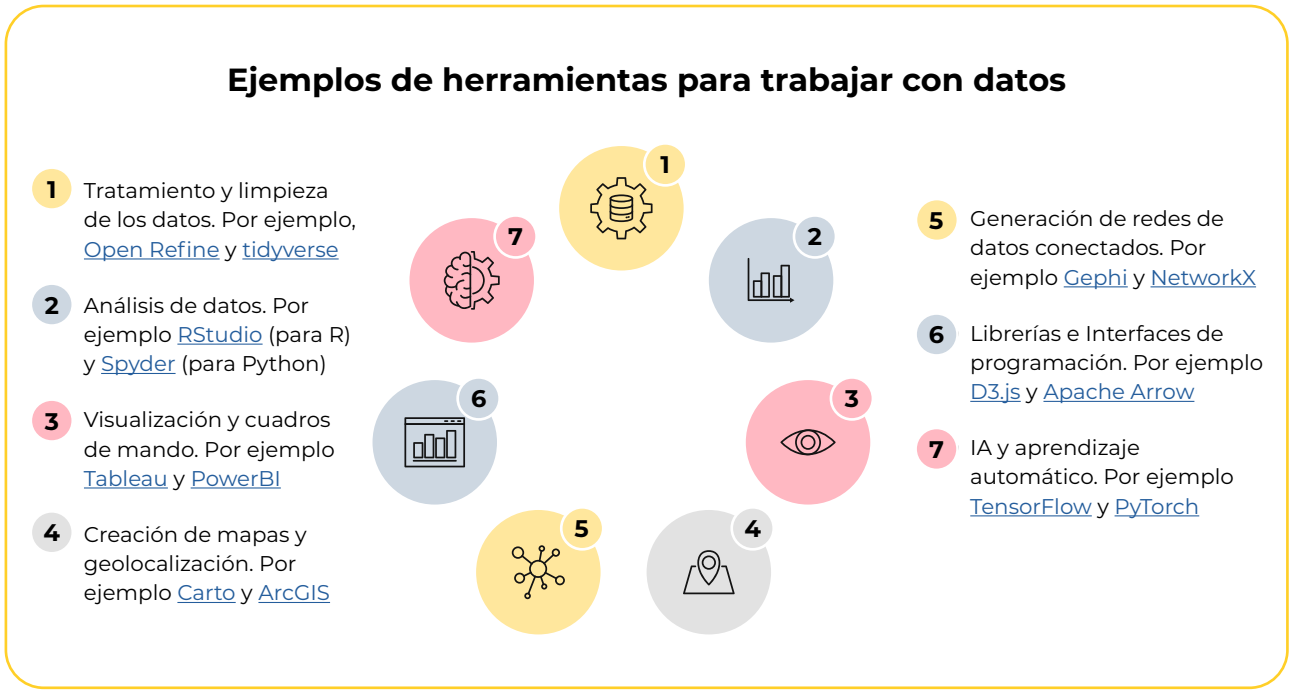



***Figura 7.*** *Ejemplos de herramientas para trabajar con datos. Fuente: elaboración propia - datos.gob.es.*

Desde datos.gob.es se han creado también otros recursos de gran valor para quienes buscan herramientas, metodologías y buenas prácticas en el tratamiento responsable y eficiente de los datos, como la guía sobre las herramientas de procesamiento y visualización de datos, la guía práctica de introducción al análisis exploratorio de datos en Python o en R, la guía sobre anonimización de datos y el decálogo del científico de datos.

# 8. FORMARSE Y EXPERIMENTAR CON DATOS

El ecosistema de los datos evoluciona con enorme rapidez, por lo que la formación continua y la experimentación práctica se han convertido en elementos esenciales tanto para quienes reutilizan datos públicos como para quienes trabajan en ciencia de datos, analítica, IA o gobernanza del dato. La **Estrategia Digital Europea** subraya la importancia de las *digital skills* y, en particular, de las capacidades relacionadas con el **uso, análisis y gestión de datos.** De hecho, numerosos proyectos de investigación financiados por la Comisión Europea incorporan ya actividades específicas de capacitación, alfabetización digital y ciencia de datos.

