Web Semántica & Linked data

Grupo 11

- Aizensztain Goltz, Uriel
- Camporredondo Portela, Alberto
 - Fernández Rico, Víctor
 - García Espeso, Eduardo
 - Salas Sancho, Olga

Datasets elegidos

Data sets de datos.madrid.org:

- Agenda eventos culturales y de ocio en Madrid
- Centros culturales de Madrid
- Juntas de los distritos de Madrid

Licencia:

http://datos.madrid.es/egob/catalogo/aviso-legal

Análisis de datasets

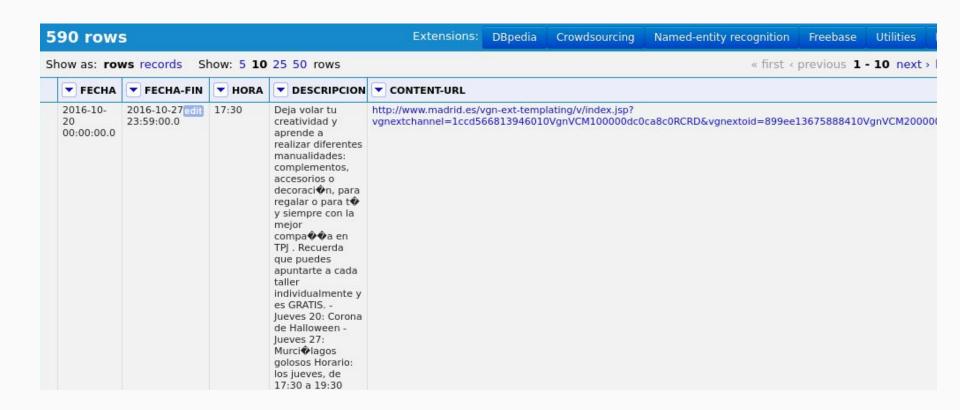
Datos dificiles de leer

Facules de explotar

••

Open → .

Análisis csv's con lodrefine



Análisis datasets: Centros culturales

Nombre	Tipo de val	or Descripción	Rango de valores	5
PK	Entero	campo clave, dentro de la web municipal, para uso interno	[0-9]	
Nombre	Cadena	Denominación de la instalación	[a-Z][0-9]	
Descripción-entida	ad Cadena	Horario de apertura		
Equipamiento	Cadena	Detalle de espacios internos especializados	[a-z][A-Z]	
Transporte	Cadena	estaciones de metro e indicación de línea; líneas de bus; estaciones de Renfe.	[a-z][A-Z][0-9]	
Descripción	Cadena	Información adicional; por ej.: otros equipamientos con los que se comparte la sede física; información de venta de entradas	[a-z][A-Z][0-9]	
Accesibilidad	Booleano	Si es accesible o no para personas con minusvalías	[a-z][A-Z][0-9]	
Content-URL	Cadena	Nombre del recurso en internet	[a-z][A-Z][0-9]	
Nombre vía	Cadena		[a-z][A-Z]	
Clase vía	Cadena	Tipo de vía: calle, avenida, etc	[a-z][A-Z]	
Tipo num	Cadena	V, S/N, KM		[A-Z
Núm	Entero	número del equipamiento en el vial	[0-9]	
Planta	Entero	Nivel que ocupa el inmueble	[0-9]	
Puerta	Entero	Puerta que ocupa en el inmueble	[0-9]	
Escalera	Entero	Escalera de acceso	[0-9]	
Orientación	Cadena	Detalles descriptivos de la dirección	[a-z][A-Z]	
Localidad	Cadena	Municipio del lugar	[a-z][A-Z]	
Provincia	Cadena	Provincia a la que pertenece la localidad	[a-z][A-Z]	
Código postal	Entero	Código postal del centro cultural	[0-9]	
Barrio	Cadena	Barrio donde esta el centro cultural	[a-z][A-Z]	
Distrito	Cadena	Distrito donde esta el centro cultural	[a-z][A-Z]	
Coordenada X	Entero	coordenada de 6 caracteres sobre plano de la ciudad	[0-9]	
Cordenada Y	Entero	coordenada de 7 caracteres sobre plano de la ciudad	[0-9]	
Latitud	Decimal	coordenada latitud en el sistema de referencia WGS84	[0-9].[0-9]	
Longitud	Decimal	coordenada longitud en el sistema de referencia WGS84	[0-9].[0-9]	
Teléfono	Entero	Teléfono del centro cultural	[0-9]	
Fax	Entero	Fax del centro cultural	[0-9]	
Email	Cadena	Correo electrónico del centro cultural	[a-z][A-Z][0-9]	
Tipo	Cadena	Descripción de la tipología de la instalación	[a-z][A-Z][0-9]	

