

Generación de datos estadísticos enlazados del Instituto Aragonés de Estadística

Oscar Corcho, Universidad Politécnica de Madrid (ocorcho@fi.upm.es)

Hugo Lafuente, Localidata (hlafuente@localidata.com)

David Portolés, Idearium Consultores (dportoles@idearium-consultores.com)

José María Subero, Gobierno de Aragón (jmsubero@aragon.es)

César Cano, IAEst - Gobierno de Aragón (ccano@aragon.es)

M^a José Laplana, IAEst - Gobierno de Aragón (mjlaplana@aragon.es)

Resumen

En esta ponencia se presenta el trabajo realizado durante los años 2015 y 2016 para el tratamiento, transformación y publicación según el paradigma de los datos enlazados de una gran parte de los informes de estadística local publicados por el Instituto Aragonés de Estadística (IAEst - <http://www.aragon.es/iaest/EstadisticaLocal>).

Se parte de los informes en formato CSV publicados por el sistema de publicación de informes del IAESt. Estos informes son tratados, cada vez que se producen cambios, por un proceso automático que realiza su transformación al formato RDF de acuerdo con el vocabulario W3C RDF DataCube. Asimismo, el proceso identifica de manera global las dimensiones utilizadas por la globalidad de todos los informes y genera dichas dimensiones como listas de códigos de acuerdo con el vocabulario SKOS. Estos datos generados en RDF, así como las descripciones de los informes (cubos de datos de acuerdo con el estándar RDF Datacube), se cargan de manera automática en el triple store del Gobierno de Aragón, cuyo punto de acceso SPARQL está disponible en <http://opendata.aragon.es/sparql>, y se regeneran y vuelven a cargar de manera automática cada vez que se detectan cambios en cualquiera de los informes.

Todos los datos se encuentran además publicados como datos enlazados, con una API REST que da acceso a todos los informes y sus descripciones, todas las listas de códigos, y todas las observaciones, cada uno de los cuales tiene su propio identificador universal (URI), lo que permite hacer referencia a cualquiera de ellos. Todo el proceso se gestiona utilizando GitHub como plataforma de compartición de datos y de gestión de versiones (<https://github.com/aragonopendata/local-data-aragopedia>).

1. Introducción: situación de partida en la Estadística Local

Desde el año 2007 se venía trabajando en el Instituto Aragonés de Estadística en dar solución al problema de la dispersión de la información. Se disponía de una gran cantidad de bases de datos con información territorial de carácter local, todas ellas independientes y de diversos ámbitos temáticos. La gestión de esa información se basaba en la administración independiente de distintas bases de datos relacionales. Todo ello implicaba una gestión no centralizada de clasificaciones y nomenclaturas (territorio, nacionalidades, CNAE, etc.) que al final derivaba en una gran heterogeneidad en las publicaciones.

Al mismo tiempo se había iniciado un proyecto destinado a la difusión de información estadística a nivel local, tratando el territorio como objeto de consulta más que como dimensión de una determinada operación. Es decir, se estaba gestando una nueva operación a la que llamamos **Estadística Local**.

