



CIUDADES
ABIERTAS

Adaptaciones y configuraciones necesarias que aseguren el cumplimiento de la NTI y la federación con datos.gob.es - Conjuntos de datos de alto valor

ACTUACIÓN D5 – Implantación de Soluciones para la Publicación de los Conjuntos de Datos Abiertos

INICIATIVA PLATAFORMA DE GOBIERNO ABIERTO, COLABORATIVA E INTEROPERABLE (121/17-SP)

Citar como: Corcho O, De Pablo V (2021) Conjuntos de datos de alto valor. doi: 10.5281/zenodo.5942571



red.es



UNIÓN EUROPEA

Fondo Europeo de Desarrollo Regional
"Una manera de hacer Europa"



Este documento ha sido elaborado en el marco de la iniciativa 'Plataforma de Gobierno Abierto, Colaborativa e Interoperable' cofinanciada por el Ministerio de Economía y Empresa, a través de la Entidad Pública Empresarial Red.es, y por los ayuntamientos de A Coruña, Madrid, Santiago de Compostela y Zaragoza y con la cofinanciación del Fondo Europeo de Desarrollo Regional (FEDER), dentro de la 'II Convocatoria de Ciudades Inteligentes'.

El presente documento está clasificado como "Interno". Esta clasificación habilita a su distribución entre los participantes en la iniciativa 'Plataforma de Gobierno Abierto, Colaborativa e Interoperable'. El receptor tendrá derecho al uso de la información contenida en el documento para los fines para los que el proyecto la ha facilitado, y ello sin perjuicio del cumplimiento de la normativa sobre propiedad intelectual y sobre protección de datos de carácter personal.

RESUMEN EJECUTIVO

El concepto de “datos de alto valor” o *high-value datasets* en inglés comienza a ser utilizado de manera generalizada en Europa con la publicación de la Directiva Europea 2019/1024 (comúnmente conocida como Directiva de Datos Abiertos u *Open Data Directive*). Los datos de alto valor se definen como aquellos conjuntos de datos que pueden tener un gran potencial para generar beneficios para la ciudadanía (y de manera general para la sociedad), considerando aspectos no sólo económicos, sino también relacionados con el medio ambiente y la economía.

A partir de la publicación de esta directiva, se han realizado varios estudios sobre el impacto de los datos de alto valor, con el objetivo de permitir su identificación, y en las transposiciones de la directiva que están realizando los distintos estados miembros (y que deberían haber estado todas disponibles en julio de 2021) se están siguiendo distintas estrategias para su identificación y catalogación. Por ejemplo, en el caso de España este trabajo aún no ha dado sus frutos y se encarga en la transposición la determinación de los conjuntos de datos de alto valor a la División Oficina del Dato.

En este informe se analizan los procesos seguidos en algunos países europeos (por ejemplo, Alemania y Holanda) y fuera de Europa (por ejemplo, Canadá) para la determinación de este catálogo en distintos niveles administrativos, y se revisan también los trabajos que previamente habían sido realizados para la identificación de estos conjuntos de datos por parte de las administraciones locales, por grupos de trabajo voluntarios como los de la norma UNE 178301:2015, la iniciativa OjoAIData100 y el grupo de trabajo de datos abiertos de la Red de Entidades Locales por la Transparencia y la Participación Ciudadana de la Federación Española de Municipios y Provincias.

Basándose en todo este trabajo previo, se propone un protocolo para la identificación y catalogación de conjuntos de datos que se podrían considerar de alto valor para las entidades locales y se utiliza este protocolo para el análisis de aproximadamente 100 conjuntos de datos que surgen de los identificados y refinados por la iniciativa OjoAIData100 y por el trabajo realizado en algunos otros países. A partir de este catálogo inicial se realiza una valoración subjetiva, basada en la opinión agregada de tres expertos, sobre distintas dimensiones de cada uno de ellos, y se realiza una primera propuesta de 50 conjuntos de datos que podrían ser considerados de alto valor, que podría servir como base para la puesta en común con otras ciudades, preferiblemente en el contexto del grupo de trabajo de la FEMP mencionado anteriormente.

Finalmente, se ofrece una propuesta para realizar la descripción de estos conjuntos de datos y para los siguientes pasos a realizar para su normalización posterior, con el objetivo de facilitar la homogeneización en la publicación de datos abiertos por parte de las administraciones locales españolas.

Este documento debe ser considerado como un documento de trabajo inicial que podrá ser cuestionado por grupos de trabajo específicos y refinado de manera periódica.

TABLA DE CONTENIDOS

RESUMEN EJECUTIVO	2
TABLA DE CONTENIDOS	3
1. CONJUNTOS DE DATOS DE ALTO VALOR: CONTEXTO Y ESTADO ACTUAL	4
1.1 CONJUNTOS DE DATOS DE ALTO VALOR EN ESPAÑA	7
1.2 CONJUNTOS DE DATOS DE ALTO VALOR EN OTROS PAÍSES DE EUROPA Y DEL MUNDO	10
2. PROTOCOLO DE CATALOGACIÓN DE DATOS DE ALTO VALOR	13
3. SELECCIÓN INICIAL, CLASIFICACIÓN Y ANÁLISIS DE CONJUNTOS DE DATOS	16
3.1 CONJUNTOS DE DATOS EN LA CATEGORÍA GEOESPACIAL	17
3.2 CONJUNTOS DE DATOS EN LA CATEGORÍA OBSERVACIÓN DE LA TIERRA Y MEDIO-AMBIENTE	18
3.3 CONJUNTOS DE DATOS EN LA CATEGORÍA METEOROLOGÍA	19
3.4 CONJUNTOS DE DATOS EN LA CATEGORÍA ESTADÍSTICA (INCLUYENDO ECONOMÍA)	19
3.5 CONJUNTOS DE DATOS EN LA CATEGORÍA DE SOCIEDADES	20
3.6 CONJUNTOS DE DATOS EN LA CATEGORÍA MOVILIDAD	21
3.7 CONJUNTOS DE DATOS EN OTRAS CATEGORÍAS (SALUD, SEGURIDAD, CULTURA Y DEPORTE)	22

1. CONJUNTOS DE DATOS DE ALTO VALOR: CONTEXTO Y ESTADO ACTUAL

La Directiva Europea 2019/1024¹ (comúnmente conocida como Directiva de Datos Abiertos u *Open Data Directive*) define los datos de alto valor como aquellos conjuntos de datos que pueden tener un gran potencial para generar beneficios para la ciudadanía (y de manera general para la sociedad). Son, por tanto, y según esta norma: "datos cuya reutilización está asociada 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, aplicaciones y puestos de trabajo nuevos, dignos y de calidad, y al número de beneficiarios potenciales de los servicios de valor añadido y aplicaciones basados en tales conjuntos de datos". La generación de beneficios se entiende de una manera amplia en el contexto de esta directiva, considerando aspectos no sólo económicos, sino también relacionados con el medio ambiente y la economía.

