



DIPUTACIÓN DE CASTELLÓN

GUÍA DE NORMALIZACIÓN DEL MODELO DE DATOS

30 de abril de 2019

**Servicio de Administración e
Innovación Pública**



Contenido

1	Introducción	5
1.1	Definición de datos abiertos	6
1.2	Objeto de la guía	6
2	Normalización de datos	7
2.1	Principios generales	7
2.2	Generación de ficheros para su publicación con datos abiertos	7
2.3	Codificación de algunos tipos de datos	8
2.3.1	Datos de fechas.....	8
2.3.2	Codificación de años	8
2.3.3	Codificación de meses	8
2.3.4	Codificación de días	8
2.3.5	Codificación de días de la semana.....	8
2.3.6	Datos económicos	9
2.3.7	Provincias	9
2.3.8	Municipios.....	9
2.3.9	Núcleos de población	9
2.3.10	Representación de magnitudes.....	9
2.3.11	Codificación de coordenadas	9
2.3.12	Codificación de países	10
2.3.13	Distancias kilométricas	10
2.3.14	Descripción de regímenes de explotación de equipamientos	10
2.3.15	Codificación de género.....	10
2.3.16	Codificación de edad	10
3	Principios que rigen la gestión de datos.....	11
4	Muestreo de calidad de datasets.....	13

4.1	Máximos y mínimos.....	13
4.2	Nulos	13
4.3	Direcciones.....	13
4.4	Actualización.....	13
4.5	Muestreo de relación	14
4.6	Duplicados.....	14
4.7	Normalización aplicada	14
4.8	Máscaras aplicadas	15
4.9	Codificación errónea.....	15
4.10	Ámbito territorial	15
4.11	Datos personales.....	15
4.12	Seguridad	15
4.13	Auditoría	16
4.14	Instrucción de auditoría de privacidad y datos personales.....	16
5	Contratación	17
5.1	Contratación de desarrollo de aplicaciones.....	17
5.1.1	Contratación de soluciones y contratación de software as a service (SaaS).....	18
5.1.2	Subcontratación de servicios globales	18
5.2	Recomendaciones para los formatos técnicos.....	18
5.3	Cláusula Pliegos de Condiciones referentes a la transparencia y el libre acceso a la información del Ayuntamiento de Zaragoza “Cláusula Open Data”.....	19
6	Anexos I. Definiciones	20
7	Anexo II. Temáticas de información de acuerdo a NTI-RISP y de acuerdo a perfil de aplicación DCAT (DCAT-AP)	22

Índice de tablas

Tabla 1: Comparativa temáticas NTI-RISP / DCAT-AP 24

1 Introducción

La Diputación de Castellón tiene asignada la gestión y el soporte a más de 100 entidades municipales de la provincia. La actividad de gestión pública vinculada a un grupo de tales dimensiones ha llevado a los órganos de gobierno a plantear una política propia para con la gestión de los datos.

De cara a la publicación de datos abiertos la falta de un modelo estandarizado de gestión sobre los datos de la institución podría conllevar problemas múltiples que van desde las duplicidades en las tareas y los datos recogidos hasta el conocimiento interno de la información en custodia. Dicho de otra manera, la organización se encuentra sujeta a unas rutinas que no favorecen la eficiencia y, sin duda alguna, limitan la capacidad de garantizar los estándares de rapidez, transparencia y apertura que emanan de las exigencias de los vecinos, del marco normativo y de la sociedad digital actual.

Por todo esto la Diputación de Castellón planteó la necesidad de implementar pautas y criterios internos para instalar una gestión interna de los datos homogénea. Siendo esta guía el documento base a partir del cual se establecen los principios fundamentales para llevar a cabo esta tarea.

Con esta finalidad, los objetivos de esta guía se concretan en:

- Normalizar los principales datos tanto publicados como a publicar dentro del portal de datos abiertos y promover iniciativas para que su gestión interna sea también normalizada
- La interoperabilidad se considera el estadio inevitable para escalar las actuaciones gracias a establecer una sistemática y estándares comunes que reducen las barreras de acceso a los trámites y procedimientos, a la vez que incrementa la eficiencia de su gestión.