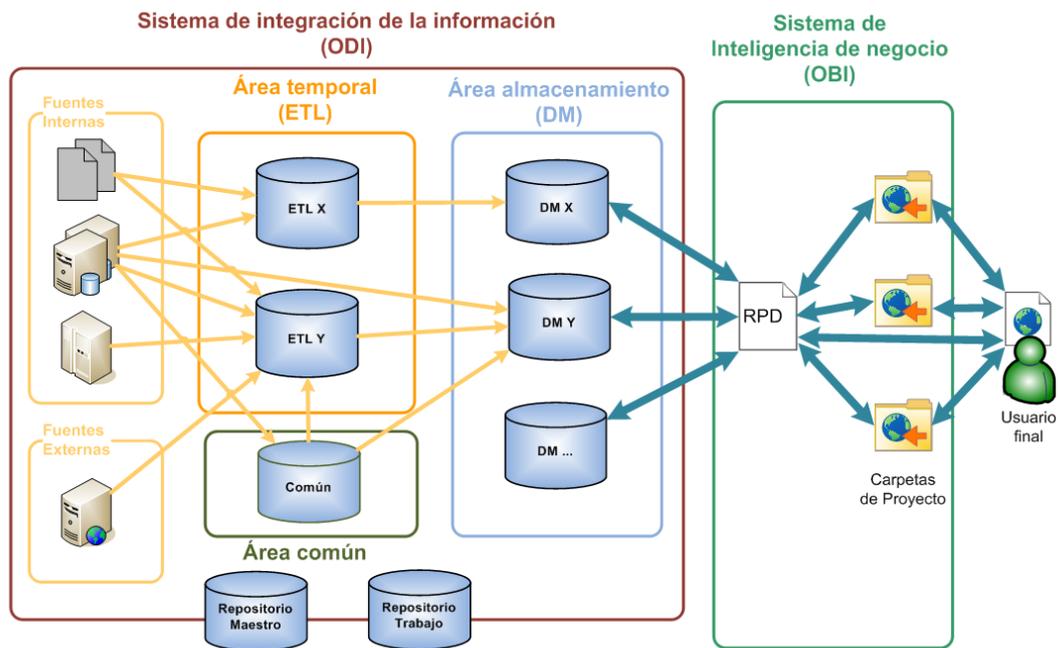
Estadística Local nació como un proyecto web en el que el usuario podía obtener toda la información disponible de un territorio concreto. Para ello se crearon fichas territoriales en formato Excel y PDF para cada municipio y cada comarca de Aragón, siendo actualizadas periódicamente (3 o 4 veces al año). Las fichas se hicieron extensivas a las tres provincias y a la comunidad autónoma para garantizar la comparabilidad. Este producto tuvo una gran aceptación por parte de los usuarios del IAEST, y al mismo tiempo se demandaba completarlo con la disposición de los datos de la ficha en forma de base de datos, por lo que se colgaron las matrices de datos en formato Excel. El punto flaco de Estadística Local eran las arduas tareas de mantenimiento.

1.1 Implantación de un Data Warehouse

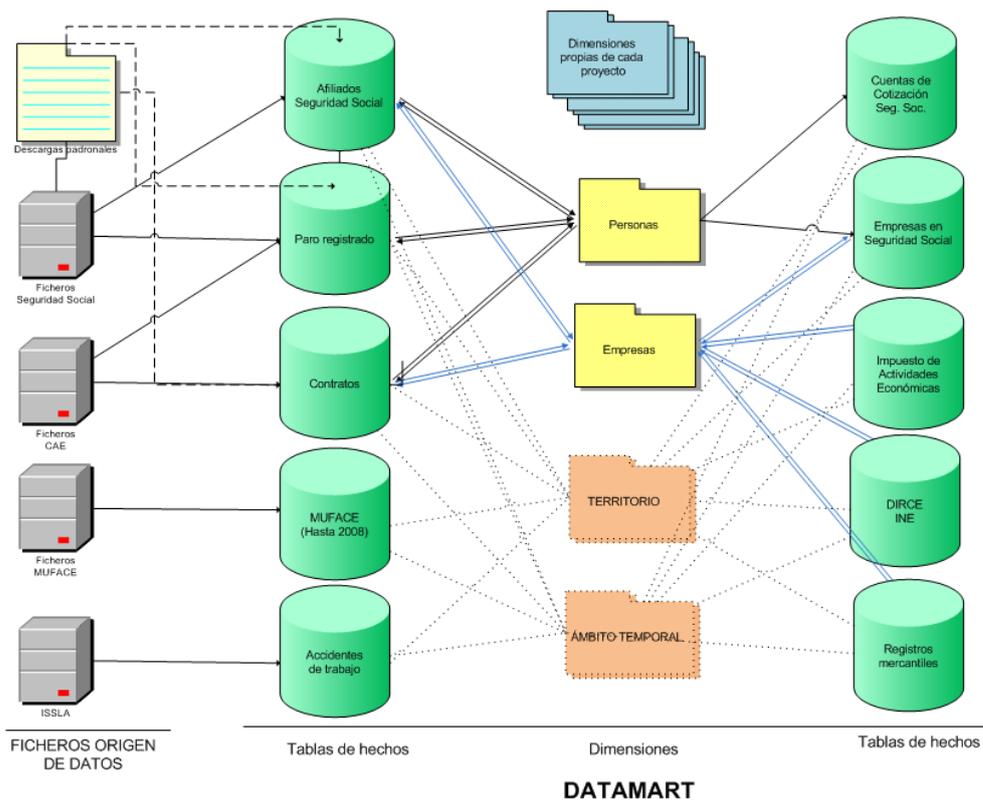
Vista la necesidad de dinamizar Estadística Local y de la gestión centralizada de los microdatos en el IAEST, se implantó una herramienta basada en la tecnología *Data Warehouse*, cuyo desarrollo exige una buena definición y diseño en la base de datos (dimensiones, jerarquías, variables, etc.), pero que a cambio produce muy buenos resultados en el análisis utilizando software específico orientado a la agregación de grandes cantidades de datos.