Esta directiva identifica un grupo inicial de categorías de conjuntos de datos como prioritarias para la identificación de los conjuntos de datos de alto valor: el geoespacial, el medio-ambiental, el meteorológico, el estadístico, la información sobre sociedades mercantiles y la movilidad. A modo de ejemplo, apunta a algunos conjuntos de datos que podrían pertenecer a este listado: códigos postales y mapas nacionales y locales (categoría geoespacial), consumo de energía e imágenes de satélite (medio-ambiental), datos de instrumentos de medida y pronósticos meteorológicos (meteorología), indicadores demográficos y económicos (estadístico), registros de empresas y sus identificadores (sociedades mercantiles) y señales de carreteras y canales de agua navegables (movilidad). Asimismo, deja abierta la posibilidad de que dicho conjunto de categorías pueda ampliarse como resultado de la transposición de la directiva en cada uno de los Estados Miembros de la Unión Europea. Esta transposición estaba inicialmente prevista como muy tarde para julio de 2021 y en el caso de España ha sido finalmente incluida como parte del Real Decreto-ley 24/2021, del 2 de noviembre².

Aunque parte de los fundamentos de la Directiva de Datos Abiertos, según el Comisario de Mercado Interior, Thierry Breton, pasan por "facilitar la creación de productos y servicios de información en toda la UE y garantizar el uso transfronterizo eficaz de datos del sector público", la iniciativa de identificación de conjuntos de datos de alto valor también tiene como fin contribuir a importantes objetivos sociales, como la responsabilidad y la transparencia de las empresas. Para realizar esta identificación de conjuntos de datos de alto valor, se debe evaluar el potencial de los datos, identificarlos y definirlos, analizar los beneficios potenciales y costes incurridos en su publicación, evaluar el impacto en el presupuesto de los propietarios e intermediarios de los datos y asegurar que la reutilización respeta la normativa sectorial sobre publicación de conjuntos de datos y otras reglas horizontales, como las de protección de datos.

Con idea de hacer hincapié en qué son datos de alto valor y ayudar en su definición, data.europa.eu publicó en julio de 2020 el *Analytical Report 15: High-value datasets: understanding the perspective of data providers*³, en el que define el valor de datasets específicos y trata de dar respuesta a la complejidad y al papel de los proveedores de datos de los distintos estados miembros.

El valor de los datos se percibe de manera diferente desde los distintos puntos de vista de proveedores y de usuarios. No son lo mismo las industrias y empresas -que difieren también en función de su actividad- que la ciudadanía que, al mismo tiempo, puede tener el papel de usuario y de proveedor, a través de la cesión voluntaria de sus datos para el bien común (*data for good*) y sin ánimo de lucro. El papel de las administraciones, en este caso, cubre todas las etapas, desde proveedor de conjuntos de datos a usuario de los mismos, siendo también quien publica (portales de datos abiertos), quien establece pautas legales y, a partir de las directrices de la propia Comisión Europea, quien gestiona y centraliza el funcionamiento de los espacios de datos de alto valor: "Los gobiernos centrales", como se indica en el ya mencionado *Analytical Report 15* del *European Data Portal*, "también deberían conectar a los editores de datos, coordinar iniciativas en torno a los conjuntos de datos de alto valor y

¹ <http://data.europa.eu/eli/dir/2019/1024/oj>

² <https://www.boe.es/eli/es/rd/2021/11/02/24>

³ https://data.europa.eu/sites/default/files/analytical_report_15_high_value_datasets.pdf

(doi:

10.2830/363773)

trabajar como intermediarios en debates a nivel europeo". En Holanda, por ejemplo, donde el gobierno central tiene un enfoque más holístico a la hora de compartir datos y percibe su apertura como una oportunidad para el intercambio de datos en general, el gobierno no sólo proporciona el marco legal y técnico para la apertura de datos, sino para todo el abanico de opciones de intercambio de datos.

Definir ese valor añadido a los conjuntos de datos es una tarea compleja, ya que se deben tener en cuenta factores como la calidad, la interoperabilidad, sus limitaciones de uso, etc. Identificar el número de descargas de un conjunto de datos con su calidad no es real, ya que pueden existir conjuntos de datos de potencial muy alto aún no publicados o no localizados. Sin embargo, sí se pueden identificar esos conjuntos de datos de alto valor a través de una serie de indicadores, entre los que destacaría:

- Su potencial para generar beneficios sociales o medioambientales significativos.
- Su potencial para generar beneficios económicos y nuevos ingresos.
- Su potencial para generar servicios innovadores.
- Su potencial en cuanto a número de usuarios beneficiados, con atención particular a las PYMES.
- Su capacidad para ser combinados con otros conjuntos de datos.

La propia *Open Data Directive* da un serie de ideas que permiten entender qué conjuntos de datos podrían ser seleccionados por su alto valor (FIGURA 1), atendiendo a las temáticas anteriormente mencionadas:

- Geoespacial (datos geoespaciales).
- Observación de la Tierra y medio ambiente (observación de la tierra y datos medioambientales).
- Meteorología (datos meteorológicos).
- Estadística (datos estadísticos).
- Sociedades y propiedad de sociedades (registros empresariales).
- Movilidad (datos de transporte).



FIGURA 1: Datos de alto valor según la Open Data Directive (fuente: datos.gob.es)

Algunas de estas temáticas, además, están reguladas a través de normativa propia, como por ejemplo la Directiva 2007/2/CE sobre datos espaciales (INSPIRE)⁴, la Directiva 2003/4/CE sobre información medioambiental⁵ y la Directiva 2010/40/UE sobre datos de transporte⁶.