1.1 Definición de datos abiertos

En el marco de este nuevo modelo de gestión y relación entorno a lo público, los Datos Abiertos, según la Open Knowledge International, se definen como¹:

1 <http://opendatahandbook.org/guide/es/what-is-open-data/>

“Los datos abiertos son datos que pueden ser utilizados, reutilizados y redistribuidos libremente por cualquier persona, y que se encuentran sujetos, cuando más, al requerimiento de atribución y de compartirse de la misma manera en que aparecen.”

El término abierto, se define como un:

“Formato cuya especificación esté disponible públicamente y de manera gratuita, además, para su uso no se deben imponer restricciones de tipo monetario u otras”.

De modo, que el dato abierto para lograr ser reutilizado no debe tener limitaciones de licencia, derechos de autor o patentes, además de que no haya ninguna variable que limite su acceso. Siendo así, sus características principales son la disponibilidad, el acceso, la reutilización, la redistribución y la ausencia de restricción tecnológica.

1.2 Objeto de la guía

Si los consideramos como la materia prima del siglo XXI, los datos pueden resultar en una herramienta eficiente para crear servicios públicos de impacto. Sin embargo, la gestión basada en los datos es un campo relativamente reciente que necesita de organizaciones y equipos con el conocimiento y habilidades adecuadas para poner al descubierto el valor de los datos.

Considerando esta necesidad, el objeto de esta guía consiste en proporcionar las pautas necesarias para alcanzar una gestión eficiente de los datos en el conjunto de servicios involucrados en la publicación de datos abiertos que conforman la diputación de Castellón.

2 Normalización de datos

2.1 Principios generales

En general se acudirá a estandarizaciones internacionales y nacionales que lo sea por organizaciones sin ánimo de lucro y con gestión abierta (p.e. ISO, UNE, etc)

En caso de no existir una normalización aplicable se recomienda codificar la información e incluirla como una tabla dentro del portal de datos abiertos bien como tabla publicada o solo para uso interno.

En el caso de normalizaciones se recomienda codificación en Mayúsculas, mediante acrónimos no acentuados y sustituir los espacios por guiones. Evitar el uso de caracteres no ASCII.

Ej.

Normalización de tipos de asociaciones
organización no gubernamental → ONG

2.2 Generación de ficheros para su publicación con datos abiertos

Se recomienda solo tener una fila de cabecera con todos los nombres de los campos

No se deben incluir subtotaes o resúmenes en el medio de los datos

La codificación geográfica se incluirá en campos separados de latitud y longitud o bien en un único campo separado por una ',' ambas magnitudes. Codificación preferente en Sistema de referència WGS84, salvo que el GIS corporativo use otra codificación. En este caso utilizar para los nombres de campo LAT para latitud y LON para longitud.

Si las coordenadas son X, Y utilizar codificación ETRS89.

En el caso de ficheros de hojas de cálculo incluir una única pestaña

La separación preferida en los ficheros csv será por ',' para evitar problemas con las , incluidas en campos de texto.

Preferiblemente los campos de texto irán limitados por comilla doble

Los datos con porcentaje incluirán el símbolo %

En la medida de lo posible sustituir campos en texto libre por campos categorizados

2.3 Codificación de algunos tipos de datos

Esta es una tabla que progresivamente ha de hacerse crecer, está basado solo en los datos disponibles para el portal de datos abiertos.

2.3.1 Datos de fechas

Estandarización final mediante ISO 8601

para aquellas fechas en otros formatos durante la transición a la normalización, elegir preferentemente

AAAA-MM-DD

2.3.2 Codificación de años

Se codificará siempre con 4 cifras

Ej. 2019

2.3.3 Codificación de meses

Se codificará siempre con 2 cifras

Ej. Mes de marzo → 03

2.3.4 Codificación de días

Se codificará siempre con 2 cifras

Ej. Cuarto día del mes → 04

2.3.5 Codificación de días de la semana

Se codificarán completos

Lunes, Martes, Miércoles, Jueves, Viernes, Sábado y Domingo

2.3.6 Datos económicos

Representación con 3 letras mediante ISO 4217

2.3.7 Provincias

Nomenclatura según tabla INE de provincias

Disponible en : http://www.ine.es/daco/daco42/codmun/cod_provincia_estandar.htm

2.3.8 Municipios

Nomenclatura según tabla INE de municipios

<https://www.ine.es/daco/daco42/codmun/codmunmapa.htm>

2.3.9 Núcleos de población

Nomenclatura según tabla interna de núcleos de población disponible en portal de datos