Análisis dataset: Agenda eventos culturales

Dataset: a	genda ev	entos culturales	
Nombre	Tipo de valor	Descripción	Rango de valores
id-evento Título Precio Gratuito	String Integer	Es un campo clave, dentro de la web municipal, para uso interno Denominación del evento al que se refiere Coste del evento/actividad. Si son gratuitas, el campo aparece vacío Figurará un 1 o 0 si es gratuito o no	[0-9]{1,8} Cualquier carácter alfanumérico Cualquier número
Larga duración	Booleano	Figurará un 1 o 0 si es un evento que se realiza a lo largo del tiempo	[0-1]
Días semana		Iniciales de cada día de la semana en los que se desarrolla el evento. L,M,X,J,V. Esta codificación no es válida para eventos de larga duración.	L,M,X,J,V
Días excluidos	Date	En eventos de larga duración, son los días concretos que no se celebra la misma. Es una fecha en formato D/M/AAAA. Si existiesen varios días que no se celebran, se mostrarían las fechas separadas por el carácter ";"	Cualquier fecha
Hora Descripción	Date Date Date	Campo fecha del inicio del evento, con estructura "aaaa-mm-dd hh:mm:ss" Campo fecha del fin del evento, con estructura "aaaa-mm-dd hh:mm:ss" Campo en el que se indica la hora de inicio del evento, con la estructura "hh:mm" Campo con la descripción del evento	Cualquier fecha Cualquier fecha [0-23]:[0-59] Cualquier cadena de carácteres
Content- URL	URL	Nombre del recurso en internet del evento en concreto, en la <u>web municipal</u>	"http://www.madrid.es/vgn-ext-templating/v/index.jsp?vgnextchannel=*" donde asterisco es cualquier identificador o cadena de carácteres
Título Actividad	String	denominación del evento o actividad a la que esté asociado	Cualquier cadena de carácteres
URL- Actividad URL-		Nombre del recurso en internet, en la <u>web municipal</u>	"http://www.madrid.es/vgn-ext-templating/v/index.jsp?vgnextchannel=*" donde asterisco es cualquier identificador o cadena de carácteres "http://www.madrid.es/vgn-ext-templating/v/index.jsp?vgnextchannel=*"
Instalación	URL	Nombre del recurso centro en internet, en la <u>web municipal</u>	donde asterisco es cualquier identificador o cadena de carácteres
Nombre instalación	3	Denominación de la instalación donde se desarrolla el evento	Cualquier cadena de carácteres
Coordenada X		Coordenada de 6 caracteres sobre el plano de la ciudad	[0-9]{1,6}
Coordenada Y	Integer	Coordenada de 7 caracteres sobre el plano de la ciudad	[0-9]{1,6}
Latitud	float		[-+]?[0-9].[0-9]
Longitud	float	Coordenada longitud en el sistema de referencia WGS84. NOTA: si se abre en fichero excel o csv se muestran mal por temas de conversión.	[-+]?[0-9]*\.?[0-9]*
Distrito		Nombre del distrito donde se desarrolla la actividad	Cualquier cadena de carácteres
Tipo		Descripción de la tipología de la actividad. Se encuentra al final de la cadena de texto. En muchos conjunto de datos, el valor que contenga será igual, pero por ejemplo en otros, como Actividades, si ue permite una clasificación.	Cualquier cadena de carácteres

