|  |
| --- |
| [Escribir el nombre de la compañía] |
| Acceso, transformación y explotación de datos abiertos sobre contrataciones públicas en el Ayuntamiento de Madrid |
|  |
|  |
| **Raul A. del Águila Escobar** |
| **Febrero de 2017** |

|  |
| --- |
| Este trabajo se ha realizado utilizando los datos abiertos del Ayuntamiento de Madrid. |

# Introducción

Como recoge la Open Knowledge Foundation[[1]](#footnote-2), los datos abiertos son una pieza fundamental para la transparencia en las organizaciones. La transparencia implica que un ciudadano pueda ejercer el control sobre las acciones de los gobiernos e instituciones mediante al análisis de los datos que éstas han publicado. Es, por tanto, una buena base para comenzar la lucha contra el fraude.

El fraude implica engaño. La voluntad de ocultar una acción deshonesta para con la organización o con la sociedad es, sin duda, una de las principales características de quien perpetra un fraude.

Tras la reciente crisis económica, los ciudadanos, empresas y agentes sociales han tomado conciencia del grave impacto financiero (en adelante, perjuicio económico) que el fraude puede ocasionar a organizaciones privadas y gobiernos; el fraude es, sin duda, un ataque directo a la cuenta de resultados de cualquier tipo de organización.

Sin embargo, la lucha contra el fraude no es algo nuevo:

Instituciones como la ACFE (Association of Certified Fraud Examiners) llevan formando y elaborando estándares de facto en lo relativo a procedimientos y tipología de fraude.

Existen firmas de servicios profesionales tienen departamentos especializados en la investigación y prevención de fraude como, por ejemplo, EY. Estas firmas utilizan procedimientos de análisis financiero, social network analytics, o minería de datos, entre otros, en el contexto de grandes investigaciones de fraude.

Existen normativas como la FCPA americana o la UK Bribery Act británica que establecen sanciones para todas aquellas organizaciones que no hayan puesto los medios necesarios para prevenir la corrupción. A nivel nacional, las sucesivas reformas del Código Penal español desde el 2010 se han ido alineando con la normativa internacional, llegando la jurisprudencia nacional a establecer como requerimiento la existencia de controles informáticos para dicha prevención.

La adjudicación de contratos es una de las principales áreas de análisis en la lucha contra el fraude. El presente trabajo tiene por objeto el acceso a una fuente de datos abiertos relacionadas con la adjudicación de datos, su transformación de acuerdo con un modelo ontológico, su publicación y un ejemplo de explotación de estos datos.

La explotación de estos datos se realizará de acuerdo a la búsqueda de patrones de fraude sencillos. En cualquier caso, las conclusiones que se obtengan del análisis de estos datos deben considerarse como una aproximación académica y su finalidad no es la conclusión sobre la eventual existencia de fraude, aspecto para lo que se deberían tener en cuenta otras fuentes de datos (alta de proveedores, procedimientos de licitación, tesorería, entre otros) y realizarse procedimientos de muy diversa naturaleza (análisis forense, entrevistas formales, etc).

# Fuente de datos abiertos

Se ha accedido al portal de datos abiertos del Ayuntamiento de Madrid[[2]](#footnote-3) y se ha accedido a los datos contratos inscritos en 2016[[3]](#footnote-4). La imagen siguiente contiene una captura de pantalla de esta página web, accedida el 9 de enero de 2018.

|  |
| --- |
|  |

Imagen Captura de pantalla del acceso a la página web de datos abiertos del gobierno americano

En este sentido, los datos que vamos a descargar, enlazar y explotar son los relativos a los contratos ***menores*** contratados por el Ayuntamiento de Madrid.