Un *Data Warehouse* (DW) es la colección de datos, organizados, integrados, historizados y disponibles para facilitar la generación de informes estadísticos. Desde el punto de vista de su construcción, un DW es una colección de bases de datos integradas no volátiles y orientadas a ciertos temas, diseñada para apoyar la toma de decisiones, donde cada unidad de datos es relevante en un momento de tiempo. Los datos se ofrecen al usuario desde la capa de presentación, que se adapta a las necesidades de información y que es independiente de cómo estén almacenados físicamente los datos.

Los datos del Data Warehouse están almacenados en una base de datos, en nuestro caso Oracle, y se cargan los datos desde las distintas fuentes mediante procesos ETL (Extracción, Transformación y Carga) con el software de integración Oracle Data Integrator (ODI), realizando cargas iniciales o incrementales de cada DataMart. Una vez están disponibles los datos en los DataMart (entendido como subconjuntos de datos agrupados por temas, utilizando tablas de hechos, dimensiones, jerarquías, etc.), se diseña la capa de presentación de los datos que ofrecerá el software de análisis, Oracle Business Intelligence (OBI) en nuestro caso, con el que se explota la información por parte de los técnicos estadísticos del IAEST.



Ejemplo de integración de diversas fuentes en el IAEST:



Los técnicos del IAEST utilizan OBI para diseñar las consultas contra el DataMart; OBI permite hacer una presentación formateada, en forma de tablas, gráficos, textos aclaratorios, etc., a los que, con los permisos adecuados, los usuarios finales pueden acceder desde la web del IAEST conocidos como “Informes” o “Análisis”. Todo ello, supone que la información publicada en el IAEST es dinámica, es decir, en el momento que se cargan nuevos datos con ODI en el DataMart de un proyecto, los Informes publicados en la Web del IAEST mostrarán automáticamente toda la información actualizada de dicho proyecto. El OBI también permite un acceso fácil y rápido a los datos, para resolver consultas a medida o necesidades de información dentro del propio Instituto. Internamente, los análisis se pueden almacenar en un catálogo organizado como un sistema de carpetas dentro de la propia herramienta OBI.

La implantación del Data Warehouse ha solucionado algunos aspectos de la gestión de las bases de datos en el IAEST:

- La gestión centralizada y única que se hace de nomenclaturas y clasificaciones que garantiza homogeneidad en todos nuestros proyectos, y en consecuencia en la publicación de datos.
- La conexión de distintas fuentes mediante variables comunes para extraer la información.

1.2 Banco de datos de Estadística Local

Una vez puesto en marcha el acceso de usuarios finales a los análisis de OBI, se procedió a la reestructuración de Estadística Local. Se añadió un banco de datos ligado a Entidades Locales y

que obtuviera la información de forma dinámica de los diferentes proyectos de Data Warehouse.

