

Publicación del Catálogo del Centro de Fotografía de la Intendencia de Montevideo con Linked Data

Antecedentes

El Centro de Fotografía (CDF) de la Intendencia de Montevideo posee un acervo importante de fotografías. Algunas de ellas son digitales pero la mayoría son en papel, negativo o incluso vidrio.

La mayor parte de esas fotografías están catalogadas siguiendo normas archivísticas (ISAD(G): General International Standard Archival Description¹) dictadas por el International Council on Archives² (ICA) y teniendo en cuenta su alineamiento con Dublin-Core³. En base a estas normas el CDF ha definido un subconjunto de propiedades que son utilizadas para describir cada fotografía. Además se cuenta con un tesoro de términos que son utilizados como valores de algunas de estas propiedades.

Por otra parte la Intendencia de Montevideo viene llevando adelante políticas de publicación de Datos Abiertos en Uruguay⁴. La publicación de parte de este catálogo es un paso más en esas políticas.

Actualmente, los datos del catálogo se encuentran representados en una base de datos relacional, mientras que la mayoría de los metadatos (ej: descripción de propiedades, tesoro) se encuentra en documentos.

Descripción del Proyecto

El objetivo del proyecto es desarrollar una estrategia a seguir para la publicación del catálogo del CDF en Linked Data y probar esa estrategia con la publicación de parte del catálogo. 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 catálogo del CDF.

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

- una revisión de vocabularios y/o ontologías existentes directamente relacionados a ISAD-(G)
- una revisión de otros 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 actualmente almacenados en un RDBMS y su transformación y publicación como Linked Data, utilizando el vocabulario y la ontología definidos en el punto 1. Dicha arquitectura debe incluir la funcionalidad de SPARQL *endpoint* y permitir realizar consultas SPARQL 1.1 sobre los datos publicados, de ser posible este *endpoint* deberá ser capaz de soportar diferentes regímenes de inferencia (*entailment regimes*). Además sería deseable contar con alguna interfaz gráfica que permita navegar sobre los datos publicados.

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:

- Una descripción del proceso de definición de la ontología.

1 <http://www.ica.org/10207/standards/isadg-general-international-standard-archival-description-second-edition.html>

2 <http://www.ica.org/3/homepage/home.html>

3 <http://dublincore.org/>

4 <http://www.montevideo.gub.uy/institucional/datos-abiertos/introduccion>

- Una descripción del procedimiento de publicación.
- Un prototipo implementado y funcional de una parte del catálogo.
- Una descripción de las potencialidades de disponer de esta publicación.

Algunas preguntas que pueden quedar. (FAQ)

- ***Por qué definir un vocabulario y una ontología? No son básicamente lo mismo?***
 - Un **vocabulario**, es solamente un conjunto de términos con un uso definido para cada término. Ese uso está dado de diferentes formas. Algunos están dados como descripciones textuales, otros a través de una descripción en OWL 2, otros usando XMLSchema.
 - Una **ontología**, se puede pensar como una estructura lógica similar a una estructura de primer orden o como un esquema de una base de datos. Allí se definen conjuntos de datos (en términos lógicos, predicados unarios) y relaciones entre esos predicados (en términos lógicos, predicados n-arios con $n > 1$) con restricciones de integridad entre todos ellos que regulan que instancias son válidas (en términos lógicos, modelos) y que instancias no lo son. Una ontología, induce un vocabulario, pero la ontología define la forma en que dicho vocabulario debe ser interpretado y permite definir reglas que luego son utilizadas en la etapa de razonamiento.