En el contexto de la contratación de carácter público y de acuerdo con la Ley de Contratos del Sector Público[[4]](#footnote-5) los contratos menores son aquellos inferiores a 50.000, si son contratos de obras, ó 18.000 euros si son contratos de servicios o suministros (importes sin IVA). Es necesario señalar que la si bien la última modificación de la Ley es de noviembre de 2017, la versión vigente en el momento de las contrataciones recogidas en el fichero que contiene los contratos menores la aprobada a través de Real Decreto Legislativo 3/2011[[5]](#footnote-6).

En cuanto a la aprobación y tramitación del expediente, la citada Ley recoge que en los contratos menores la tramitación del expediente sólo exigirá la aprobación del gasto y la incorporación al mismo de la factura correspondiente, que deberá reunir los requisitos que las normas de desarrollo de esta Ley establezcan; en el caso del contrato menor de obras, deberá añadirse, además, el presupuesto de las obras, sin perjuicio de que deba existir el correspondiente proyecto cuando normas específicas así lo requieran el trabajo afecte a la estabilidad, seguridad o estanqueidad de la obra.

Este aspecto es muy relevante. Un típico patrón de fraude a analizar es la fragmentación de contratos, mediante sucesivos contratos, novaciones o renovaciones de los mismos, con objeto de no tener que pasar los requerimientos establecidos para contratos no menores, en concreto, la obligatoriedad de establecer un proceso licitador.

En cuanto a la contratación, de acuerdo con el citado Real Decreto, los contratos menores podrán adjudicarse directamente a cualquier empresario con capacidad de obrar y que cuente con la habilitación profesional necesaria para realizar la prestación, cumpliendo con las normas que se acaban de describir.

Una vez expuesta la fuente de datos que vamos a analizar y adelantado el objeto posterior del análisis, pasamos a describir los datos

## Descripción de los datos

La web desde la que se ha descargado el catálogo de datos, no especifica el formato de cada uno de los ficheros. La única descripción que recoge de los datos es la siguiente:

*"Contratos menores contabilizados en 2017: listado generado a 15/12/2017, diferenciando los contratos menores que fueron aprobados y contabilizados el año 2016 y los que tras la carga de plurianuales tienen gasto en 2017, de los generados a partir del 1 de enero de 2017 (estos últimos tendrán vacío el campo "Documento 2016")."*

Dado que no se dispone de un descriptivo de los datos, es necesario realizar un análisis en profunidad sobre los mismos. Si bien se puede utilizar OpenRefine para realizar el análisis de los datos, en vez de utilizar esta herramienta para realizar este análisis descriptivo inicial se ha utilizado un script SQL que analiza la estructura, rango y valores de las tablas. A continuación se presenta el resultado de este análisis:

Tabla Análisis de los datos: descripción

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Tipo de dato** | | **Valor mínimo** | | **Valor máximo** | **Valores únicos** | **Número de nulos** | **Número cercanos a los límites** |
| Importe | float | 0 | | 60.428 | | 3.773 | 0 | 0 |
| Ce gestor | float | 1002 | | 509.087 | | 41 | 0 | 83 |
| Fe contab | datetime | 17/01/2017 | | 30/11/2017 | | 215 | 0 | NULL |
| Fech apro | datetime | 17/01/2017 | | 30/11/2017 | | 217 | 0 | NULL |
| Nº expediente | float | 5031500510 | | 14520170006194 | | 3.260 | 1.438 | 2 |
| Descripción | nvarchar | 5 | | 40 | | 41 | 0 | 199 |
| Documento 2016 | nvarchar | 10 | | 10 | | 2 | 4.248 | 488 |
| NIF | nvarchar | 9 | | 13 | | 2.128 | 20 | 4732 |
| Tercero | nvarchar | 5 | | 108 | | 2.150 | 0 | 3 |
| Tipo de expediente | nvarchar | 16 | | 40 | | 6 | 0 | 291 |
| Título del expediente | nvarchar | 8 | | 89 | | 4.559 | 0 | 11 |