Cuando se planteó el banco de datos municipal, se consideraron una serie de requisitos:

- Debería ser de **fácil acceso y navegación** para el usuario. El éxito (medido por el nivel de utilización por parte de usuarios) dependía en gran medida de este punto
- Se debería poder **obtener resultados en forma matricial**: municipios × variables o comarcas × variables.
- El usuario debería poder saber de una forma sencilla **qué información hay disponible y para qué periodos**.
- Se mostrará como mínimo el **último periodo disponible**, pero el usuario podrá acceder a la información histórica dentro de la misma página.
- La **definición y fuente** de cada variable disponible en la base de datos debería ser de fácil acceso. Es decir, no se puede hacer una metodología única, puesto que las consultas son sobre variables concretas.

Finalmente se optó por el desarrollo de una aplicación propia de gestión de publicación de los análisis obtenidos con OBI. La aplicación Estadística Local se basa en un árbol temático de información disponible, que se gestiona en una base de datos de ORACLE, el menú ofrecido depende del tipo de territorio que el usuario ha seleccionado y en el caso de que sea municipal de su población -consecuencia de la disponibilidad y del secreto estadístico-.

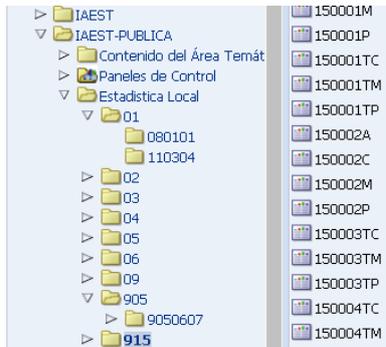
[Indicaciones de uso y contenido temático del banco de datos](#)

Paso 1: Seleccione un ámbito territorial

The screenshot shows a web application interface with a navigation menu on the left and a main content area. The navigation menu includes options for 'Municipal', 'Comarcal', 'Provincial', and 'Aragón'. The main content area displays 'Paso 2: Selección de territorio' with a search bar and a list of municipalities. Below this, 'Paso 3: Acceda a la información' is visible. The main content area shows a table of economic activities for San Miguel del Cinca (Municipio código: 22903) from 2009 to 2013. The table is titled 'Actividades económicas en el territorio según sector y agrupación de actividad (CNAE-2009). Municipio' and includes a source and methodology note.

Sector	Rama de actividad	Año	2009	2010	2011	2012	2013
Agricultura, ganadería y pesca	Agricultura, ganadería, silvicultura y pesca (cnae 01, 02, 03)		51	48	46	46	49
Industria y energía	Industrias extractivas (cnae 05, 06, 07, 08, 09)		1	1	1	1	1
	Industria de la alimentación, bebidas y tabaco (cnae 10, 11, 12)		1	1	1	1	2
	Industria de la madera y corcho, papel y artes gráficas (cnae 16, 17, 18)		4	4	4	5	4
	Metalurgia y fabricación de productos metálicos, excepto maquinaria y equipo (cnae 24, 25)		2	2	2	1	1
	Suministro de energía eléctrica, gas, vapor y aire acondicionado (cnae 35)		1	1	1	1	1
	Suministro de agua, actividades de saneamiento, gestión de...		1	1	2	2	2

Los estadísticos del IAEST crean los análisis, que irán parametrizados por territorio, que se organizan dentro del catálogo de OBI. Con una gestión de permisos adecuada los informes pueden ser visualizados y descargados desde el exterior. El usuario final solo puede ver el informe resultante sin posibilidad de modificarlo (salvo alteración de posición de filas, columnas, etc.) con ello queda garantizada la seguridad para la información de las bases de datos.



2. Objetivos del proyecto de generación y publicación de datos enlazados

Una vez descrita la situación de partida, se exponen los objetivos del proyecto que se ha abordado en el último año para la transformación de los datos de estadística local expuestos a partir del Data Warehouse al lenguaje de representación RDF y su publicación como datos enlazados.

Teniendo en cuenta la forma en la que se exponían hasta la fecha los datos de Estadística Local anteriormente mencionados, surgió la necesidad y oportunidad de exponer estos datos de una manera más estructurada y normalizada, para así facilitar su reutilización por terceros más allá del uso de los informes tal y como se presentaban en formatos como Excel o PDF. Aprovechando el hecho de que el Gobierno de Aragón ya había trabajado en el pasado con tecnologías de datos enlazados para la exposición de datos sobre municipios (en lo que se conoce como AragoDBpedia - <http://opendata.aragon.es/portal/desarrolladores/api-aragodbpedia>), se propuso la posibilidad de aumentar la cantidad de datos expuestos en estos formatos con los datos procedentes del IAEst.