<https://dipc.as.opendatasoft.com/explore/dataset/nucleos-de-poblacion/table>

2.3.10 Representación de magnitudes

La representación de magnitudes con datos decimales será de la forma:

EEE.DD. En otro campo se incluirán las unidades

E parte entera

D parte decimal

Separación mediante un .

Ej. 10.5

2.3.11 Codificación de coordenadas

Según normalización disponible en ARCGIS

2.3.12 Codificación de países

Según ISO ISO_3166-1

2.3.13 Distancias kilométricas

Codificar según magnitudes con las unidades en otro campo

2.3.14 Descripción de regímenes de explotación de equipamientos

- Público → PUB

- Privado → PRI

2.3.15 Codificación de género

Sexo : Hombre → H, Mujer → M, O → Otras opciones, N → No especificado

2.3.16 Codificación de edad

Se codificará siempre con la cifra completa con 3 cifras

Ej. Edad 46 años → 046

3 Principios que rigen la gestión de datos

Una gestión de datos ideal debe cumplir con unos determinados principios si queremos que dicha gestión produzca beneficios al aplicarla en la Diputación de Castellón.

Los principios serán los recogidos en la estrategia provincial de datos abiertos para la Diputación de Castellón y que se replican aquí.

- **Único:** Se evitará la duplicidad de los datos y documentos y se establecerán mecanismos de trabajo para evitar la generación de copias replicadas de la misma información.
- **Compartido:** La información deberá estar disponible de forma abierta por defecto, salvo aquella afectada por regulaciones específicas de privacidad o seguridad, en cuyo caso se procesará adecuadamente para estar en condiciones de ser publicada.
- **Accesible:** Se facilitará el acceso a los datos mediante los recursos técnicos disponibles y en general mediante la utilización de direcciones web persistentes y amigables.
- **Abierto:** Se utilizarán formatos abiertos, de uso libre para la gestión y puesta a disposición, siempre que sea posible, y en todo caso para la publicación y deberán cumplir la definición de estándar abierto según el Anexo, punto k) de la ley 18/2015, de modificación de la Ley 37/2017, sobre la reutilización de la información en el sector público. Los formatos deberán, además, ser apropiados para permitir el acceso de la ciudadanía y la reutilización de la información. En general no se recomienda el uso del formato pdf salvo que los datos ya estén disponibles en otro formato reutilizable.
- **Georreferenciado:** en la medida de lo posible siempre se incluirá la geolocalización del dato, incluyendo además de un sistema de coordenadas utilizable para su representación geográfica y de acuerdo a las directrices establecidas por la Directiva INSPIRE, y texto adicional para precisión del dato.
- **Documentado:** Los datos deberán estar explicados en su significado, rangos válidos y relación con otros juegos de datos. Deberá documentarse estas relaciones y definiciones. En la medida de lo posible se utilizarán metadatos y representaciones de los mismos para facilitar información adicional.
- **Licencia de uso abierta:** Los datos deberán incorporar información sobre las posibilidades de reutilización, en el caso de que esta no sea completa y cumpliendo la definición de conocimiento abierto. Sin perjuicio de que tenga

que cumplir los requisitos de la licencia tipo establecida en la citada Ley 18/2015 como la no desnaturalización de la información, la fecha de actualización, así como la exención de responsabilidad por el uso de la misma y la no adscripción de la misma. La Diputación se reserva el derecho al cambio de los términos de la licencia en cualquier momento, previo aviso de tres meses.