⁴ <https://eur-lex.europa.eu/legal-content/ES/LSU/?uri=celex:32007L0002>

⁵ <http://data.europa.eu/eli/dir/2003/4/oj>

⁶ <http://data.europa.eu/eli/dir/2010/40/oj>

La dificultad de determinar con claridad qué y cuáles son los conjuntos de datos que se pueden considerar de alto valor, llevó a la Comisión Europea a encargar, en 2019, la elaboración de un informe de impacto, en el contexto de un *Call for Tenders: Impact Assessment Study on High Value Datasets*⁷ para poder definir, con más detalle, este aspecto. Como resultado de este trabajo se elaboró el informe *Impact Assessment study on the list of High Value Datasets to be made available by the Member States under the Open Data Directive*⁸, elaborado conjuntamente por Deloitte, Open Data Institute, The Green Land y Lisbon Council. En este informe se hace un análisis de literatura en el que se consideran las siguientes seis dimensiones o macro-características para tener en cuenta si un conjunto de datos puede ser considerado de alto valor o no: beneficios económicos, beneficios medio-ambientales, generación de servicios innovadores y de innovación (por ejemplo, inteligencia artificial), reutilización, y la mejora, fortalecimiento y soporte a las autoridades públicas para llevar a cabo su misión (servicios públicos y labores sociales). Se tienen en cuenta 126 indicadores cualitativos y cuantitativos alrededor de estas dimensiones.

Los resultados de esta consulta aparecen recopilados en el *Summary Report on the open public consultation on the European strategy for data*⁹, un breve informe en el que se indica que el objetivo primero de la encuesta era "obtener información para una serie de iniciativas planificadas y relacionadas con el acceso y reutilización de datos, de manera particular el marco legislativo sobre datos europeos comunes espacios de datos, así como elaborar una lista de conjuntos de datos de alto valor". El punto 2.2 recoge los resultados relativos a los datos de alto valor, indicando que el 82,2% de los 761 participantes considera éstos una buena forma de garantizar que los datos del sector público puedan tener un impacto positivo en la economía y la sociedad de la UE. Para ello, se hace necesario que los datos de alto valor estén a disposición de particulares y empresas de forma gratuita y sin restricciones siendo, además, de fácil acceso. Los participantes aseguraron que es importante la buena disponibilidad de conjuntos de datos en condiciones uniformes en toda la Unión Europea: este acceso es uno de los factores más importantes a la hora de definir qué se puede considerar conjuntos de datos de alto valor. Los formatos estandarizados de datos y metadatos, las licencias y otros términos aplicables a la reutilización de datos fueron, junto al acceso, los principales elementos que se han tenido en cuenta para definir qué son "conjuntos de datos de alto valor".

Los conjuntos de datos de alto valor son conjuntos de datos que obran en poder del sector público y que se empiezan a considerar imprescindibles para crear espacios de datos, esenciales para el crecimiento económico, la competitividad, la innovación, la creación de empleo y el progreso social en general. Para obtener más detalle sobre el contexto de los espacios de datos, se recomienda leer la sección 2.5 del informe de Ciudades Abiertas de título "Adaptaciones y configuraciones necesarias para que el portal de datos abiertos del ayuntamiento de Madrid cumpla con la NTI y la federación con datos.gob.es en el futuro – en el contexto de la Estrategia Europea de Datos", donde se describe en más detalle la evolución y rol que los portales de datos abiertos locales tendrán en el contexto de los avances que se proponen en la Estrategia Europea de Datos, con la llegada de los espacios de datos.

Los conjuntos de datos identificados, hasta la fecha, por la UE están relacionados con determinados dominios de información, como ya se ha enumerado, y para ser considerados como tal tienen que ser datos reutilizables de manera gratuita a través de aplicaciones disponibles en formato legible por máquina y descargables a gran escala siempre que sea posible. Asimismo, se considera que los datos de alto valor podrían ampliarse en un futuro muy próximo a otros sectores de gran importancia para el desarrollo económico como la salud o la energía. De este modo, los sectores para la creación de espacios de datos de alto valor podrían ser: datos de salud, datos geoespaciales, datos de medio ambiente, datos de meteorología, datos de movilidad, datos de empresa y estadística. La FIGURA 2: "Datos de alto valor y Espacios de Datos" recoge tanto los sectores en los que se podrían centrar los datos de alto valor en el futuro próximo, como los elementos más destacados de los espacios de datos. Ambos elementos ayudarán a establecer un protocolo de búsqueda y catalogación de datos para crear conjuntos de datos de alto valor.

⁷ <https://inspire.ec.europa.eu/news/call-tenders-impact-assessment-study-high-value-datasets>

⁸ <https://www.access-info.org/wp-content/uploads/Deloitte-Study-2020.pdf> (doi: 10.2759/493091)

⁹ https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68611



FIGURA 2: Datos de alto valor y Espacios de Datos

1.1 CONJUNTOS DE DATOS DE ALTO VALOR EN ESPAÑA

Como se ha comentado anteriormente, la *Open Data Directive* debía haber sido transpuesta a la legislación nacional de cada estado miembro como muy tarde en julio de 2021. En el caso de España, esto no ha ocurrido hasta noviembre del 2021, en el Real Decreto-ley 24/2021, del 2 de noviembre (libro III). A continuación se copia el texto del artículo que hace referencia a los conjuntos de datos de alto valor:

"Artículo 3.ter. *Conjuntos de datos de alto valor.*

1. Además de la lista de conjuntos de datos específicos de alto valor que, en su caso, establezca la Comisión Europea, se podrán determinar a nivel nacional otros conjuntos de datos adicionales seleccionados en relación a su potencial para generar beneficios socioeconómicos o medioambientales importantes y servicios innovadores; beneficiar a un gran número de usuarios, en concreto pymes; contribuir a generar ingresos, y la posibilidad de ser combinados con otros conjuntos de datos.