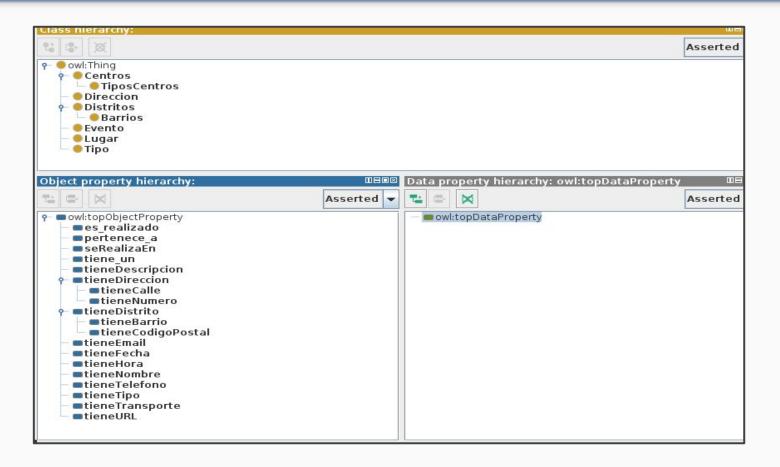
Análisis dataset: Juntas distrito

Dataset: jun	tas dis	trito	
	Integer Date String String	Descripción de quien organiza la actividad Es un campo clave, dentro de la web municipal, para uso interno Campo donde se describe el horario de la actividad Descripción de como llegar a la ubicación mediante el transporte público Pequeña descripción de la actividad Nivel de acceso para personas descapacitadas	Cualquier cadena de carácteres [0-9]
Content-URL	URL	Nombre del recurso en internet del evento en concreto	"http://www.madrid.es/vgn-ext-templating/v/index.jsp?vgnextchannel=*" donde asterisco es cualquier identificador de carácteres
Provincia Disrito	String	Municipio de la localidad Provincia a la que pertenece la localidad Distrito donde se encuentra la junta del distrito	[a-z][A-Z] [a-z][A-Z] [a-z][A-Z]
Coordenada X	Integer	Coordenada de 6 caracteres sobre el plano de la ciudad	[0-9]{1,6}
Coordenada Y	Integer	Coordenada de 7 caracteres sobre el plano de la ciudad	[0-9]{1,6}
Latitud Longitud	float float	Coordenada latitud en el sistema de referencia WGS84 Coordenada longitud en el sistema de referencia WGS84. NOTA: si se abre \$	[-+]?[0-9].[0-9] [-+]?[0-9]*\.?[0-9]*
Teléfono Fax Email Tipo	Entero Cadena	Teléfono del centro cultural Fax del centro cultural Correo electrónico del centro cultural I Descripción de la tipología de la instalación	[0-9] [0-9] [a-z][A-Z][0-9] [a-z][A-Z][0-9]

Naming strategy

- Se creó una URI para referenciar los términos de la ontología creada por nosotros.
 - Los objetos empiezan por mayúsculas:
 - https://www.madridcultural.es/ontology.owl#Centro
 - Las propiedades empiezan por minúsculas
 - https://www.madridcultural.es/ontology.owl#longitud
- Para los individuos (recursos) se utilizó las URIs que ya venían en los csv, o se buscaron en diferentes grafos de conocimientos.
 - https://www.dbpedia.org/resource/Canillas_(Madrid)