Tanto la formación como la experimentación práctica son esenciales para aprovechar plenamente el potencial de los datos abiertos y para desarrollar competencias clave en la economía del dato. Hoy en día existe una amplia oferta de formación, desde cursos introductorios hasta programas avanzados. Entre las opciones disponibles se encuentran:

- **MOOC y formación abierta:** plataformas como Coursera, edX, Udacity o FutureLearn ofrecen numerosos cursos sobre análisis de datos, gobernanza del dato o IA, varios de ellos impartidos por universidades de referencia.
- **Cursos especializados:** grandes empresas tecnológicas como Google, Microsoft, IBM o Amazon ofrecen algunos itinerarios formativos completos y gratuitos.
- **Certificaciones profesionales:** proporcionan un marco reconocido para comprender la gestión del ciclo de vida del dato, su calidad, su gobernanza y su valor, como puede ser el caso de DAMA CDMP (*Certified Data Management Professional*).
- **Recursos de formación:** disponibles a través del canal de conocimiento de datos.gob.es donde se ofrece todo tipo de recursos relacionados con los datos, datos abiertos y su reutilización, incluyendo: ejercicios de ciencia de datos, infografías, informes y estudios, vídeos, etc.

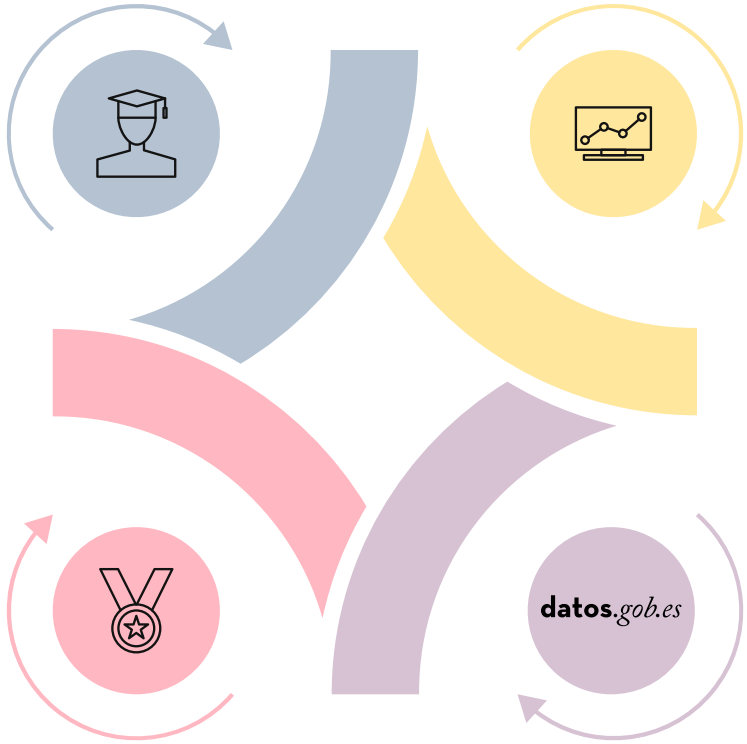
Por otra parte, el blog, los vídeos y el canal de noticias de datos.gob.es son también una de las fuentes más actualizadas sobre formación, jornadas, eventos y novedades del dato en el sector público. Además, el Encuentro Nacional de Datos Abiertos proporciona un **foro de discusión y de diálogo** abierto y carácter anual a todas las personas implicadas en la apertura de datos.