2. Dichos conjuntos de datos de alto valor, tanto los establecidos a nivel europeo como nacional:

a) Estarán disponibles gratuitamente, a reserva de lo previsto en el artículo 7.9.a).

b) Serán legibles por máquina

c) Se suministrarán a través de interfaz de programación de aplicaciones (API), y

d) Se proporcionarán en forma de descarga masiva, cuando proceda.

Se podrán especificar acuerdos organizativos relativos a la publicación y de reutilización de los tipos de conjuntos de datos de alto valor. Esos acuerdos serán compatibles con las licencias tipo abiertas. Los acuerdos podrán incluir condiciones aplicables a la reutilización, el formato de los datos y los metadatos, así como acuerdos técnicos para la difusión.

3. El Ministerio de Asuntos Económicos y Transformación Digital **aprobará la lista de los conjuntos de datos de alto valor nacionales que se publicará mediante Resolución de la Secretaria de Estado de Digitalización e Inteligencia Artificial. La selección y actualización de los conjuntos de datos incluidos en dicha lista se realizará a través de la División Oficina del Dato contando con la colaboración de los actores interesados, tanto públicos como privados, a través de los órganos y mecanismos que se establezcan**¹⁰

Por tanto, teniendo en cuenta la muy reciente creación de la Oficina del Dato, se puede concluir que este proceso de búsqueda e identificación de conjuntos de datos de alto valor requerirá aún de trabajo adicional en los próximos meses. Los protocolos de catalogación no han sido propuestos aún, y la selección de actores que participarán en este proceso no ha sido realizada aún tampoco.

De esta manera, por tanto, la referencia más reciente de un proceso de catalogado de conjuntos de datos de alto valor en el contexto de las administraciones locales es el trabajo realizado en el contexto de las siguientes tres iniciativas, en las que han participado diversos miembros del proyecto Ciudades Abiertas:

- Durante el año 2014, el grupo de trabajo que elaboró la norma técnica UNE 178301:2015 sobre datos abiertos y ciudades inteligentes, donde se identificaron 11 conjuntos de datos de interés, basándose en las peticiones de reutilizadores recibidas por administraciones públicas locales y regionales, fundamentalmente.
- Durante los años 2015 (a partir del 17 de abril de 2015) y 2016, el grupo de trabajo de la iniciativa abierta OjoAlData100, que surgió a partir del trabajo realizado en el grupo anterior. Este grupo pretendía multiplicar por 10 el número de conjuntos de datos del catálogo anterior, llegando aproximadamente a los 100 conjuntos de datos que deberían ser publicados preferentemente por las administraciones locales en sus portales de datos abiertos. Este grupo de trabajo realizó reuniones periódicas en Medialab-Prado, y pasó por varias fases en las que se comenzó con 250 conjuntos de datos aproximadamente y finalmente se consiguió reducir el catálogo (eliminando duplicados y refinando el trabajo inicial) a 93 conjuntos de datos en la versión de enero del 2017. Para cada uno de estos conjuntos de datos se ofrecía una breve descripción y ejemplos de conjuntos de datos publicados en distintos portales de datos abiertos. Desafortunadamente la Web que se creó (<http://ojoaldata100.okfn.es/>) para este grupo de trabajo no está ya activa, aunque sí están disponibles los resultados de refinamiento en hojas de cálculo: <https://docs.google.com/spreadsheets/d/1-qEZbB23zGcEIRdSsDfGnzFRiGS9idN5-wNFPgpoYJU/edit#gid=1707546358> (este enlace debería ser archivado adecuadamente, para evitar que se pierda).
- Grupo de trabajo de datos abiertos asociado a la Red de Entidades Locales por la Transparencia y la Participación Ciudadana de la Federación Española de Municipios y Provincias. En el contexto de este grupo de trabajo se elaboró la guía de datos abiertos para entidades locales del 2017 y en 2019 se creó una lista de cuarenta conjuntos de datos¹⁰, basada en el trabajo de la iniciativa OjoAlData100. Este es el catálogo más reciente de conjuntos de datos abiertos relevantes para entidades locales, que se resume en la siguiente figura.

¹⁰ <http://femp.femp.es/files/3580-1938-fichero/DATOS%20ABIERTOS%20FEMP%202019.pdf>



FIGURA 3: Conjuntos de datos identificados en la Guía de Datos Abiertos de la FEMP, edición 2019

1.2 CONJUNTOS DE DATOS DE ALTO VALOR EN OTROS PAÍSES DE EUROPA Y DEL MUNDO

Los pasos que se han seguido en el estudio sobre conjuntos de datos de alto valor en distintos países de Europa han sido muy diversos. Algunos países han realizado estudios más o menos detallados sobre los conjuntos de datos de alto valor que serían aplicables en su contexto, teniendo en cuenta la propuesta inicial de la directiva, y como trabajo preparatorio para la transposición de la directiva a su legislación correspondiente. Las técnicas utilizadas para la determinación de los conjuntos considerados de alto valor han pasado por análisis de literatura y trabajo de revisión (desk research) de acuerdo con algunas características, a encuestas online, workshops, *focus groups*, hackatones, etc.

Queda fuera del alcance de este informe realizar el análisis pormenorizado de los trabajos preparatorios en este sentido para todos los países europeos, por lo que se han escogido únicamente algunos para ilustrar el proceso seguido. Este análisis ha sido posible a partir de entrevistas personales y contactos con la red de nodos del Open Data Institute, así como con responsables de data.europa.eu, que han ayudado durante el proceso de identificación de iniciativas.

En el caso de Alemania, el Ministerio de Asuntos Económicos y Energía (BMWi) publicó en febrero de 2021 el estudio "High Value Datasets in Germany (HOCHWERTIGE DATENSÄTZE IN DEUTSCHLAND)"¹¹, preparado conjuntamente por Fraunhofer FOKUS, Institut der deutschen Wirtschaft Köln e.V. y iRights.Lab. Este informe tomó como punto de partida la propia directiva y el informe sobre análisis de impacto mencionado en la sección anterior. A continuación, realizó una encuesta online que permitió identificar 60 conjuntos de datos para un análisis más detallado, desde las perspectivas técnica, económica y legal. Se realizaron workshops con expertos de dominio en los distintos campos y se contó con un comité asesor para analizar los conjuntos de datos, su potencial y los retos asociados a su publicación.

