

Publicación de Datos de Compras Estatales en Linked Data

Antecedentes

La Agencia de Compras y Contrataciones del Estado (ACCE) es una agencia del estado uruguayo cuya misión es "promover la mejora de gestión y la transparencia del sistema de compras y contrataciones del sector público, brindando asesoramiento para la generación de políticas públicas, el marco normativo, la tecnología, la capacitación necesaria y amplio acceso a la información generada".¹

Esta agencia brinda acceso a los datos de compras y licitaciones que se realizan en el estado.² El acceso a los mismos se provee a través de un formulario web³, el cual permite seleccionar algunos parámetros del conjunto de datos a descargar. En la Figura 1 se presenta una captura de pantalla de dicho formulario.

Interfaz de Consulta de Publicaciones

A través de esta interfaz usted puede obtener todos los datos en tiempo real relativos a las compras que se publican en el sitio web de Compras Estatales, en *formato XML*.

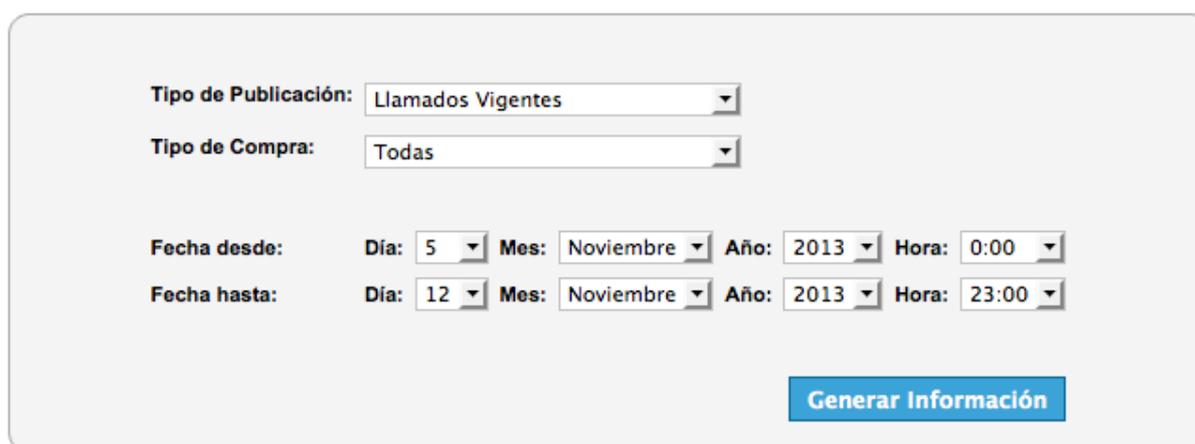


Figura 1: Formulario para obtener datos de compras estatales

Los datos se devuelven en un documento XML, y para cada solicitud se genera una URL donde se indican todos los parámetros de la consulta. A modo de ejemplo se presenta a continuación la URL generada por el formulario de la Figura 1

https://www.comprasestatales.gub.uy/comprasenlinea/jboss/generarReporte?tipo_publicacion=lv&tipo_compra=&dia_inicial=05&mes_inicial=11&anio_inicial=2013&hora_inicial=0&dia_final=12&mes_final=11&anio_final=2013&hora_final=23&Generar+Archivo=Generar+Informaci%F3n

Esta interfaz de consulta impone una restricción importante respecto a los datos que se pueden descargar en cada solicitud: el período a consultar no puede ser mayor a 10 días. Dado que este es el único punto de acceso a los datos, esto implica que no es posible realizar una descarga del juego de datos completo y por lo tanto, si bien la información está publicada y en un formato procesable, no es consultable directamente dado que es necesario generar y descargar cada archivo XML para luego procesarlo.

1 <http://www.comprasestatales.gub.uy/inicio/institucional/plan-estrategico/>

2 <http://www.comprasestatales.gub.uy/inicio/datos-abiertos/>

3 <https://www.comprasestatales.gub.uy/inicio/datos-abiertos/interfaz-consulta-publicaciones/interfaz-consultas-publicaciones>

Descripción del Proyecto

El objetivo del proyecto es desarrollar estrategias para la publicación de los datos de compras estatales en Linked Data, permitiendo la realización de consultas en SPARQL sobre los mismos, y probar esas estrategias con la publicación de parte del conjunto de datos disponible. A continuación se describen los componentes de la solución a desarrollar:

1- Como etapa previa a la publicación propiamente dicha es necesario definir un **vocabulario** y una **ontología** compatibles, con RDFS (RDFS u OWL 2), que modelen adecuadamente los conceptos y propiedades que hoy maneja el sitio de Compras Estatales.

El proceso de definición de estos artefactos debe ser documentado y deberá incluir:

- una revisión de vocabularios existentes que sean pertinentes (DC, FOAF, etc.)
- un relevamiento de los mecanismos de razonamiento (o clasificación) que se esperan aplicar luego sobre los datos publicados.

2 – Se debe proponer e implementar una arquitectura que permita la extracción de los datos y metadatos actualmente accesibles via el formulario web descrito anteriormente, y su transformación y publicación como Linked Data, utilizando el vocabulario y la ontología definidos en el punto 1. Dado que los datos del sistema fuente pueden cambiar se desea explorar mecanismos de generación de triplas dinámicos. Para esto hay al menos dos estrategias:

- Utilizar productos existentes que permiten el acceso directo a un grafo virtual y consultarlo directamente. Ej: Virtuoso usando Sponger.
- Utilizar productos que permiten realizar la transformación de documentos XML en triplas RDF y luego almacenar el resultado en un triplestore. (ej: extensión para Google Refine⁴)
- Programar una solución desde cero, en base a lenguajes de programación web (ej: php), que realicen la transformación y devuelvan triplas RDF.

Hay que tener en cuenta que se debería publicar tanto los datos como los metadatos de cada dataset que se publique o acceda. Uno de los objetivos del proyecto es evaluar estas opciones y hacer al menos una prueba de concepto de una de ellas.

Alcance y Resultados

El alcance del proyecto estará regulado por los resultados intermedios que se obtengan. Sin embargo, es fundamental contar con uno o más informes que justifiquen la toma de decisiones.

También se espera disponer de:

- La definición de la ontología (o vocabulario) involucrado en la publicación.
- La descripción de los mecanismos utilizados de forma que se puedan reproducir los resultados.
- Una prueba de concepto de cada procedimiento.

4 <http://refine.deri.ie/>