Tabla Análisis de los datos: distribución de valores

| **Variable** | **Valor** | **% sobre total** |
| --- | --- | --- |
| [Ce#gestor] | 1067 | 11,44 |
| [Ce#gestor] | 508067 | 10,16 |
| [Ce#gestor] | 1027 | 8,55 |
| [Ce#gestor] | 1098 | 8,26 |
| [Ce#gestor] | 1214 | 5,05 |
| [Ce#gestor] | 1097 | 4,79 |
| [Ce#gestor] | 1057 | 3,91 |
| [Ce#gestor] | 503027 | 3,86 |
| [Ce#gestor] | 1209 | 2,70 |
| [Ce#gestor] | 501016 | 2,58 |
| [Ce#gestor] | 1016 | 2,47 |
| [Ce#gestor] | 506057 | 2,30 |
| [Ce#gestor] | 1205 | 2,03 |
| [Ce#gestor] | 1217 | 1,90 |
| [Ce#gestor] | 1208 | 1,86 |
| [Ce#gestor] | 1204 | 1,82 |
| [Ce#gestor] | 1210 | 1,79 |
| [Ce#gestor] | 1002 | 1,60 |
| [Ce#gestor] | 1087 | 1,60 |
| [Ce#gestor] | 1215 | 1,54 |
| [Ce#gestor] | 1203 | 1,48 |
| [Ce#gestor] | 1206 | 1,48 |
| [Ce#gestor] | 1220 | 1,44 |
| [Ce#gestor] | 1201 | 1,41 |
| [Ce#gestor] | 1047 | 1,33 |
| [Ce#gestor] | 1077 | 1,27 |
| [Ce#gestor] | 1202 | 1,22 |
| [Ce#gestor] | 1012 | 1,20 |
| [Ce#gestor] | 1037 | 1,08 |
| [Ce#gestor] | 1218 | 1,03 |
| [Ce#gestor] | 1211 | 1,01 |
| [Ce#gestor] | 1213 | 0,9923 |
| [Ce#gestor] | 1221 | 0,8868 |
| [Ce#gestor] | 1212 | 0,8023 |
| [Ce#gestor] | 1216 | 0,6967 |
| [Ce#gestor] | 1207 | 0,6545 |
| [Ce#gestor] | 1120 | 0,5701 |
| [Ce#gestor] | 1028 | 0,5489 |
| [Ce#gestor] | 1219 | 0,5067 |
| [Ce#gestor] | 509087 | 0,1478 |
| [Ce#gestor] | 1010 | 0,0211 |
| [Descripción] | SALUD, SEGURIDAD Y EMERGENCIAS | 11,44 |
| [Descripción] | MADRID SALUD | 10,16 |
| [Descripción] | EQUIDAD, DERECHOS SOCIALES Y EMPLEO | 8,55 |
| [Descripción] | CULTURA Y DEPORTES | 8,26 |
| [Descripción] | DISTRITO DE MORATALAZ | 5,05 |
| [Descripción] | MEDIO AMBIENTE Y MOVILIDAD | 4,79 |
| [Descripción] | ECONOMÍA Y HACIENDA | 3,91 |
| [Descripción] | AGENCIA PARA EL EMPLEO DE MADRID | 3,86 |
| [Descripción] | DISTRITO DE MONCLOA-ARAVACA | 2,70 |
| [Descripción] | INFORMÁTICA DEL AYUNTAMIENTO DE MADRID | 2,58 |
| [Descripción] | GERENCIA DE LA CIUDAD | 2,47 |
| [Descripción] | AGENCIA TRIBUTARIA MADRID | 2,30 |