Unterka- tegorie	Exemplarisch unter- suchter Datensatz (ID)	Föderale Ebene	Beschreibung	T	R	Ö
Administ- ration	Katasterdaten (Geo_01)	Land	Die Katasterdaten beinhalten Darstellungen und Beschreibungen zu Flurstücken, Nutzungen und Gebäuden. Zusätzliche Punktinformationen sind u. a. Angaben zur Punktart, Abmarkung und Lagegenauigkeit. Im vorliegenden Beispieldatensatz werden die Katasterdaten von Sachsen betrachtet.	Green	Green	Yellow
Administ- ration	Postleitzahlen (Geo_02)	Bund	Der Datensatz enthält sämtliche Postleitzahlen und Orte Deutschlands sowie zusätzlich die Ortsteile.	Yellow	Red	Yellow
Administ- ration	Postleitregionen (Geo_02_alt) ⁴³	Bund	Im vorliegenden Datensatz werden die Postleitregionen (2-stellig) mit regionaler Zugehörigkeit nach Fläche, Bevölkerung und Bevölkerungsdichte vom 31.12.2018 betrachtet.	Red	Green	Green
Administ- ration	Bodenschätzungsdaten (Geo_03)	Land	Die Bodenschätzungsdaten beinhalten Darstellungen und Beschreibungen zu Merkmalen und Eigenschaften des Bodens, wie z.B. Bodenart, Humus- und Kalkgehalt, Hydromorphie-merkmale etc. Im vorliegenden Beispieldatensatz werden die	Yellow	Green	Yellow

FIGURA 4: Ejemplo de análisis técnico, económico y legal de algunos conjuntos de datos en Alemania

En el contexto de Holanda, se han hecho distintos estudios a nivel nacional, regional y municipal. Por su interés para este informe, nos centramos en el trabajo en los conjuntos de alto valor a nivel municipal, que se remonta al los meses de octubre y noviembre de 2016, donde se realizan tres reuniones para la creación de una primera

11

https://cdn0.scrvt.com/fokus/660f375a7dd3619b/e5f5ef206057/Studie_Hochwertige_Datensaeetze_in_Deutschland.pdf

lista de datos municipales de alto valor, a partir de una colaboración del portal de datos abiertos de Holanda (overheid.nl), VNG/KING y la DSA (Digital City Agenda). Esta lista fue actualizada con posterioridad en el 2020. En la siguiente tabla se muestran, en su idioma original, los conjuntos de datos incluidos en la lista:

Thema	Werkgebied	Dataset
		Landelijke datasets [wel high value; geen directe publicatie actie nodig bij gemeenten]
Bestuur	Uitvoering van de Wettelijke taak	Bekendmakingen
		Verkiezingsuitslagen
Ruimte & Infrastructuur	Uitvoering van de Wettelijke taak	Vergunningen
Economie	Mogelijkheden/mate hergebruik	Werklocaties/bedrijventerreinen
	Mogelijkheden/mate hergebruik	Financiële verantwoordingsdata
Huisvesting	Mogelijkheden/mate hergebruik	Energie-labels
Landbouw	Mogelijkheden/mate hergebruik	CO2 uitstoot per gemeente
		Energieverbruik
	Mogelijkheden/mate hergebruik	Luchtkwaliteit binnen gemeente
		Kinderopvang en BSO-locaties
	Sociaal-economische waarde	Criminaliteits-cijfers
	Mogelijkheden/mate hergebruik	Gebiedsindelingen BAG
		Gemeentelijke datasets [high value, worden door individuele gemeenten geleverd]
		Raadsverslagen (Raadsinformatie)
		Handhaving en uitvoeringsinformatie
	Sociaal-economische waarde	Gemeentelijke adressengids
	Mogelijkheden/mate hergebruik	Wob-verzoeken, besluiten en bijlagen inclusief metadata.
Cultuur & recreatie	Uitvoering van de Wettelijke taak	Evenementen
		Horeca & winkellocaties

	Sociaal-economische waarde	Sportvoorzieningen
		Openbare Kunst
	Mogelijkheden/mate hergebruik	Historische Akten (geboorte-uittreksels e.d.)
		Monumenten
Financien	Uitvoering van de Wettelijke taak	Subsidies
Natuur en Milieu	Sociaal-economische waarde	Afvalcontainers
		Afvalkalender
		Bomen
Onderwijs en wetenschap	Mogelijkheden/mate hergebruik	Onderwijsdata (kwaliteit-en risicoprofielen)
Openbare orde & veiligheid	Uitvoering van de Wettelijke taak	Meldingen openbare ruimte
		Openbare toiletten
		Hondenuitlaatplaatsen
	Mogelijkheden/mate hergebruik	Strooiroutes
	Sociaal-economische waarde	Invalideparkeerplaatsen
		Openbare verlichting
Verkeer	Sociaal-economische waarde	Fietsenstallingen
	Mogelijkheden/mate hergebruik	Wegwerkzaamheden
		Verkeersintensiteiten/verkeerstellingen
Zorg en gezondheid	Uitvoering van de Wettelijke taak	Locatie van AED's
	Mogelijkheden/mate hergebruik	Gezondheidsvoorzieningen

TABLA 1: Lista de datos de alto valor municipal de Holanda

Fuera del contexto de la Unión Europea se puede mencionar Suiza, donde el Grupo de Trabajo sobre Digitalización dependiente del Ministerio de Industria y Comercio ha identificado más de 70 conjuntos de datos considerados necesarios para impulsar la reutilización de datos abiertos. Entre los participantes en el proceso de identificación figuran empresarios del mundo del automóvil y de las telecomunicaciones.