## Formación para impulsar el uso de datos abiertos

**MOOC y formación abierta.**
Cursos introductorios y avanzados en plataformas como Coursera, edX, Udacity o FutureLearn

**Cursos especializados.**
Itinerarios formativos gratuitos ofrecidos por Google, Microsoft, IBM o Amazon.




**Certificaciones profesionales.**
Marcos reconocidos como DAMA CDMP para dominar gestión, calidad y gobernanza del dato

**Recursos de datos.gob.es.**
Ejercicios, infografías, estudios, vídeos y materiales para aprender y practicar con datos abiertos

***Figura 8.*** *Formación para impulsar el uso de datos abiertos. Fuente: elaboración propia - datos.gob.es.*

# 9. EL VALOR DE LOS DATOS

El valor de los datos crece año tras año, no solo como motor económico, sino también como un bien social esencial para impulsar la transparencia, mejorar los servicios públicos y afrontar de manera colectiva los grandes retos de nuestra época. La capacidad de acceder, combinar y reutilizar datos es también fundamental para avanzar en los Objetivos de Desarrollo Sostenible, abarcando ámbitos como el cambio climático, la movilidad sostenible, la salud, la educación, la digitalización de las pymes o la planificación territorial. Por ello, garantizar **un acceso equitativo a los datos,** evitando brechas de uso y desigualdades en cuanto a capacidades digitales, es un elemento central de las estrategias europeas y nacionales.

La posibilidad de que tanto los ciudadanos como las empresas puedan acceder y utilizar cada vez un mayor volumen y variedad de datos está facilitando el nacimiento de un **nuevo mercado** **basado en la creación de productos y servicios de valor añadido** a partir de esa información. Son ya varios los estudios económicos, tanto a nivel nacional como a nivel europeo, que indican que las estimaciones del volumen potencial de negocio y de generación de empleo en torno a los datos no paran de crecer año tras año. Existen además múltiples ejemplos reales de empresas y aplicaciones basadas en datos abiertos donde podemos encontrar casos de uso en ámbitos como la movilidad, el turismo, la energía o el urbanismo, que nos muestran de forma tangible cómo los datos se transforman en soluciones reales.

Sin embargo, cuando hablamos de las posibilidades de negocio de los datos abiertos lo primero que debemos aclarar es que no existe un modelo de negocio específico, ya que la clave sigue siendo la misma de siempre: **conocer cuáles son las necesidades de las personas y tratar de darles respuesta** de la mejor forma posible. La diferencia es que ahora tenemos la oportunidad de acceder a un creciente volumen de datos que nos servirá como materia prima fundamental a la hora de explotar nuestra creatividad y capacidad analítica.

Si se desea ir un paso más allá y dar el salto entre la experimentación y el negocio, una opción muy recomendable es los concursos, retos y premios de reutilización de datos que organizan distintas administraciones, muy útiles a la hora de estimular la creatividad, conectar perfiles técnicos con necesidades públicas concretas y favorecer la aparición de prototipos, startups y nuevos servicios basados en los datos abiertos. También se puede contar con el apoyo de los cada vez más numerosos planes de apoyo a los emprendedores ofrecidos por un gran número de regiones y localidades. Estos programas generalmente ofrecen un soporte muy variado que va desde el **asesoramiento personalizado,** la **formación** o las **redes de intercambio** hasta el apoyo a la financiación mediante **ayudas, subvenciones** o proporcionando **locales adecuadamente equipados** para poder desarrollar una actividad.

## Los nuevos nichos de valor

Los **espacios europeos de datos sectoriales** (salud, movilidad, turismo, agricultura, energía, finanzas, clima, administración pública, entre otros) representan una nueva infraestructura de colaboración que permite compartir datos de manera segura, gobernada e interoperable. Estos ecosistemas abren la puerta a nuevos modelos de negocio basados en datos de calidad, fiables y reutilizables, y permiten que empresas de cualquier tamaño puedan innovar sobre una base común de interoperabilidad y reglas claras. Para muchas pymes y startups, estos espacios de datos suponen una oportunidad histórica de competir en igualdad de condiciones y de acceder a nuevos mercados europeos.

Por otro lado, la inteligencia artificial multiplica aún más este potencial al permitir extraer patrones, automatizar tareas, construir predicciones y desarrollar servicios cada vez más personalizados a partir de los datos. La combinación de IA con datos abiertos y datos compartidos en espacios sectoriales tiene el potencial de generar nuevas soluciones en ámbitos como la salud digital, la movilidad inteligente, la eficiencia energética, la predicción climática o la gestión de recursos públicos.