En las siguientes secciones se describe el proceso llevado a cabo y los resultados obtenidos y que se encuentran en la actualidad ya en producción.

3. Generación de datos enlazados sobre estadística local

Como ya se ha comentado anteriormente, los datos de estadística local expuestos por el IAEst están disponibles para su visualización en HTML, así como para su descarga en formatos como Excel o CSV, entre otros, que son los que los potenciales reutilizadores de los datos utilizarían para hacer uso de los mismos.

3.1 Ejemplo de informe: viviendas según antigüedad en el municipio de Jaraba

A continuación se presenta un ejemplo de uno de estos informes que se hacen disponibles a partir de la aplicación de Estadística Local. Concretamente, seleccionamos el informe de “Edificios según antigüedad” correspondientes a la localidad de Jaraba, cuyo código de informe es el 030092M.

Mostrar información de un municipio:

Jaraba 

Mostrar información de todos los municipios de:

- Aragón
- Provincia de Huesca
- Provincia de Teruel
- Provincia de Zaragoza

Paso 3: Acceda a la ficha territorial o al banco de datos

Ficha municipal

-  [Descargar en formato PDF](#)
-  [Descargar en formato XLS](#)

Banco de datos

- + Territorio
- Demografía y Población
 - Cifras de población y Censos demográficos
 - + Padrón
 - Censos de Población
 - Censos de población y viviendas. Año 2011
 - + Residentes en viviendas principales
 - + Hogares y familias
 - Edificios y viviendas
 - Edificios según tipo**
 - Estado de los edificios
 - Edificios según antigüedad
 - Tipo de vivienda
 - Tamaño de la vivienda
 - Régimen de tenencia de la vivienda
 - Superficie útil de las viviendas
 - Equipamientos de las viviendas
 - + Censos de población y viviendas. Año 2001
 - + Datos históricos
 - + Indicadores demográficos
- + Educación y Formación

Edificios y viviendas
Edificios según antigüedad. Municipios.

Unidades: número de edificios destinados principal o exclusivamente a viviendas.
Fuente: Censos de población y viviendas, 2011. Aragón. IAEST.
[Metodología](#)

50129 Jaraba

Año de construcción	Total edificios
Total	269
Antes 1900	15
1900 a 1920	3
1921 a 1940	4
1941 a 1950	91
1951 a 1960	24
1961 a 1970	18
1971 a 1980	19
1981 a 1990	40
1991 a 2001	9
2002	7
2003	7

PDF 3
Excel 9
PowerPoint 6
Archivo Web (.mht) 4
Datos 5

Formato CSV 2
Formato Delimitado por Tabuladores
Formato XML

[Imprimir](#) - [Exportar](#)

En la opción de Exportar, que se muestra en la figura anterior, se pueden observar los distintos formatos de exportación en los que se pueden hacer los datos disponibles. Concretamente, nuestro interés está en el formato CSV, puesto que es el más sencillo de tratar automáticamente sin necesidad de acceder a las bases de datos del Data Warehouse directamente (y por tanto asegurando las condiciones de acceso impuestas por el IAEST a determinados tipos de datos).

Este informe, como se puede observar, contiene 19 filas, donde se especifica el año de construcción y el número de viviendas correspondientes en esta localidad. Por tanto, se trata de un informe sencillo donde sólo se cuenta con una dimensión (el año de construcción) y una medida (el número de viviendas). Este es el informe que utilizaremos a partir de este momento para la descripción del trabajo de transformación realizado.

3.2 Transformación de datos al formato W3C RDF DataCube

El vocabulario RDF DataCube (<https://www.w3.org/TR/vocab-data-cube/>) es una recomendación del consorcio W3C para la publicación de datos multi-dimensionales, como es el caso de los datos estadísticos. Numerosas organizaciones encargadas de producir y publicar estadísticas oficiales están utilizando este formato para la exportación de sus datos, además de otros formatos como SDMX, PC-Axis o formatos tabulares como los de las hojas de cálculo.

