

# Datos enlazados

Los pasos que se detallan a continuación permiten **transformar datos tabulares en formato CSV a RDF usando vocabularios comunes**. Con ello se logra que aquellos conjuntos de datos abiertos sobre el mismo dominio que sean publicados por organizaciones distintas, **utilicen esquemas similares para representar los datos en la Web**, facilitando de esta forma su reutilización y explotación a mayor escala por parte de los reutilizadores.

La metodología que se describe utiliza **vocabularios controlados**, se apoya en **herramientas como OpenRefine para la limpieza, preparación y enlazado de datos del CSV de partida, RML para la implementación de reglas de transformación y Morph-KGC para la generación del RDF**.

**Paso 1: Seleccionar un vocabulario** (esquema de datos) estándar que represente de forma consensuada el dominio de los datos de entrada.

**Tareas:**

- Identificar un vocabulario **compatible con el dominio** de los datos de entrada.
- Analizar y comprender el vocabulario, las **relaciones de las clases y sus propiedades**.
- Examinar en profundidad los posibles **ejemplos de las tripletas RDF** que se deben generar.
- Asegurar que el vocabulario representa la información del dominio de los datos de entrada.

**Paso 2: Limpiar y preparar los datos** de entrada en CSV, ajustando a los requisitos del formato que define el vocabulario, utilizando **OpenRefine**.

**Tareas:**

- **Transformar los valores de las columnas** de acuerdo a buenas prácticas de publicación de archivos CSV utilizando las **funciones "GREL"** disponibles en la herramienta.
- Preparación de los datos de entrada para la generación de **listas controladas (SKOS)**.
- **Enlazar las columnas del CSV con entidades** definidas en bases de conocimiento como Wikidata para **enriquecer los datos**.
- Descargar el CSV modificado.

**Paso 3: Construir un documento de reglas de transformación** que siga unas normas ya preestablecidas dónde se relacionen las clases y propiedades definidas en el vocabulario escogido con cada una de las columnas del fichero CSV de entrada.

**Tareas:**

- Conocer la **sintaxis de YAML**.
- Entender y comprender la **semántica de cada campo del CSV así como de las clases y propiedades del vocabulario**.
- Describir las **reglas de transformación utilizando YARRRML**. Extracto de documento (no completo) con ajustes que se explican en la guía →

**Paso 4: Generar el RDF a partir de los datos de entrada aplicando las reglas de transformación creadas.**

**Tareas:**

- **Validación de las reglas de transformación** creadas.
- **Generación de RDF** con Morph-KGC.

Una vez generado el RDF, se puede publicar como **distribución del conjunto de datos** en el portal de datos abiertos. Igualmente, es recomendable hacer **disponible el conjunto de reglas en YARRRML o RML** usadas para generar el RDF.

```

prefixes:
  rr: http://www.w3.org/ns/r2rml#
  foaf: http://xmlns.com/foaf/0.1/
  xsd: http://www.w3.org/2001/XMLSchema#
  rdfs: http://www.w3.org/2000/01/rdf-schema#
  rev: http://purl.org/stuff/rev#
  schema: http://schema.org/
  escom: http://vocab.ciudadesabiertas.es/def/comercio/tejido-comercial
  escom-skos: http://vocab.linkeddata.es/datosabiertos/kos/comercio
...
mappings:
  terraza:
    sources:
      - [/data/OPEN_DATA_Terrazas202107.csv~csv]
    s: http://datos.madrid.es/recurso/terrazas/${id_terrazas}
    po:
      - [a, escom:Terraza]
      - [escom:superficie, $(Superficie_ES), xsd:float]
      - [escom:numeroMesasAutorizadas, $(mesas_es), xsd:integer]
      - [escom:numeroSillasAutorizadas, $(sillas_es), xsd:integer]
      - [schema:openingHours, Estacional Lun-Juev $(hora_ini_LJ_es) - $(hora_fin_LJ_es)]
      - [schema:openingHours, Estacional Vier-Dom $(hora_ini_VS_es) - $(hora_fin_VS_es)]
      - [schema:openingHours, Anual Lun-Juev $(hora_ini_LJ_ra) - $(hora_fin_LJ_ra)]
      - [schema:openingHours, Anual Vier-Dom $(hora_ini_VS_ra) - $(hora_fin_VS_ra)]
      - [escom:periodoFuncionamiento, escom-skos:periodo-funcionamiento/${id_periodo_terrazas}~iri]
      - p: escom:perteneceA
        o:
          - mapping: localcomercial
      - p: schema:address
        o:
          - mapping: direccionpostal
    localcomercial:
      sources:
        - [/data/OPEN_DATA_Terrazas202107.csv~csv]
      s: http://datos.madrid.es/recurso/localcomercial/${id_local}
      po:
        - [a, escom:LocalComercial]
        - p: geosparql:hasGeometry
          o:
            - mapping: point
    point:
      sources:
        - [/data/OPEN_DATA_Terrazas202107.csv~csv]
      s: http://datos.madrid.es/recurso/localcomercial/${id_local}/geometry
      po:
        - [a, sf:Point]
        - [geo_core:xETRS89, $(coordenada_x_local), xsd:float]
        - [geo_core:yETRS89, $(coordenada_y_local), xsd:float]
    direccionpostal:
      sources:
        - [/data/OPEN_DATA_Terrazas202107.csv~csv]
      s: http://datos.madrid.es/recurso/direccionpostal/${id_vial_edificio}-${num_edificio}-${Cod_Postal}
      po:
        - [a, esdir:DireccionPostal]
  
```

# Extracto de documento de reglas de transformación en YARRRML