- **Tiempo real:** En la medida de lo posible y de las características de generación de los datos, éstos se harán disponibles en la mayor brevedad y en tiempo real si es posible.
- **Automatizable:** Se evitarán todas aquellas recogidas de datos cuyo resultado no pueda ser automatizable, bien en su acceso o en su reutilización tanto interna como externa.
- **Fiable:** Incluir mecanismos dentro de las agrupaciones y procesados de los datos que permitan verificar la autoría y la exactitud del dato respecto de la fuente original.
- **Adaptado al uso:** *Se deberá recoger información de cómo se utilizan los distintos datos para confirmar que el formato utilizado es el adecuado y cómo posibles mejoras lo harían más utilizable tanto de forma interna como externa a la organización.*

4 Muestreo de calidad de datasets

Se incluyen aquí algunos elementos para comprobar la calidad mediante muestreo de datasets de cara a su publicación, requiere tener los criterios de conformidad para los distintos tipos de datos que compongan el dataset.

4.1 Máximos y mínimos

Identificar los valores máximos y mínimos de todas las columnas. Contrastar con el responsable del dataset si dichos valores encajan en los valores posibles.

En caso de no encajar, documentar las posibles causas y registrarlo para la generación de micro proyectos de mejora.

4.2 Nulos

Realizar en todas las columnas del dataset una búsqueda de valores nulos y/o errores. Contrastar con el responsable del dataset si dichos valores encajan en los valores posibles.

En caso de no encajar, documentar las posibles causas y registrarlo para la generación de micro proyectos de mejora.

4.3 Direcciones

En el caso de que tenga direcciones ordenar el dataset por el campo de dirección. Buscar duplicados codificados de diferente forma P.e Avinguda / av/ etc.

En caso de que impacte en la gestión del dataset documentar las posibles causas y registrarlo para la generación de micro proyectos de mejora.

4.4 Actualización

En el caso de que exista un campo de actualización buscar el valor más antiguo posible. Contrastar con el responsable si dicha fecha se corresponde con los procedimientos internos.

En caso de que impacte en la gestión del dataset documentar las posibles causas y registrarlo para la generación de micro proyectos de mejora.

4.5 Muestreo de relación

En el caso de que el dataset tenga relación con otros elegir una muestra de registros (5 + número de ceros del numero de registros del dataset)

Ej. si hay 1300 registros. $1000 (10^3) < 1300 < 10000 (10^4)$ → será 5 (termino fijo) + 4

ej. si hay 230.000 registros. $100.000 (10^5) < 230.000 < 1.000.000 (10^6)$ → será 5 (termino fijo) + 6

Comprobar que los campos que establecen la relación con otros datasets se corresponden con los que se encuentran en aquellos datasets.

En caso de detectar discrepancias intentar averiguar la causa, documentar las posibles causas y registrarlo para la generación de micro proyectos de mejora.

4.6 Duplicados

En el caso de que el dataset no permita duplicados ordenar por diversos criterios el dataset y verificar que no existen duplicados.

En caso de detectar discrepancias intentar averiguar la causa, documentar las posibles causas y registrarlo para la generación de micro proyectos de mejora.

NOTA: una forma sencilla de comprobar este caso en el caso de datasets de una única tabla es crear una tabla dinámica y agrupar y contar por cada campo.

4.7 Normalización aplicada

En el caso de que en el cuestionario se declare que alguno de los campos está normalizado, realizar un muestreo (con el protocolo definido en 4.5) y comprobar el cumplimiento de la normalización.

En caso de detectar discrepancias intentar averiguar la causa, documentar las posibles causas y registrarlo para la generación de micro proyectos de mejora.

4.8 Máscaras aplicadas

En el caso de que el dataset declare que existe algún mecanismo para las prevención de introducción de datos erróneos (p.e. máscaras de fechas, mascarar de formato) comprobar mediante muestreo que se cumple dicho protocolo.

En caso de detectar discrepancias intentar averiguar la causa, documentar las posibles causas y registrarlo para la generación de micro proyectos de mejora.

4.9 Codificación errónea

En el caso de que existan campos de texto realizar una búsqueda por caracteres no imprimibles o provenientes de una conversión de textos errónea. (p.e. & por &, caracteres > o < provenientes de etiquetas html, o bien de conversión de letras acentuadas).