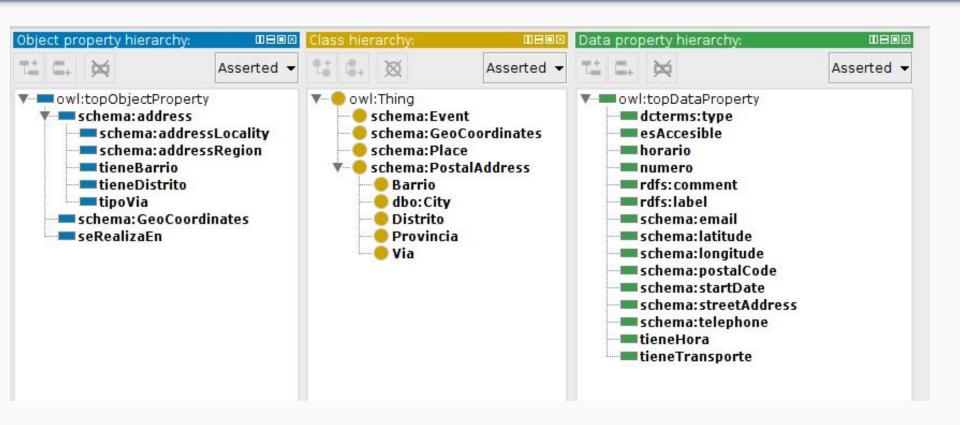
Ontología



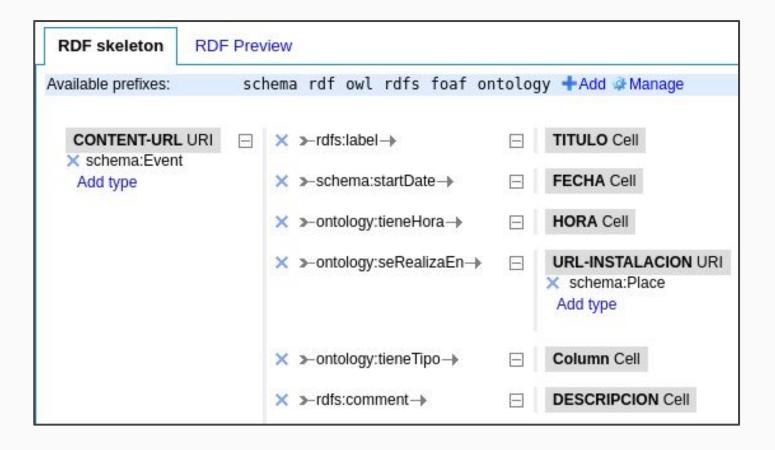
Generación RDF's

```
<http://www.madridcultural.es/139451> <http://www.madridcultural.es/tieneNombre> "Auditorio y sala de exposiciones Paco de Luc
        <http://www.madridcultural.es/tieneDistrito> "LATINA" :
        <http://www.madridcultural.es/tieneBarrio> "LAS AGUILAS" :
        <http://www.madridcultural.es/tieneCodigoPostal> "28044" ;
        <http://www.madridcultural.es/tieneTransporte> " Metro: Aluche (linea 5) Bus: 17, 31, 117, 121, 131, 138, 139, 155
        <http://www.madridcultural.es/tieneEmail> "ruizdm@madrid.es" ;
        <http://www.madridcultural.es/tieneTelefono> "915 120 890 / 915 120 891" .
<http://www.madridcultural.es/1956> <http://www.madridcultural.es/tieneNombre> "Casa de América" ;
        <http://www.madridcultural.es/tieneDescripcion> " Antiquo Palacio de Linares, diseñado por Carlos Colubí en 1872. De
        <http://www.madridcultural.es/tieneDistrito> "SALAMANCA" ;
        <http://www.madridcultural.es/tieneBarrio> "RECOLETOS";
        <http://www.madridcultural.es/tieneCodigoPostal> "28014" ;
        <http://www.madridcultural.es/tieneTransporte> " Metro: Banco de España (línea 2) Bus: 1, 2, 5, 9, 10, 14, 15, 20, 27
        <http://www.madridcultural.es/tieneTelefono> "915 954 800 (información) " .
<http://www.madridcultural.es/7320916> <http://www.madridcultural.es/tieneNombre> "Casa de la Juventud de Arganzuela" ;
        <http://www.madridcultural.es/tieneDistrito> "ARGANZUELA" ;
        <http://www.madridcultural.es/tieneBarrio> "CHOPERA" ;
        <http://www.madridcultural.es/tieneCodigoPostal> "28045";
        <http://www.madridcultural.es/tieneTransporte> " Metro: Legazpi (líneas 3, 6), Pirámides (línea 5) Bus: 6, 8, 18, 19
        <http://www.madridcultural.es/tieneTelefono> """915 132 661
0.0.0
<http://www.madridcultural.es/4605237> <http://www.madridcultural.es/tieneNombre> "Centro Casa Asia - Madrid" ;
        <http://www.madridcultural.es/tieneDistrito> "CENTRO" ;
        <http://www.madridcultural.es/tieneBarrio> "PALACIO" ;
        <http://www.madridcultural.es/tieneCodigoPostal> "28013";
        <http://www.madridcultural.es/tieneTransporte> " Bus: 3, 148, L3 " ;
        <http://www.madridcultural.es/tieneEmail> "casaasia@casaasia.es" :
        chttn://www.madridcultural.es/tieneTelefonos """913 600 194
```