Finalmente, es importante destacar que además de la iniciativa holandesa no se ha encontrado ninguna otra iniciativa parecida a la realizada en el contexto de las actividades para la norma UNE 178301:2015, OjoAlData100 y grupo de datos abiertos de la FEMP en el entorno local en Europa, lo que refuerza la intuición inicial sobre la novedad del trabajo realizado en estos grupos de trabajo voluntarios. Sí se han encontrado algunos trabajos similares en Canadá, por ejemplo, con la iniciativa de "high-value municipal datasets" de Québec¹², realizada desde el año 2018, donde se establecen seis grupos de criterios para analizar el valor de los datos y determinar con ello 18 conjuntos de datos que deberían priorizarse. Estos grupos de criterios son: tratan aspectos socio-económicos y medio-ambientales, fomentan la innovación y el desarrollo económico sostenible, responde a una demanda comunitaria fuerte. Asimismo, es importante destacar que se analiza si el conjunto de datos ha sido estandarizado o no. Los 18 conjuntos de datos considerados prioritarios son los siguientes:

- Requêtes citoyennes
- Catalogue des données
- Budget annuel
- Contrats octroyés
- Paiements effectués
- Infractions légales - Contravention
- Licences des animaux
- Permis de construction
- Adresses
- Zonage
- Accidents de la route
- Statistiques sur le crime
- Offres de services récréatifs
- Géobase
- Travaux routiers
- Avis d'ébullition préventive
- Arbres et espaces verts
- Collecte des matières résiduelles

2. PROTOCOLO DE CATALOGACIÓN DE DATOS DE ALTO VALOR

Como ya se ha comentado en la sección anterior, el concepto de datos de alto valor fue acuñado en el año 2019 en el contexto de la directiva europea de datos abiertos (*Open Data Directive*), tomando como base algunos informes previos generados por el grupo de expertos en compartición de datos entre negocios y administraciones (*Business to Government*). Teniendo en cuenta el ámbito de aplicación de la directiva y su proceso de transposición a la legislación nacional de los Estados Miembros, a partir de ese momento se crearon varios grupos de trabajo para determinar cuáles serían los datos de alto valor que cada país podría identificar, a nivel nacional, en los dominios de aplicación prioritaria identificados (asociados a los posteriores espacios de datos que se están desarrollando en el contexto de la Estrategia Europea de Datos): salud, manufacturero, agricultura, finanzas, movilidad, medio ambiente y energía. Este proceso se espera que culmine en los próximos meses con las propuestas de dichos conjuntos de datos por parte de cada uno de los Estados Miembros. Estaba previsto contar con dichas propuestas hacia julio de 2021, pero las condiciones surgidas por la pandemia de la COVID-19 están retrasando el proceso en la mayor parte de los países, incluido España, donde la transposición de la directiva se ha realizado en noviembre de 2021, y donde se ha dejado para más adelante la selección de los conjuntos de datos de alto valor.

¹² <https://open.canada.ca/data/en/dataset/20ed36e5-85fb-4409-b272-b1fa1d8c92b3>

Como se ha podido observar en las descripciones de la sección anterior, no existe un protocolo ampliamente aceptado en el estado del arte para determinar cómo realizar la selección de los conjuntos de datos que pueden ser considerados de alto valor para una ciudad, un conjunto de ciudades o un espacio de datos de carácter urbano (vertical u horizontal). En los casos en que esta priorización se ha abordado a nivel nacional, tampoco existe un acuerdo común sobre el protocolo a seguir. Así, en países como Alemania se ha realizado un análisis cualitativo de tres dimensiones (técnica, legal y económica). En otros se ha optado por ampliar las características a tener en cuenta, definiendo y evaluando indicadores tanto cuantitativos como cualitativos. En otros, la base del análisis ha sido la de realizar encuestas para determinar las prioridades de los proveedores y consumidores de datos, y a partir de ellas derivar un listado inicial que ha sido refinado mediante reuniones de distintos tipos (workshops, focus groups, hackatones).

Los casos que se pueden considerar más cercanos son los de Holanda y Canadá, presentados en la sección 1.2. Sólo en el caso de Holanda se establecen unos principios¹³ para la selección de conjuntos de datos de alto valor a nivel municipal, como sigue:

- La lista no es una lista absoluta que cambiará con el tiempo;
- La lista tomará una forma dinámica dependiendo de los desarrollos relacionados con la apertura de datos por municipios;
- La lista sirve como lista de inspiración, no para ejercer presión sino para estimular;
- La lista ayuda a eliminar el "miedo" y / o los obstáculos para acceder a los datos;
- La lista hace lo más concreto posible lo que se requiere y cómo se puede acceder a los datos;
- La lista es pública y el procesamiento o modificación se hará con transparencia.

Es decir, se trata de una serie de principios generales, pero en ningún caso de un protocolo con instrucciones o pasos claros a seguir: la selección de conjuntos de datos se hace por un grupo de trabajo, se valida en reuniones públicas y es refrendada por el Ministerio correspondiente.

Como regla general para este tipo de estudios y análisis puede ser recomendable seguir algunas de las guías ofrecidas por el Digital Curation Center (DCC)¹⁴, que consideran que un buen sistema de identificación y catalogación de conjuntos de datos (de alto valor, en nuestro caso) es uno que esté basado en pautas previas de evaluación y selección de conjuntos de datos. Estas pautas deberían haber sido identificadas con antelación por la propia institución o conjunto de instituciones que quieran elaborar el catálogo correspondiente. A la hora de establecer estas pautas o protocolos de inclusión en el catálogo, se debe tener en cuenta cuál será la política de valoración y selección, que debe garantizar una toma de decisiones coherente, transparente y responsable, de modo que los compromisos se puedan contabilizar y rastrear, además de que esta debe cumplir con los requisitos legales, relacionados con privacidad y derechos de propiedad intelectual, entre otros. Esta política proporcionará la base para una evaluación adicional de los conjuntos de datos, que se basará en criterios generales como los indicados en las citadas Guías de Digital Curation Center (DCC), entre otros:

- Valor comercial o social: se debe tener en cuenta qué valor podrán proporcionar los datos una vez publicados y tratados.
- Singularidad: la medida en la que el recurso es la única o más completa fuente de información que puede derivarse de él.
- Relevancia: el contenido del conjunto de datos se ajusta a lo que es necesario para que se le pueda sacar partido.
- Potencial de redistribución: se puede determinar la confiabilidad, integridad y usabilidad de los archivos de datos.
- Coste económico de conservación y publicación: el coste que supondría la administración y conservación del recurso, y si saldría rentable en un futuro proporcionando beneficios.
- Documentación: información necesaria para facilitar el descubrimiento, acceso y reutilización de los datos, así como de los metadatos sobre la procedencia del recurso y el contexto de su creación y de su uso.