3.2.1 Creación de las estructuras de los cubos de datos

En nuestro trabajo hemos realizado la transformación de los datos disponibles en los CSVs identificados anteriormente a RDF DataCube. Para ello, en primer lugar se ha creado una estructura del cubo de datos (*Data Structure Definition* en términos de RDF DataCube) para los informes de viviendas por antigüedad, que es aplicable no sólo a informes correspondientes a

municipios, como el seleccionado para nuestro ejemplo, sino también a informes de comarcas, provincias y de todo Aragón. Esta estructura de datos tiene la URI <http://opendata.aragon.es/recurso/iaest/dsd/03-02010301-030092>¹, donde está disponible y publicada como datos enlazados, y el dump completo de todas las estructuras de datos que han sido publicadas a partir de los datos de estadística local está disponible en <https://github.com/aragonopendata/local-data-aragopedia/blob/master/data/dump/DatosTTL/dataStructures/dsd.ttl>.

Las siguientes tripletas describen esta estructura de datos, indicando que tiene como componentes del cubo de datos las dimensiones de área y periodo (común en todos los cubos de datos), y año de construcción (ano-de-construccion), y la medida de número de edificios (n-edificios):

```
<http://opendata.aragon.es/recurso/iaest/dsd/03-02010301-030092> a
qb:DataStructureDefinition ;
    rdfs:label "Estructura de los cubos de datos que se
corresponden con los informes 03-02010301-030092, Edificios según
antigüedad"@es ;
    skos:notation "DSD-03-02010301-030092" ;
    rdfs:comment "Esta estructura afecta a las areas: TC TM TP A" .
```

```
<http://opendata.aragon.es/recurso/iaest/dsd/03-02010301-030092>
qb:component _:node79egmfx1 .
_:node79egmfx1 qb:dimension sdmx-dimension:refArea .
```

```
<http://opendata.aragon.es/recurso/iaest/dsd/03-02010301-030092>
qb:component _:node79egmfx7 .
_:node79egmfx7 qb:dimension sdmx-dimension:refPeriod .
```

```
<http://opendata.aragon.es/recurso/iaest/dsd/03-02010301-030092>
qb:component _:node79egmfx2 .
_:node79egmfx2 qb:dimension iaest-dimension:ano-de-construccion .
```

```
<http://opendata.aragon.es/recurso/iaest/dsd/03-02010301-030092>
qb:component _:node79egmfx5 .
_:node79egmfx5 qb:measure iaest-measure:n-edificios .
```

3.2.2 Generación de los datos de los cubos de datos

Los datos RDF generados para este informe están disponibles en la siguiente URI: <http://opendata.aragon.es/recurso/iaest/dataset/03-02010301-030092TM> , y el RDF completo se puede obtener en el siguiente enlace:

<https://github.com/aragonopendata/local-data-aragopedia/blob/master/data/dump/DatosTTL/informes/03-02010301-030092TM.ttl>

¹ Para que todas las URIs presentadas en esta ponencia produzcan resultados, debe estar almacenada una API key como cookie en el navegador, o incluirla en la llamada utilizando: ?api_key=e103dc13eb276ad734e680f5855f20c6 (tal y como se explica en la documentación correspondiente: <http://opendata.aragon.es/portal/desarrolladores/api-aragodbpedia>)

Por ejemplo, las siguientes tripletas describen una de las líneas del informe anterior (la que dice que en el municipio de Jaraba hay 15 viviendas construidas antes del 1900):

```
<http://opendata.aragon.es/recurso/iaest/observacion/03-02010301-030092TM/11948e35-5725-3b7f-ad87-210875955322> a qb:Observation ;
    qb:dataSet <http://opendata.aragon.es/recurso/iaest/dataset/03-02010301-030092TM>;
    iaest-dimension:ano-de-construccion "Antes 1900"^^xsd:string;
    sdmx-dimension:refArea
<http://opendata.aragon.es/recurso/territorio/Municipio/Jaraba> ;
    iaest-measure:n-edificios "15"^^xsd:int;
    sdmx-dimension:refPeriod
<http://reference.data.gov.uk/id/year/2011> .
```

Como se puede observar, además de los datos que aparecen en el CSV correspondiente, se ha añadido el año al que se refiere el informe (2011), algo que se ha tenido que configurar manualmente para cada informe disponible en la base de datos de Estadística Local.