En caso de detectar discrepancias intentar averiguar la causa, documentar las posibles causas y registrarlo para la generación de micro proyectos de mejora.

4.10 Ámbito territorial

Respecto al ámbito territorial. Cualquier referencia a direcciones del territorio deberá quedar codificada de acuerdo con el sistema territorial que mantiene la cartografía municipal.

4.11 Datos personales

Respecto al ámbito de los datos de identificación personal. Cualquier referencia a datos de identificación personal contará con una estandarización a la hora de definir los datos.

4.12 Seguridad

Para garantizar la actualización y la correcta utilización de las bases de datos, aplicaciones y accesos del sistema de información corporativo será necesario respetar y cumplir el circuito de seguridad y acceso recogido en los procedimientos de la Diputación.

A tal efecto, el departamento de sistemas de información requiere de la cooperación efectiva de los distintos servicios para garantizar que el acceso a los datos con requisitos de seguridad cumple las normativas internas. Periódicamente se lanzarán auditorías de seguridad o procesos de actualización para garantizar la actualización de los datos.

4.13 Auditoría

Periódicamente se ejecutarán auditorías de seguridad de información, de protección de datos personales, y de actualización de los permisos de acceso a los distintos conjuntos de datos.

Los errores, inconsistencias de la información, infrautilización y utilización errónea detectados en los procesos de depuración serán comunicados a los departamentos responsables de las aplicaciones y bases de datos para su solución.

Los servicios deberán dar respuesta a los requerimientos de los procesos de depuración en un plazo no superior a seis meses desde su comunicación por parte de la auditoría.

Para la ejecución de esta auditoría pueden ser de utilidad las recomendaciones del esquema nacional de seguridad.²

4.14 Instrucción de auditoría de privacidad y datos personales

Cualquier conjunto de datos, aplicación, archivo informático y formulario web que contengan datos de carácter personal deberá estar convenientemente registrado según la Ley Orgánica de Protección de Datos de Carácter Personal (LOPD). Hay que recordar que cada dos años en la Diputación se debe realizar una auditoría al respecto.

A partir del momento de aplicación de esta guía en los servicios del proyecto piloto la incorporación de nuevos juegos de datos habrá de cumplir el protocolo descrito en puntos anteriores.

A medida que se vaya extendiendo su aplicación a más servicios será de aplicación obligatoria su incorporación al inventario de datos con los requisitos de seguridad y protección de datos ya incorporados.

² Guía de seguridad (CCN-STIC-802). Esquema Nacional de seguridad. guía de auditoría https://www.ccn-cert.cni.es/publico/seriesCCN-STIC/series/800-Esquema_Nacional_de_Seguridad/802-Auditoria_ENS/802-Auditoria_ENS-jun10.pdf

5 Contratación

Las aplicaciones externas deberán cumplir los mismos requerimientos en materia de gestión de datos que el resto de sistemas de la Diputación de Castellón.

Para adquirir cualquier aplicación de terceros será necesario el visto bueno del departamento de sistemas de información para garantizar la perfecta integración de la misma con la plataforma tecnológica, las integraciones necesarias con otras aplicaciones o módulos y del departamento de sistemas de información por sí con posterioridad es necesaria su explotación a través de aplicaciones corporativas de la Diputación y requieren relacionarse con otros datos.

Estas aplicaciones también deberán contar con un responsable y/o interlocutor y estar perfectamente catalogadas, siguiendo los criterios que determine el servicio de sistemas de información. Dentro de esta catalogación se incluirán las aplicaciones externas ya existentes. A tal efecto, el departamento de sistemas de información determinará y establecerá la coordinación necesaria para poder proceder a registrar y documentar estas bases de datos. Los departamentos municipales que dispongan de aplicaciones externas deberán facilitar la información que les sea requerida.

En caso de contar con datos de carácter personal deberán cumplir las mismas exigencias recogidas en esta instrucción. El responsable de la aplicación externa será el encargado de seguir el procedimiento establecido en la institución para garantizar el cumplimiento de la LOPD.

Se incluye a continuación recomendaciones específicas para cada entorno de contratación.