¹³ https://data.overheid.nl/sites/default/files/uploaded_files/High%20Value%20DataIijst%20Procedure.pdf

¹⁴ <https://www.dcc.ac.uk/guidance>

Con respecto al protocolo concreto a seguir en el contexto de este informe, estos son los pasos que se han propuesto, considerando el marco global marcado por la Open Data Directive, los trabajos previos realizados en el contexto nacional (norma UNE 178301:2015, OjoAlData100 y grupo de trabajo de la FEMP), y el contexto municipal sobre el que se está trabajando para este informe:

Paso 1. Selección inicial de un catálogo de al menos 100 conjuntos de datos sobre el que comenzar a realizar el trabajo. Este catálogo estará fuertemente relacionado con los trabajos previos realizados en España (especialmente la iniciativa OjoAlData100, que ya hizo esta selección de manera colaborativa con un grupo de actores variado) y considerando también algunos de los conjuntos de datos municipales identificados en iniciativas de otros países, siempre que sean relevantes.

Paso 2. Clasificación de los conjuntos de datos con respecto a las grandes temáticas identificadas en la directiva europea (geoespacial, observación de la Tierra y medio ambiente, meteorología, estadísticas, sociedades y propiedad de sociedades, y movilidad). También se tendrán en cuenta otras categorías en caso de que los conjuntos de datos iniciales no coincidan con dichas categorías predefinidas como de alto valor.

Paso 3. Análisis de las siguientes características de cada conjunto de datos:

- Valor de reutilización interna (del 1 al 5). Se analizará si se puede reutilizar internamente el conjunto de datos, para dar soporte a funcionalidades no previstas inicialmente.
- Valor comercial (del 1 al 5). Se analizará el potencial valor comercial que puede tener el conjunto de datos cuando sea reutilizado por terceros.
- Valor social (del 1 al 5). Se analizará el potencial valor social que puede tener el conjunto de datos cuando sea reutilizado por terceros o por la propia institución. En este punto también se tiene en cuenta el valor medio-ambiental del conjunto de datos.
- Singularidad (del 1 al 5). Se analizará el grado de singularidad del conjunto de datos. ¿Es un conjunto de datos que no puede ser aportado por alguna otra entidad regional, autonómica o nacional? 1 si hay otras entidades que lo puedan aportar, y 5 si no las hay. Valores intermedios si pueden aportarse partes de los datos.
- Enlazado con otros conjuntos de datos (del 1 al 5). Se analizará si el conjunto de datos se puede enlazar con otros (por ejemplo, el de accidente se puede enlazar con el callejero).
- Coste económico de mantenimiento (del 1 al 5). Se analizará el coste que puede tener el mantenimiento del conjunto de datos (1 si el coste es muy grande, 5 si es muy pequeño).
- Facilidad de generación de una API (del 1 al 5). Dado que los conjuntos de datos de alto valor deben ser proporcionados por una API, se analizará la facilidad de que pueda ser ofrecido de esta manera.

Es importante destacar que estos parámetros son distintos a los utilizados en el catálogo de la FEMP que sirve de base para este trabajo, puesto que en ese catálogo, en primer lugar, el objetivo no era determinar si un conjunto de datos era de alto valor o no, y se utilizaban criterios como el valor de reutilización (más general y ambiguo), la complejidad del conjunto de datos, si es utilizable en el contexto de la transparencia, si está afectado por la normativa de protección de datos y si es aplicable a ciudades de cualquier tamaño o no, entre otros. Estas características se mantendrán en cualquier caso, de manera general, en el catálogo que se propone en este informe.

Paso 4. Ranking basado en la puntuación total obtenida.

Paso 5. Propuesta de un número de conjuntos de datos (50 para este informe), y generación de la información detallada de cada uno de dichos conjuntos, con una descripción más completa basada en la usada en la guía FEMP, atributos esperados, ejemplos, etc.

Paso 6. Actualización en un repositorio abierto (por ejemplo, GitHub) de la lista de conjunto de datos, con los detalles correspondientes a cada uno de ellos.

3. SELECCIÓN INICIAL, CLASIFICACIÓN Y ANÁLISIS DE CONJUNTOS DE DATOS

La creación de este catálogo ha tenido en cuenta el proceso indicado en la sección 2. La siguiente tabla ofrece el análisis de las características de cada uno de los conjuntos de datos inicialmente seleccionados. La puntuación ofrecida para cada una de las dimensiones asociadas a cada conjunto de datos se ha realizado por parte de tres expertos, escogiéndose el valor entero más cercano a la media de todos los valores proporcionados.

El contenido siempre actualizado de todas estas tablas se encuentra en <https://docs.google.com/spreadsheets/d/1rZbOQIVH0eqbgs52EUmvJTd2HiCZZNYWBEXiY13vaZg/edit#gid=0>, con el objetivo de permitir su edición posterior en futuros grupos de trabajo donde se puedan revisar estos conjuntos de datos en el futuro.

3.1 CONJUNTOS DE DATOS EN LA CATEGORÍA GEOESPACIAL

Dataset	Valor reutilización interna	Valor comercial	Valor social	Singularidad	Enlazado	Coste mantenimiento	Facilidad API	Total
Callejero (nomenclátor geográfico)	5	5	3	3	5	5	5	31
Cartografía topográfica de áreas urbanas	4	2	3	1	4	5	3	22
Unidades administrativas	5	3	3	1	5	5	5	29
Catastro de parcelas	5	5	3	1	5	5	5	29
Edificios, incluyendo su nivel de protección	4	4	3	1	5	5	5	27
Suelo urbanizable y no urbanizable	4	5	3	1	5	5	5	29
Infraestructuras y redes de transporte	5	5	5	1	5	5	4	30

3.2 CONJUNTOS DE DATOS EN LA CATEGORÍA OBSERVACIÓN DE LA TIERRA Y MEDIO-AMBIENTE

Dataset	Valor reutilización interna	Valor comercial	Valor social	Singularidad	Enlazado	Coste mantenimiento	Facilidad API	Total
Calidad del aire	5	4	5	1	4	1	5	25
Calidad del agua	5	4	5	1	4	1	5	25
Contaminación acústica	5	3	5	1	5	1	5	25
Arbolado	4	3	5	1	5	4	5	27
Contenedores, puntos limpios (fijos y móviles), recogida de basuras	5	4	5	5	3	5	5	32
Áreas verdes, parques y jardines	5	2	5	2	5	4	5	28
Control de plagas	4	3	4	2	4	4	5	26
Hidrantes de agua potable y de incendios	5	3	5	5	5	5	5	33
Recogida de residuos peligrosos	5	3	3	3	5	4	4	27
Estaciones depuradoras	5	4	4	2	5	5	5	30
Consumo energético	5	5	5	3	4	1	4	27