3.2.3 Generación de las propiedades correspondientes a dimensiones y medidas

Asimismo, se han generado automáticamente las propiedades correspondientes a dimensiones y medidas utilizando el lenguaje RDF Schema. El fichero de definición de estas dimensiones y medidas, que se regenera en el caso de que aparezcan nuevas dimensiones o medidas, o desaparezcan o se modifiquen las existentes, está disponible en la siguiente URL:

<https://github.com/aragonopendata/local-data-aragopedia/blob/master/data/dump/DatosTTL/codelists/properties.ttl>

Por ejemplo, a continuación se puede observar la definición de la propiedad de dimensión para el año de construcción de las viviendas. Estas propiedades pueden ser compartidas por varios informes.

```
iaest-dimension:ano-de-construccion a qb:DimensionProperty ,
rdf:Property, qb:CodedProperty ;
    rdfs:label "Año de construcción"@es ;
    rdfs:comment "Año de construcción"@es ;
    rdfs:range xsd:string .
```

Todas estas propiedades se encuentran también disponibles en su propia URI, con el objetivo de poder ser consultadas y recuperadas en cualquier momento por cualquier usuario o máquina.

3.2.4 Generación de listas de códigos

También se han generado listas de códigos (*codelists* de acuerdo con la nomenclatura habitualmente usada en el modelo SKOS) para aquellos conjuntos de valores que normalmente toman valores de una lista determinada. Este no es el caso en el informe seleccionado, pero recurrimos a otro informe distinto (viviendas agrupadas por tipo) para mostrar un ejemplo de lista de códigos:

```
<http://opendata.aragon.es/kos/iaest/tipo-vivienda> a
skos:ConceptScheme ;
    skos:notation "tipo-vivienda";
```

```

    rdfs:label "Tipo vivienda";
    skos:hasTopConcept <http://opendata.aragon.es/kos/iaest/tipo-
vivienda/residencias-de-trabajadores>;
    skos:hasTopConcept <http://opendata.aragon.es/kos/iaest/tipo-
vivienda/instituciones-religiosas>;
    skos:hasTopConcept <http://opendata.aragon.es/kos/iaest/tipo-
vivienda/vacias>;
    skos:hasTopConcept <http://opendata.aragon.es/kos/iaest/tipo-
vivienda/internados-academias-y-escuelas-militares-seminarios>;
    skos:hasTopConcept <http://opendata.aragon.es/kos/iaest/tipo-
vivienda/secundarias>;
    skos:hasTopConcept <http://opendata.aragon.es/kos/iaest/tipo-
vivienda/convencionales>;
    skos:hasTopConcept <http://opendata.aragon.es/kos/iaest/tipo-
vivienda/instituciones-penitenciarias>;
    skos:hasTopConcept <http://opendata.aragon.es/kos/iaest/tipo-
vivienda/asilos-o-residencias-de-ancianos>;
    skos:hasTopConcept <http://opendata.aragon.es/kos/iaest/tipo-
vivienda/colegios-mayores-residencias-de-estudiantes>;
    skos:hasTopConcept <http://opendata.aragon.es/kos/iaest/tipo-
vivienda/otro-tipo-de-colectivo>;
    skos:hasTopConcept <http://opendata.aragon.es/kos/iaest/tipo-
vivienda/otro-tipo>;
    skos:hasTopConcept <http://opendata.aragon.es/kos/iaest/tipo-
vivienda/instituciones-para-personas-con-discapacidades>;
    skos:hasTopConcept <http://opendata.aragon.es/kos/iaest/tipo-
vivienda/albergues-para-marginados-sociales>;
    skos:hasTopConcept <http://opendata.aragon.es/kos/iaest/tipo-
vivienda/hospitales-de-larga-estancia>;
    skos:hasTopConcept <http://opendata.aragon.es/kos/iaest/tipo-
vivienda/establecimientos-militares>;
    skos:hasTopConcept <http://opendata.aragon.es/kos/iaest/tipo-
vivienda/hoteles-pensiones-albergues>;
    skos:hasTopConcept <http://opendata.aragon.es/kos/iaest/tipo-
vivienda/otras-instituciones-de-asistencias-social-a-la-infancia-
juventud>;
    skos:hasTopConcept <http://opendata.aragon.es/kos/iaest/tipo-
vivienda/hospitales-psiquiatricos>;
    skos:hasTopConcept <http://opendata.aragon.es/kos/iaest/tipo-
vivienda/alojamientos>;
    skos:hasTopConcept <http://opendata.aragon.es/kos/iaest/tipo-
vivienda/hospitales-generales-y-especiales-de-corta-estancia>.