5.1 Contratación de desarrollo de aplicaciones

En la contratación de aplicaciones se incluirá dentro de los requisitos de la documentación a elaborar:

- Documentación de los principales conjuntos de datos, definiendo los campos que los componen, su tipo de datos y los valores válidos del mismo, así como los mecanismos incluidos para la prevención de introducción de valores.
- En la elaboración del modelo de datos se coordinará con el servicio de gestión de datos de la Diputación para la validación de los campos y su conexión con campos existentes en el inventario de datos, y, en su momento, con el diccionario de datos.

5.1.1 Contratación de soluciones y contratación de software as a service (SaaS)

En la contratación de soluciones y SaaS se deberá requerir:

- 1) En caso de que la personalización de la solución determine nuevos campos de datos y estos sean configurables deberán evaluarse conforme a los requisitos de se establecen para la contratación de desarrollo de aplicaciones.
- 2) En caso de que la personalización no permita nuevos campos de datos o éstos no sean configurables, deberá solicitarse en la documentación la relación de los mismos junto con sus rangos y valores (documentación suficiente para su inclusión en el diccionarios de datos).

5.1.2 Subcontratación de servicios globales

Debe incluirse una cláusula de cesión de datos, con las prevenciones normalizadas para los datos afectados por LOPD y otras regulaciones, donde se establezca la cesión automática y en tiempo real de los mismos así como la licencia para que la Diputación pueda publicar los datos como datos abiertos

5.2 Recomendaciones para los formatos técnicos

Los formatos a utilizar deben permitir la interconexión entre distintos conjuntos de datos, y por tanto, en la medida de lo posible deben ser abiertos, estructurados y no deben requerir el uso de herramientas específicas para su acceso.

Sin embargo, también somos conscientes que actualmente en la Diputación de Castellón hay otros formatos que, aunque no sean abiertos, su uso es muy extendido. Por ejemplo este es el caso del formato Excel para las hojas de cálculo.

Ejemplos de formatos que sugerimos utilizar:

- csv (comma separated values) (para hojas de cálculo en formato texto)
- txt (texto)
- ods open document spreadsheet format (para hojas de cálculo)
- rdf (resources description framework)
- xml (extensible markup language)

5.3 Cláusula Pliegos de Condiciones referentes a la transparencia y el libre acceso a la información del Ayuntamiento de Zaragoza “Cláusula Open Data”

Se incluye a continuación esta cláusula como ejemplo de referencia para su aplicación en la Diputación de Castellón.

1. En virtud de lo dispuesto en la Ordenanza sobre transparencia y libre acceso a la información, el Ayuntamiento de Zaragoza podrá publicar o poner a disposición de quien la solicite toda la información relativa a la presente licitación, con la única excepción de la información técnica aportada por las empresas licitadoras que quede cubierta por el secreto comercial.

2. El Ayuntamiento de Zaragoza podrá solicitar a la empresa adjudicataria cualquier información relativa al objeto del contrato y la las circunstancias de su ejecución cuando ésta sea de interés para los ciudadanos, debiendo la empresa facilitarla en un formato apropiado y en el plazo máximo de una semana, salvo que por su volumen o complejidad se justificara su ampliación. Si la empresa considera que es de aplicación alguna de las limitaciones a la publicidad previstas en la Ordenanza sobre transparencia y libre acceso a la información podrá alegarlo ante el órgano competente en materia de publicidad de la información, que resolverá en plazo de tres días hábiles. Esta obligación subsistirá durante los dos años posteriores a la finalización de las obligaciones principales del contrato.

Cláusula adicional para los contratos que incluyan la concesión de servicios públicos de titularidad municipal o la delegación de potestades administrativas

3. La empresa adjudicataria deberá proporcionar, a lo largo de todo el periodo de ejecución del contrato, la información relativa a la prestación de servicios públicos o al ejercicio de potestades administrativas delegadas que el Ayuntamiento considere que ha de ser publicada de acuerdo con lo dispuesto en la Ordenanza sobre transparencia y libre acceso a la información. Se facilitará acceso a la información en tiempo real y se cumplirá con los criterios de calidad establecidos en la normativa municipal. Según se disponga en el pliego de prescripciones técnicas, la publicación en Internet podrá ser realizada por la empresa adjudicataria, por la web municipal o por ambas. En el citado pliego se especificarán asimismo la estructura de los datos y los formatos a utilizar para la publicación y/o comunicación de los datos, debiendo coordinarse el adjudicatario con el servicio responsable de la Web Municipal para la implementación y puesta en marcha de los mecanismos técnicos que resulten precisos.