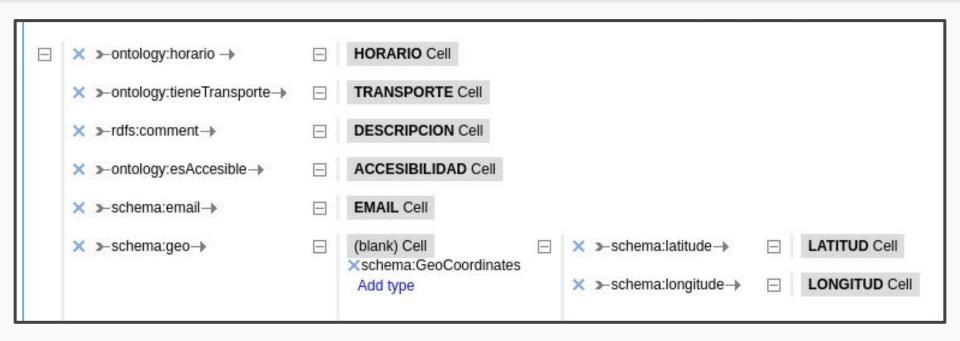
Actualizar ontología



Actualizar RDF: Agenda eventos culturales



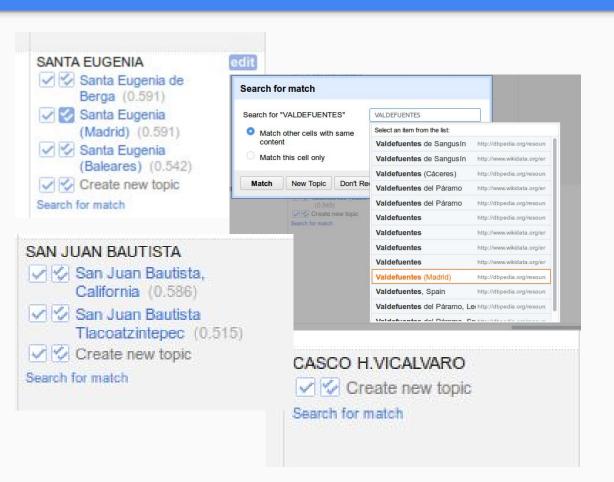
Actualizar RDF: Centros culturales



Actualizar RDF: Juntas distritos



Reconciliación con otros Datasets







SPARQL Construct + LODRefine

DBpedia Endpoint: http://dbpedia.org/sparql

```
CONSTRUCT { ?barrio rdfs:label ?nbarrio }
     WHERE {
             ?barrio rdfs:label ?nbarrio.
             ?barrio dct:subject dbc:Wards of Madrid.
 6 }
<a href="http://dbpedia.org/resource/Pac\u00EDfico_(Madrid)">http://dbpedia.org/resource/Pac\u00EDfico_(Madrid)>
   rdfs:label "Pac\u00EDfico (Madrid)"@es , "Pac\u00EDfico (Madrid)"@en , "Pac\u00EDfico (bairro)"@pt .-
<a href="http://dbpedia.org/resource/Palacio_(Madrid)>-</a>
   rdfs:label "Palacio (Madrid)"@es , "Palacio (Madrid)"@en , "Palacio"@pt , "Palacio"@nl .-
<http://dbpedia.org/resource/Palomas_(Madrid)>
   rdfs:label "Palomas (bairro de Madrid)"@pt , "Palomas (Madrid)"@en , "Palomas (Madrid)"@es .-
dbr:Palomeras Bajas
   rdfs:label "Palomeras Bajas"@es , "Palomeras Bajas"@en , "Palomeras Bajas"@pt .-
dbr:Palomeras Sureste
   rdfs:label "Palomeras Sureste"@es , "Palomeras Sureste"@pt , "Palomeras Sureste"@en .-
<http://dbpedia.org/resource/Palos de Moguer (Madrid)>
   rdfs:label "Palos de la Frontera (bairro de Madrid)"@pt , "Palos de Moguer (Madrid)"@es , "Palos de Moguer (Madr
dbr:Pavones
   rdfs:label "Pavones"@pt , "Pavones"@en , "Pavones"@es .-
<http://dbpedia.org/resource/Piovera_(Madrid)>
   rdfs:label "Piovera (bairro de Madrid)"@pt , "Piovera (Madrid)"@en , "Piovera (Madrid)"@es .-
shttp://dhpedia.org/resource/Portazgo (Madrid)>
```