# 10. USO RESPONSABLE Y ÉTICO DE LOS DATOS

El creciente volumen de datos disponibles y la incorporación acelerada de tecnologías como la inteligencia artificial están ampliando enormemente las posibilidades de reutilización, pero también los **riesgos asociados a un uso inadecuado o malintencionado.** La responsabilidad en el uso de los datos es por tanto hoy en día otro elemento esencial para asegurar que sus beneficios se compartan de forma justa y que no se agraven las desigualdades existentes o se generen nuevos impactos negativos en la sociedad o el medio ambiente.

El uso irresponsable de los datos puede manifestarse de múltiples maneras: a través de la vulneración de la **privacidad,** la explotación comercial indebida, la generación o amplificación de **sesgos,** la **exclusión** de determinados colectivos, las **decisiones automatizadas** que no se pueden explicar, la **discriminación** algorítmica o el uso indebido de **información sensible.** Además, las herramientas de IA puede tener también un **efecto multiplicador** sobre estos riesgos, especialmente cuando se alimentan de datos poco representativos o carentes de un contexto adecuado.

Por ello, cualquier reutilizador debe adoptar una actitud crítica y responsable a la hora de trabajar con los datos, teniendo en cuenta no solo la calidad o la interoperabilidad, sino también las **implicaciones sociales, éticas y ambientales.** Entre los aspectos clave de un uso responsable destacan:

- **Identificar y mitigar sesgos** en los datos utilizados, comprendiendo qué colectivos pueden estar afectados y cómo esos sesgos influyen en el resultado de los análisis o modelos.
- **Garantizar la transparencia,** tanto en el origen y la trazabilidad de los datos como en los métodos empleados para transformarlos, analizarlos o integrarlos en otros sistemas.
- **Promover un acceso equitativo,** facilitando que aquellas personas y entidades con menos recursos digitales puedan beneficiarse también del valor social y económico de los datos.
- **Evitar usos malintencionados o dañinos,** como la manipulación, la desinformación, la explotación abusiva o la creación de perfiles sensibles sin las garantías adecuadas.
- **Aplicar los principios de minimización y proporcionalidad,** especialmente en soluciones de IA que consumen grandes volúmenes de datos.
- **Tener en cuenta la sostenibilidad,** evitando prácticas que impliquen un uso excesivo de recursos computacionales que incrementen la huella de carbono sin aportar valor real.
- **Cumplir los marcos éticos y regulatorios vigentes,** incluidos las pautas para una IA ética establecidas por la Unión Europea, las obligaciones en materia de protección de datos y las recomendaciones sobre gobernanza responsable del dato.

Un enfoque verdaderamente responsable implica no solo cumplir normas, sino incorporar una cultura de **ética del dato** en cada etapa del ciclo de reutilización.




***Figura 9.*** *Claves para trabajar con datos de forma ética y responsable. Fuente: elaboración propia - datos.gob.es.*

## DEL REUTILIZADOR AL CIENTÍFICO DE DATOS

A lo largo de este Decálogo se han recorrido los principales elementos que permiten **comprender, localizar, reutilizar y poner en valor los datos del sector público** en el contexto actual de la economía del dato, la inteligencia artificial y los espacios de datos. Desde los fundamentos de los datos abiertos hasta la importancia de la calidad, la interoperabilidad, la gobernanza, la formación, la experimentación, el valor social y económico, y el uso responsable, se configura un marco práctico para avanzar hacia **un uso más estratégico, ético y transformador de los datos.**

Para quienes deseen seguir profundizando en las competencias, herramientas y buenas prácticas necesarias para extraer todo el valor de los datos en contextos más avanzados de análisis, ciencia de datos e inteligencia artificial, se recomienda también visitar el Decálogo del Científico de Datos, como complemento natural a esta guía y como hoja de ruta para desarrollar capacidades técnicas y analíticas de alto valor para el **ecosistema del dato** en el que vivimos actualmente.



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