| [Descripción] | DISTRITO DE CHAMARTÍN | 2,03 |
| [Descripción] | DISTRITO DE VILLAVERDE | 1,90 |
| [Descripción] | DISTRITO DE FUENCARRAL-EL PARDO | 1,86 |
| [Descripción] | DISTRITO DE SALAMANCA | 1,82 |
| [Descripción] | DISTRITO DE LATINA | 1,79 |
| [Descripción] | DESARROLLO URBANO SOSTENIBLE | 1,60 |
| [Descripción] | PLENO | 1,60 |
| [Descripción] | DISTRITO DE CIUDAD LINEAL | 1,54 |
| [Descripción] | DISTRITO DE RETIRO | 1,48 |
| [Descripción] | DISTRITO DE TETUÁN | 1,48 |
| [Descripción] | DISTRITO DE SAN BLAS-CANILLEJAS | 1,44 |
| [Descripción] | DISTRITO DE CENTRO | 1,41 |
| [Descripción] | PORTAVOZ, COORD. JUNTA GOB. Y REL. PLENO | 1,33 |
| [Descripción] | PARTICIP.CIUDADANA,TRANSP. Y GOB.ABIERTO | 1,27 |
| [Descripción] | DISTRITO DE ARGANZUELA | 1,22 |
| [Descripción] | COORDINACIÓN GENERAL DE LA ALCALDÍA | 1,20 |
| [Descripción] | COORD. TERRITORIAL Y COOP. PÚBLICO-SOC. | 1,08 |
| [Descripción] | DISTRITO DE VILLA DE VALLECAS | 1,03 |
| [Descripción] | DISTRITO DE CARABANCHEL | 1,01 |
| [Descripción] | DISTRITO DE PUENTE DE VALLECAS | 0,9923 |
| [Descripción] | DISTRITO DE BARAJAS | 0,8868 |
| [Descripción] | DISTRITO DE USERA | 0,8023 |
| [Descripción] | DISTRITO DE HORTALEZA | 0,6967 |
| [Descripción] | DISTRITO DE CHAMBERÍ | 0,6545 |
| [Descripción] | TRIBUNAL ECONÓMICO-ADMINISTRATIVO | 0,5701 |
| [Descripción] | POLÍTICAS DE GÉNERO Y DIVERSIDAD | 0,5489 |
| [Descripción] | DISTRITO DE VICÁLVARO | 0,5067 |
| [Descripción] | AGENCIA DE ACTIVIDADES | 0,1478 |
| [Descripción] | ALCALDÍA | 0,0211 |
| [Documento 2016] | NULL | 89,70 |
| [Documento 2016] | ANTICIPADA | 5,30 |
| [Documento 2016] | PLURIANUAL | 5,00 |
| [Tipo de expediente] | Contrato de servicios | 47,19 |
| [Tipo de expediente] | Contrato de suministros | 44,66 |
| [Tipo de expediente] | Contrato privado | 6,12 |
| [Tipo de expediente] | Contrato de obras | 1,98 |
| [Tipo de expediente] | Contrato administrativo especial | 0,0211 |
| [Tipo de expediente] | Contrato de gestión de servicios público | 0,0211 |