his will set up a new reco	nciliation service l	based on an RDF file tha	t provides entity URIs and			
ame: Barrios de Madrid						
A human readable name						
ile details						
Load file from URL:						
Upload file:	Examinar	barriosmadrid2.ttl				
File format:	Turtle 🗸					
properties will be used to performance.			m as this migh affect the			

Metadatos

Vocabulario: DCAT

- 3 catálogos
- Distribución TTL
- Licencia: cc-by-sa3.0es
- Fuente: http://datos.madrid.es/

Datasets:

- dcat:Dataset
- Título
- o Descripción

Distribuciones: TTL

- dcat:Distribution
- downloadURL
- Título
- Media-type: application/x-turtle

```
:catalog
    a dcat:Catalog ;-
    dct:title "Dataset of cultural events in Madrid"@en , "Dataset de los eventos culturales
    rdfs:label "Eventos Culturales"@es , "Cultural events"@es ;
    foaf:homepage <a href="https://github.com/albertcp/Curso2016-2017">https://github.com/albertcp/Curso2016-2017</a>;
    dct:creator <https://github.com/albertcp/Curso2016-2017> ;-
    dct:creator <https://github.com/urieljavier/Curso2016-2017> ;-
    dct:creator <https://github.com/eduman13/Curso2016-2017> ;-
    dct:creator <https://github.com/olgasalas/Curso2016-2017> ;-
    dct:creator <https://github.com/vfrico/Curso2016-2017>;
    dct:source <http://datos.madrid.es/> ;-
    dct:publisher vfrico ;-
    dct:publisher urieljavier ;
    dct:publisher eduman13 ;-
    dct:publisher olgasalas ;-
    dct:publisher albertcp ;-
    dct:license rdflicense:cc-by-sa3.0es ;-
    dct:language <a href="http://lexvo.org/id/iso639-3/spa">http://lexvo.org/id/iso639-3/spa</a>
    dcat:dataset :JuntasDistrito, :CentrosCulturales, :AgendaEventosCulturales ;
```

```
: JuntasDistrito
    a dcat:Dataset :-
    dct:title "Juntas de distritos de la ciudad de Madrid"@es ;
    dcat:keyword "Juntas Distritos"@es ;-
    dct:issued "2016-12-20"^^xsd:date ;-
 -- dct:modified "2016-12-20"^^xsd:date ;-
 -- dcat:description "Juntas de distritos de la ciudad de Madrid"@es ;-
 - dcat:distribution :JuntasDistrito-ttl :-
    dct:license rdflicense:cc-by-sa3.0es ;
:JuntasDistrito-ttl-
   a dcat:Distribution ;-
   dcat:downloadURL <a href="https://github.com/albertcp/Curso2016-2017/blob/master/HandsOn/Group11/rdf/juntas-distrito.ttl">https://github.com/albertcp/Curso2016-2017/blob/master/HandsOn/Group11/rdf/juntas-distrito.ttl</a>;
    dct:title "Juntas de distritos de la ciudad de Madrid"@es ;-
--- dcat:mediaType "application/x-turtle" ;--
dct:license rdflicense:cc-by-sa3.0es;
:CentrosCulturales
    a dcat:Dataset ;-
   dct:title "Centros culturales de la ciudad de Madrid"@es ;-
    rdfs:label "Centros Culturales"@es ;-
    dcat:keyword "Centros, Culturales"@es;
    dct:issued "2016-12-20"^^xsd:date ;-
 -- dct:modified "2016-12-20"^^xsd:date ;-
 -- dcat:description "Centros culturales de la ciudad de Madrid"@es ;-
   dcat:distribution :CentrosCulturales-ttl;
```

Fin

Grupo 11

- Aizensztain Goltz, Uriel
- Camporredondo Portela, Alberto
 - Fernández Rico, Víctor
 - García Espeso, Eduardo
 - Salas Sancho, Olga