

Mejorando la ciencia abierta utilizando datos abiertos enlazados: caso de uso de CONICET Digital

Marcos Zárate, y otros

Journal of computer science & technology, ISSN 1666-6038, Vol. 19, n. 1, 2019

La ciencia abierta es un movimiento cuyo objetivo es la accesibilidad de la investigación científica para todos los ciudadanos. La ciencia abierta incrementa y estimula la producción del conocimiento científico, innova con el uso de tecnologías, promueve el valor de compartir, reutiliza y permite que los datos, informes, y otras partes del proceso de investigación estén disponibles para todos. En este contexto, *CONICET Digital* es un repositorio de acceso abierto perteneciente al Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), la principal organización dedicada a la promoción de la ciencia y la tecnología en Argentina. CONICET Digital se creó con el objetivo de reunir, registrar, divulgar, preservar y dar acceso público a la producción científico-tecnológica realizada en el CONICET. En este artículo se presenta un trabajo en curso para publicar un subconjunto de publicaciones científicas de CONICET Digital como datos abiertos enlazados. El objetivo es mejorar la recuperación y la reutilización de datos a través de las tecnologías de la web semántica y de los datos enlazados en el ámbito de las publicaciones científicas. Para lograr estos objetivos, se han tenido en cuenta los estándares de la web semántica y los esquemas RDF (Dublin Core, FOAF, VoID, etc.). CONICET Digital adoptó la plataforma DSpace para implementar el repositorio. El proceso de conversión y publicación se basa

en las pautas metodológicas para publicar datos vinculados de gobierno. En el trabajo se describen los principales conjuntos de datos enlazados de publicaciones científicas accesibles a través de SPARQL (Springer Nature SciGraph, DBLP, Wikidata, SPAR Ontologies, OpenCitations Corpus). Explica las etapas del ciclo de vida para este trabajo, como establece el W3C Working Group para el proceso de publicación de conjuntos de datos como datos enlazados. Este proceso tiene un modelo de ciclo de vida incremental iterativo, que se basa en la mejora continua y en la extensión de los datos enlazados resultantes de realizar varias iteraciones. Se presentan estudios de casos que permiten recuperar información de diferentes conjuntos de datos. Se describe la plataforma propuesta, una arquitectura de varios niveles que separa la funcionalidad en varias capas desde el almacenamiento de datos de bajo nivel hasta los componentes de interacción del usuario. En futuros trabajos los autores planean continuar desarrollando algunos aspectos: la automatización del proceso de extracción de datos utilizando OpenRefine Python; la integración de un marco de recuperación automática de conexiones como Silk; y el desarrollo de una aplicación web para navegar por publicaciones científicas en el campo de los investigadores del CONICET.

Resumen elaborado por Natividad Escavias Extremera