```

Nuevamente, todas estas listas de códigos y las URIs específicas se encuentran publicadas en su URI correspondiente, con el objetivo de facilitar su uso por cualquier reutilizador.

4. Publicación de datos como datos enlazados y en una API

Como ya se ha comentado anteriormente, todas las URIs que se han creado, tanto para observaciones específicas como para dimensiones, medidas, listas de códigos, estructuras de datos, etc., están publicadas como datos enlazados, lo que quiere decir que se pueden recuperar haciendo peticiones HTTP GET a dichas URIs (y añadiendo la API key que sirve para poder asegurar una calidad de servicio adecuada en los datos que se sirven en todas las APIs de datos enlazados del portal de datos abiertos de Aragón). Esta publicación se ha realizado mediante la tecnología ELDA, que ya había sido utilizada en el pasado en la

publicación de datos en Aragón, y el uso de los mecanismos de autenticación y autorización proporcionados por el servicio 3scale.

Asimismo, los datos están cargados en el punto de acceso SPARQL del Gobierno de Aragón (<http://opendata.aragon.es/sparql>), lo que permite realizar consultas tanto simples como complejas, pudiendo realizar agregaciones sobre uno o varios cubos de datos. Algunos ejemplos de consultas posibles están disponibles en el repositorio GitHub del proyecto: <https://github.com/aragonopendata/local-data-aragopedia/blob/master/consultas.md>.

Finalmente, se ha creado una API REST más completa, que permite no solo consultar las URIs que se están exponiendo como datos enlazados, sino también realizar operaciones de búsqueda de cubos de datos, dimensiones, propiedades, etc., que pueden ser útiles a la hora de construir aplicaciones sobre los datos que han sido generados. La documentación de esta API se encuentra disponible en <http://opendata.aragon.es/bancodatos/doc/> (aunque esta localización podría cambiar, pero será accesible a partir de la Web general de Open Data Aragón).

5. Conclusiones

En esta ponencia se ha presentado el resultado del trabajo realizado para la generación y publicación de datos enlazados a partir del banco de datos de estadística local publicado por el Instituto Aragonés de Estadística. Esta generación se ha realizado siguiendo estándares propuestos por el W3C y adoptados por instituciones de publicación de datos estadísticos internacionales.

Durante el proceso, que se encuentra totalmente automatizado para poder actualizar los datos cada vez que se producen actualizaciones diarias de datos en el banco de datos, se han podido resolver algunas inconsistencias en el nombrado de columnas o dimensiones, así como homogeneizar la producción de algunos datos para columnas que son comunes para varios informes. Este proceso de depuración y curación de datos se sigue realizando en la actualidad, para continuar mejorando la generación de datos que se está realizando.

El proceso es replicable para cualquier otro instituto oficial de estadística que pueda proporcionar de manera generalizada CSVs de los cubos de datos que desee publicar como datos enlazados, requiriendo únicamente una configuración inicial que está parcialmente automatizada, para facilitar la generación de los ficheros de dimensiones, medidas, listas de códigos y estructuras de datos.