Como se puede observar existen muchas observaciones que no disponen de datos y otros con una frecuencia de aparición muy baja que, con independencia de la connotación estadística de los mismos, podrían considerarse como anómalos.

En el contexto de una investigación o análisis de fraude, **no** se deben alterar los datos recibidos por el cliente. Es decir, se ha de trabajar con esos datos aunque la calidad no sea adecuada. Por ejemplo, en el caso de datos con valor nulo, la gestión de esos datos debe realizarse durante la explotación y la incidencia podría devenir en una recomendación sobre calidad (dado que esas ineficiencias pueden ser aprovechadas para la comisión de un fraude).

Asimismo, este primer análisis ya nos ofrece información muy rica sobre los datos que se han de analizar. Por ejemplo, se puede comprobar que la mayor parte de los contratos menores se han concedido en el ámbito de la Salud. Asimismo, los contratos de Servicios y Suministros suponen la práctica totalidad de los contratos seleccionados.

Por otro lado, **no existen valores únicos** que sirvan como identificador. Así, por ejemplo, el expediente *115/2017/01534* aparece 30 veces, con diferente descripción.

Asimismo, en lo relativo a los importes cercanos a los límites (tabla 1) no existen muchos contratos que sean cercanos al valor del importe máximo o mínimo. Esto ya es un primer indicador de una posible irregularidad, esto es, valores extremos de importe justo por debajo del límite de contratación.

En cuanto al rango y unicidad de los datos, se pueden observar en la siguiente tabla:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **valor mínimo** | **valor máximo** | **Valores únicos** |
| [Año] | 2016 | 2016 | 1 |
| [Fecha Adjudicación] | 23/02/2015 | 29/12/2016 | 283 |
| [Fecha Formalización] | 16/03/2015 | 30/12/2016 | 274 |
| [Importe Adjudicación (IVA Incluido)] | 0 | 250.311.011 | 1.031 |
| [Número Contrato] | 201.600.001 | 201.601.154 | 1.132 |
| [Plazo] | 0 | 300 | 257 |
| [Presupuesto Total (IVA Incluido)] | 0 | 253.043.885 | 963 |
| [Acuerdo Marco] |  |  | 2 |
| [Apartado] | N/A | N/A | 4 |
| [Artículo] | N/A | N/A | 2 |
| [Criterios Adjudicación] | N/A | N/A | 2 |
| [Descripción Centro] | N/A | N/A | 2 |
| [Descripción Contrato] | N/A | N/A | 1.109 |
| [Ingreso/Coste Cero] | N/A | N/A | 1 |
| [Mes] | N/A | N/A | 12 |
| [NIF/CIF Adjudicatario] | N/A | N/A | 695 |
| [Nombre/Razón Social] | N/A | N/A | 706 |
| [Número Expediente] | N/A | N/A | 888 |
| [Observaciones] | N/A | N/A | 7 |
| [Organismo] | N/A | N/A | 44 |
| [Procedimiento Adjudicación] | N/A | N/A | 5 |
| [Tipo Contrato] | N/A | N/A | 5 |

Se puede observar que solamente el **Número de Contrato** contiene 1.132 observaciones únicas. Por tanto, éste será el número que utilizaremos como identificador unívoco en nuestra ontología.

Una vez que ya hemos accedido e inspeccionado nuestros datos, vamos a trabajar sobre los vocabularios formalizados u ontologías que pueden representar estos datos para su publicación y explotación posterior.

## Licencia

No existe especificación de licencia recogida en la página Web del Ayuntamiento de Madrid. No obstante lo anterior, sí especifica la página web[[6]](#footnote-7) las condiciones de uso de los datos[[7]](#footnote-8). Las características más relevantes son:

**"(...)**

**Autorización de reutilización y cesión no exclusiva de derechos de propiedad intelectual**

Las presentes condiciones generales permiten la reutilización de los **documentos y datos sometidos a ellas para fines comerciales y no comerciales**. (...). Por ejemplo, la reutilización autorizada incluye actividades como la copia, difusión, modificación, adaptación, extracción, reordenación y combinación de la información.  
(...)  
Esta autorización conlleva, asimismo, la cesión gratuita y no exclusiva de los derechos de propiedad intelectual, en su caso, correspondientes a tales documentos, autorizándose la realización de actividades de reproducción, distribución, comunicación pública o transformación, necesarias para desarrollar la actividad de reutilización autorizada, en cualquier modalidad y bajo cualquier formato, para todo el mundo y por el plazo máximo permitido por la Ley.

(...)