6 Anexos I. Definiciones

Dato: Según la Wikipedia³ un dato es “*una representación simbólica (numérica, alfabética, algorítmica, espacial, etc.) de un atributo o variable cuantitativa o cualitativa. Los datos describen hechos empíricos, sucesos y entidades*”. Dicho de otra forma, un dato es un valor puntual, sin contexto y sin derecho a propiedad intelectual dado que para obtenerlo/generarlo no es necesario un “esfuerzo intelectual”. Un ejemplo de datos es el valor que nos da un termómetro: 23°C.

Información: Un dato por sí mismo (sin contexto) no aporta mucho, es cuando relacionamos estos datos cuando llegamos a la información que es el paso previo para poder transmitir/recibir nuestro conocimiento. Aquí sí que rige el derecho a propiedad intelectual. La información está relacionada con el formato documento, ello implica que la información no es el objetivo de esta Guía.

Dataset: Es una palabra inglesa que significa literalmente “conjunto de datos”. Es muy utilizada en la gestión de datos para poder referirnos a conjuntos de datos con independencia del formato de éstos, su tamaño, su representación física, etc. Ejemplo, un fichero en Excel es un *dataset*, un fichero en Access también es un *dataset*, etc.

Base de datos: Es un conjunto de datos, con lo que se podría entender que es un sinónimo de *dataset*. Sin embargo, fruto del éxito de las bases de datos relacionales⁴, a menudo se confunde el término “base de datos” con el subconjunto “bases de datos relacionales”, por ello recientemente se utiliza más la palabra “Dataset” para referirse a un conjunto de datos.

Fuente de datos: Sistema u organización que genera datos. Cada dataset tiene una fuente de datos (como mínimo una).

Metadatos: Son aquellos datos que describen nuestros datos. Ejemplo: Podemos tener un dataset con los datos relacionados con los presupuestos municipales. El contenido de este dataset serán datos de presupuestos. Pero hay otros datos que

3 <https://es.wikipedia.org/wiki/Dato>

4 https://es.wikipedia.org/wiki/Base_de_datos_relacional

describen este dataset: el responsable de dicho dataset, la fecha de creación, la fecha de última actualización, la frecuencia de actualización de los datos, si contiene datos de carácter personal, en qué formato de disponen los datos, etc.

Formato de los datos: En qué sistema están representados los datos. Por ejemplo, podemos tener los datos en formato MS-Excel (.xls), en formato CSV (fichero de texto separado en comas), en formato semántico como es el RDF, etc.

7 Anexo II. Temáticas de información de acuerdo a NTI-RISP y de acuerdo a perfil de aplicación DCAT (DCAT-AP)

Identificadores temáticos DCAT-AP	Identificadores temáticos NTI-RISP
Ciencia-tecnología	Ciencia-tecnología
Agricultura, pesca, forestal y alimentación	Medio-rural-pesca
Economía y finanzas	Economía Comercio Hacienda
Educación, cultura y deporte	Educación Cultura-ocio Deporte
Energía	Energía
Medio ambiente	Medio-ambiente
Gobierno y sector público	Sector-publico
Salud	Salud
Asuntos internacionales	
Justicia, sistema legal y seguridad	Legislacion-justicia Seguridad
Regiones y ciudades	Urbanismos-infraestructuras Vivienda
Población y sociedad	Demografía Sociedad-bienestar Empleo
Transporte	Transporte
	Industria Turismo

Tabla 1: Comparativa temáticas NTI-RISP / DCAT-AP

Se recomienda clasificar a primer nivel los datos de acuerdo a la NTI-RISP y matizar o referenciar con más detalle a través de las etiquetas de cada conjunto de datos.