Alumbrado público	5	4	5	1	4	3	5	27
-------------------	---	---	---	---	---	---	---	----

3.3 CONJUNTOS DE DATOS EN LA CATEGORÍA METEOROLOGÍA

Dataset	Valor reutilización interna	Valor comercial	Valor social	Singularidad	Enlazado	Coste mantenimiento	Facilidad API	Total
Estaciones de medición climatológica	5	4	4	3	5	1	5	27
Estaciones de medición de polen	5	4	5	4	5	1	5	29

3.4 CONJUNTOS DE DATOS EN LA CATEGORÍA ESTADÍSTICA (INCLUYENDO ECONOMÍA)

Dataset	Valor reutilización interna	Valor comercial	Valor social	Singularidad	Enlazado	Coste mantenimiento	Facilidad API	Total
Datos demográficos	5	5	4	4	4	3	5	30
Padrón	5	3	4	5	4	5	5	31
Censo electoral	5	3	3	5	4	5	5	30

Presupuestos municipales	5	3	5	5	4	2	5	29
Ejecución presupuestaria	5	2	5	5	4	2	5	28
Deuda	5	2	5	5	4	2	5	28

3.5 CONJUNTOS DE DATOS EN LA CATEGORÍA DE SOCIEDADES

Dataset	Valor reutilización interna	Valor comercial	Valor social	Singularidad	Enlazado	Coste mantenimiento	Facilidad API	Total
Catálogo de empresas	4	5	5	4	4	4	5	31
Censo de locales	5	5	4	5	4	4	5	32
Mercadillos y puestos municipales	5	5	5	5	4	5	5	34
Deudas de empresas	5	5	4	5	4	2	4	29
Subvenciones y ayudas a empresas	5	5	5	5	4	2	5	31
Normativa municipal para comercios y empresas	3	3	5	4	4	5	3	27
Actividad inspectora de Hacienda	5	5	4	5	4	4	3	34

Polígonos industriales	4	5	4	4	4	4	5	30
Arrendamientos: viviendas, locales comerciales, vehículos	5	5	5	5	4	4	4	32
Contratos y licitaciones entre Ayuntamiento y empresas privadas	5	5	5	5	4	3	4	31
Registro de asociaciones, fundaciones, colectivos	5	5	5	4	4	5	5	33
Licencias de terrazas y bares	5	5	5	5	4	4	5	33
Impuestos sobre personas jurídicas	5	5	4	5	4	3	3	29
Agentes económicos que pagan impuestos	5	5	5	4	4	5	5	33

3.6 CONJUNTOS DE DATOS EN LA CATEGORÍA MOVILIDAD

Dataset	Valor reutilización interna	Valor comercial	Valor social	Singularidad	Enlazado	Coste mantenimiento	Facilidad API	Total
Tráfico a tiempo real	5	4	5	1	5	1	5	26
Accidentalidad	5	3	5	1	5	2	5	26

Autobús interurbano: líneas, paradas, tiempo real, usuarios	5	3	5	5	5	4	4	31
Autobús urbano: líneas, paradas, tiempo real, usuarios	5	3	5	5	5	4	4	31
Aparcamiento, zonas reguladas	5	4	5	1	5	5	5	30
Señalización	5	3	4	1	5	5	5	28
Bicicleta pública	4	2	4	1	5	4	4	24
Infraestructura, estaciones, combustible	4	4	4	1	3	5	5	26
Parque de vehículos	4	4	2	5	2	4	5	26
Restricciones de circulación	4	2	2	4	3	3	4	22
Sanciones y multas de tráfico	4	2	3	4	2	4	5	24
Seguridad, semáforos, radares de tráfico	4	1	4	1	3	5	5	23
Licencias de obra de uso de vía pública	5	4	4	1	4	3	5	26

3.7 CONJUNTOS DE DATOS EN OTRAS CATEGORÍAS (SALUD, SEGURIDAD, CULTURA Y DEPORTE)

Dataset	Valor	Valor	Valor	Singularida	Enlazad	Coste	Facilidad	Total
---------	-------	-------	-------	-------------	---------	-------	-----------	-------

	reutilización interna	comercial	social	d	o	mantenimient o	API	
Áreas caninas	5	1	5	5	4	5	5	30
Licencias de tenencia de animales	5	1	3	5	4	5	5	28
Equipamientos municipales	5	3	4	5	4	5	5	31
Centros de salud, ubicación y horarios de apertura	4	4	5	1	5	5	5	29
Centros de emergencia sanitaria	4	4	5	1	5	5	5	29
Estaciones de bomberos	4	3	5	2	5	5	5	29
Comisarías de policía	4	3	5	2	5	5	5	29
Protección civil	4	3	5	2	5	5	5	29
Registro de delitos	5	4	5	4	4	2	3	27
Entradas en registros	5	4	5	4	4	2	3	27
Servicios sociales y atención al ciudadano	4	3	5	2	4	5	5	28
Puntos wi-fi	5	5	5	2	5	5	5	32
Agenda de eventos y actividades	4	3	5	2	4	2	5	25

Censo de bibliotecas públicas y su dotación	4	2	5	2	4	5	5	27
Catálogo de libros de bibliotecas	4	2	5	2	4	5	5	27
Préstamos de bibliotecas	4	2	5	2	4	5	5	27
Museos, galerías, exposiciones, salas de arte	4	2	5	2	4	4	5	26
Bienes inmuebles	4	2	5	2	4	4	5	26
Lugares de interés	4	2	5	2	4	4	5	26
Salas de ocio	4	2	5	2	4	4	5	26
Instalaciones deportivas	4	2	5	2	4	5	5	27
Centros de protección animal	4	2	5	2	4	4	5	26