**Condiciones generales para la reutilización**

Son de aplicación las siguientes condiciones generales para la reutilización de los documentos sometidos a ellas:

1. Está prohibido desnaturalizar el sentido de la información.
2. Debe citarse la fuente de los documentos objeto de la reutilización. Esta cita podrá realizarse de la siguiente manera: "Origen de los datos: Ayuntamiento de Madrid (o, en su caso, órgano administrativo, organismo o entidad de que se trate)"
3. Debe mencionarse la fecha de la última actualización de los documentos objeto de la reutilización, siempre y cuando estuviera incluida en el documento original.
4. No se podrá indicar, insinuar o sugerir que el Ayuntamiento de Madrid participa, patrocina o apoya la reutilización que se lleve a cabo con la información.
5. Deben conservarse, no alterarse ni suprimirse los metadatos sobre la fecha de actualización y las condiciones de reutilización aplicables incluidos, en su caso, en el documento puesto a disposición para su reutilización.
6. En caso de información anonimizada por protección de datos personales u otros motivos, **está expresamente prohibido realizar labores de re-identificación de personas** a partir de estos datos y otras fuentes de datos e información posibles, pasadas, actuales o futuras.

**Exclusión de responsabilidad**

(...)

El Ayuntamiento de Madrid no será responsable del uso que de su información hagan las personas y/o empresas que reutilicen datos, ni tampoco de los daños sufridos o pérdidas económicas que, de forma directa o indirecta, produzcan o puedan producir perjuicios económicos, materiales o sobre datos, provocados por el uso de la información reutilizada.  
  
(...)

**Responsabilidad de las personas y/o empresas que reutilicen datos**

El agente reutilizador se halla sometido a la normativa aplicable en materia de reutilización de la información del sector público, incluyendo el régimen sancionador previsto en el artículo 11 de la Ley 37/2007, de 16 de noviembre, sobre Reutilización de la Información del Sector Público.

(...)"

Cabe señalar que si bien podemos reutilizar estos datos, estamos sujetos a la Ley 37/2007, de 16 de noviembre, sobre Reutilización de la Información del Sector Público. En concreto, es necesario señalar que la citada Ley establece que *"La utilización de los conjuntos de datos se realizará por parte de los usuarios o agentes de la reutilización bajo su responsabilidad y riesgo, correspondiéndoles en exclusiva a ellos responder frente a terceros por daños que pudieran derivarse de ella"*

## Transformación de los datos

Los importes recogidos en esta tabla son superiores a los 50.000 euros. Esto quiere decir que el importe recogido es con IVA.

Además, podemos distinguir que existen NIF y CIFS. Sería interesante diferenciar ambas situaciones. Pensemos, por ejemplo, en que los contratos con importes extremos se hayan adjudicado a una persona física o que ésta tenga una recurrencia de contratos mayor a la habitual.

La primera transformación, esto es, diferenciar los importes sin IVA, sí la vamos a realizar con OpenRefine porque sí incluye conocimiento que debería representarse en el dominio de los contratos menores (recuérdese que los límites de aprobación están en 50 y 18 mil euros). La segunda, si es empresa o no, sin embargo no es conocimiento del dominio sino un aspecto relacionado con el procesamiento posterior, no con el dominio.

Vamos a realizar esta transformación con OpenRefine:

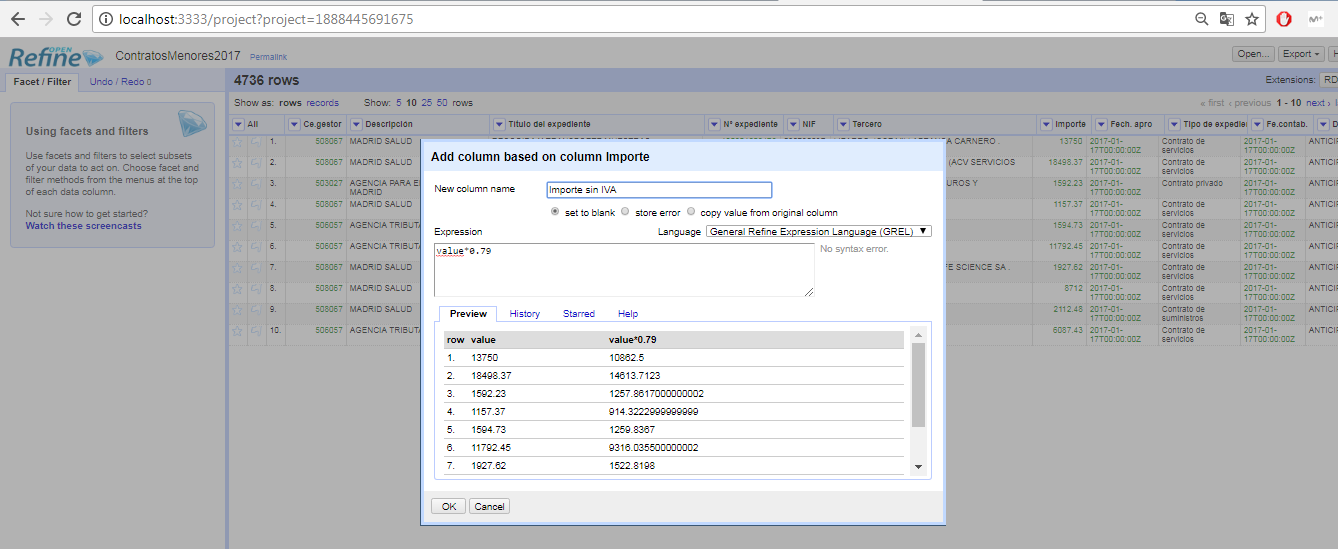


Imagen Transformación con OpenRefine

# Ontologías

Podría considerarse el diseño e implementación de una ontología específica para publicar estos datos. No obstante, una de las características fundamentales de las ontologías es su orientación hacia la reutilización, dado que representan conocimiento consensuado y compartido.

Se ha accedido al buscador LOV para acceder y buscar ontologías del dominio de la contratación pública, así como al buscador Google. Las ontologías encontradas han sido las siguientes:

http://lov.okfn.org/dataset/lov/vocabs/pc

http://lov.okfn.org/dataset/lov/vocabs/pproc

Cabe señalar que PC y PPROC cubren la totalidad de los conceptos definidos en mi dataset. Bajo esta premisa, no es necesario dado que el objeto de la reutilización de la ontología el desarrollo de una nueva podar la ontología existente.

Asimismo, cabe señalar que PPROC es una ontología específicamente diseñada para el análisis de los contratos adjudicados por Administraciones públicas. Según recoge la propia ontología, PPROC es una "Ontología creada con objeto de publicar datos sobre contratación publica española en RDF".

Por este motivo, utilizaremos la ontología PPROC para el etiquetado de nuestro dataset.

1. https://okfn.org/about/ [↑](#footnote-ref-2)
2. https://datos.madrid.es/ [↑](#footnote-ref-3)
3. http://datos.madrid.es/portal/site/egob/menuitem.c05c1f754a33a9fbe4b2e4b284f1a5a0/?vgnextoid=139afaf464830510VgnVCM1000000b205a0aRCRD&vgnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD&vgnextfmt=default [↑](#footnote-ref-4)
4. *Ley de Contratos del Sector Público 9/2017: https://www.boe.es/boe/dias/2017/11/09/pdfs/BOE-A-2017-12902.pdf* [↑](#footnote-ref-5)
5. https://www.boe.es/boe/dias/2011/11/16/pdfs/BOE-A-2011-17887.pdf [↑](#footnote-ref-6)
6. http://datos.madrid.es/portal/site/egob/menuitem.400a817358ce98c34e937436a8a409a0/?vgnextoid=eba412b9ace9f310VgnVCM100000171f5a0aRCRD&vgnextchannel=eba412b9ace9f310VgnVCM100000171f5a0aRCRD&vgnextfmt=default [↑](#footnote-ref-7)
7. http://datos.madrid.es/portal/site/egob/menuitem.3efdb29b813ad8241e830cc2a8a409a0/?vgnextoid=108804d4aab90410VgnVCM100000171f5a0aRCRD&vgnextchannel=b4c412b9ace9f310VgnVCM100000171f5a0aRCRD&vgnextfmt=default [↑](#footnote-ref-8)