

Integración de Datos y Servicios Geoespaciales entre Infraestructuras de Datos Espaciales y Repositorios de Datos Institucionales Abiertos

Roda, Fernando^{1,2,3}; Mellino, Natalia^{1*}; Casali, Ana^{1,3}; Balparda, Laura R.¹; López, Diego A. G.¹; Mata, Ramiro¹; Mestre, Diego L.¹; Bongiovani, Paola C.⁴

¹Facultad de Ciencias Exactas, Ingeniería y Agrimensura, Universidad Nacional de Rosario. ²Instituto de Altos estudios Espaciales M. Gulich (CONAE - UNC) y CONICET. ³CIFASIS (UNR - CONICET). ⁴Facultad de Humanidades y Artes, y CEIBCi
*natimellino.nb@gmail.com

Palabras claves: Información Geoespacial; IDE; Repositorios Académicos, Ciencia Abierta, Ontologías.

INTRODUCCIÓN

El trabajo aborda el problema de la fragmentación y duplicación no controlada de datos, metadatos y servicios vinculados a la **información geoespacial** en entornos académicos. Se presentan los avances de un proyecto que toma como caso de estudio **la gestión de datos geoespaciales en la Universidad Nacional de Rosario (UNR)**. En la UNR coexisten dos sistemas independientes: i) El Repositorio de Datos Académicos (**RDA**), (basado en el framework Dataverse) y ii) una Infraestructura de Datos Espaciales (**IDE**), basada en GeoServer. Sin embargo, el RDA carece de indexación espacial, no implementa perfiles de metadatos geoespaciales estándares (ej. ISO 19115) ni brinda acceso mediante servicios geoespaciales (ej. WMS, WFS, entre otros) . Asimismo, la IDE carece de ciertas funcionalidades de los repositorios de datos abiertos institucionales, como la asignación y el mantenimiento de identificadores persistentes (ej. DOI), o la integración con redes de repositorios digitales (ej. LA Referencia). Por lo tanto, ambos sistemas se complementan y una integración mediada por un Agente potenciaría las fortalezas de cada uno.

OBJETIVOS

Mejorar la eficiencia, coherencia y sostenibilidad en la publicación de datos geoespaciales, **potenciando el acceso abierto a la información científica**. A tal fin, se propone **una solución que gestiona los datos, metadatos y servicios geoespaciales de forma integrada** sin renunciar a los servicios de las plataformas RDA e IDE de la UNR .

METODOLOGÍA

La metodología consiste en el desarrollo de una **ontología** basada en vocabularios estándares de Catálogos de datos (DCAT). Dicha ontología surge de una alineación semántica entre los perfiles de metadatos del RDA y el IDE a través del modelo de catálogo **GeoDCAT-AP**; el cual es extendido para cubrir tanto la catalogación de los datos geoespaciales como de los servicios de acceso a los mismos (a cargo de la IDE). Sobre las bases de dicha ontología, el trabajo presenta el diseño de una arquitectura de sistema para la integración entre el RDA y la IDE, contemplando esquemas de migración y cosecha de metadatos implementados por un Agente de Vinculación.

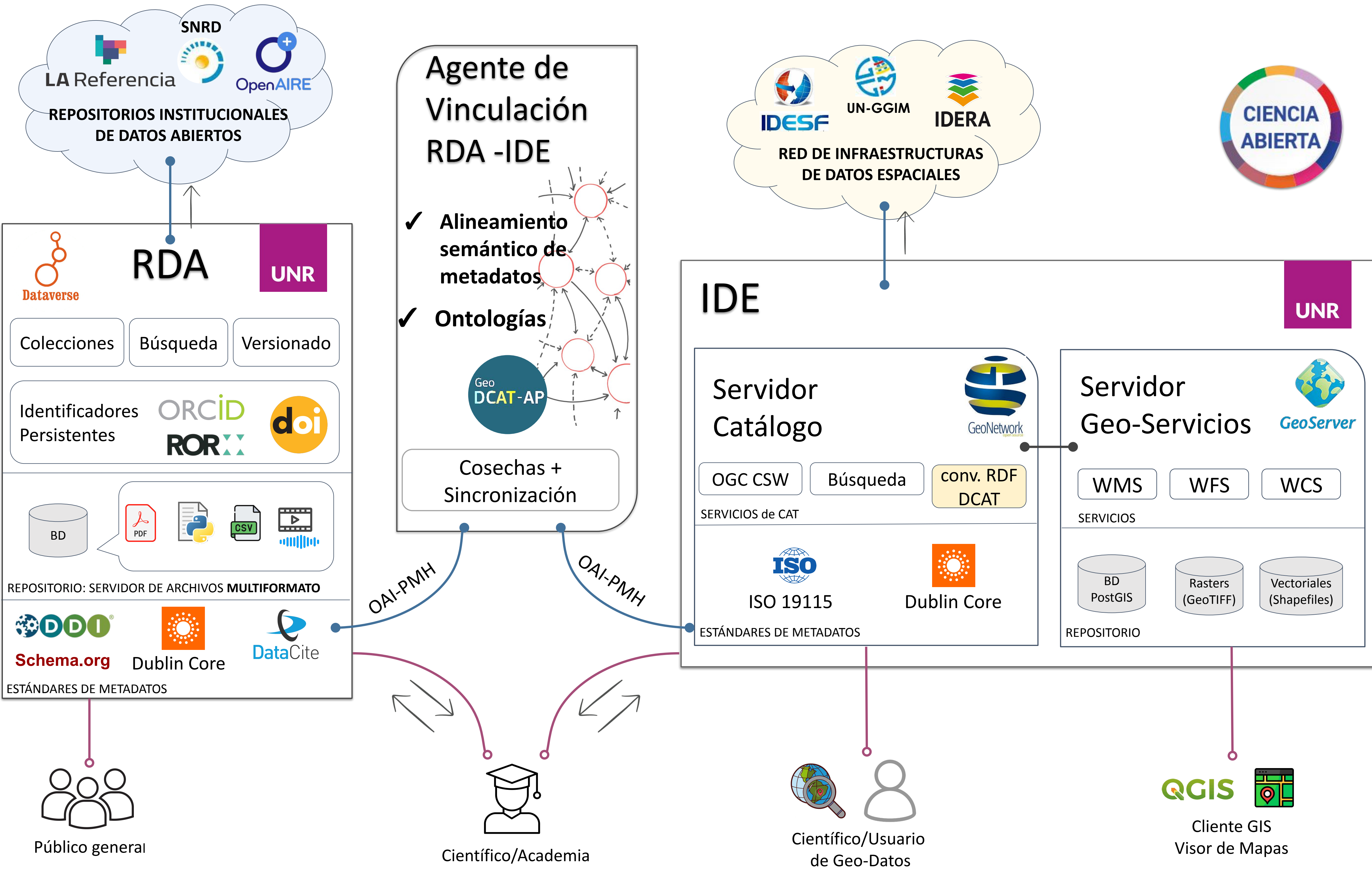


Figura 1. Arquitectura Propuesta: Agente de Vinculación basado en metadatos semánticos y cosecha automatizada.

DISCUSIÓN / PRÓXIMOS PASOS

Nuestra hipótesis establece que estos sistemas al ser integrados pueden *brindar mejores niveles de cumplimiento de los principios FAIR* en lo que refiere a la publicación de ítems académicos de proyectos cuyos resultados generan información geoespacial (coordenadas, mapas, manuscritos, etc). Como próximos pasos, se desarrollará un prototipo funcional de dicha arquitectura que será validado mediante casos de uso reales, como los datos de la Huella Urbana en Rosario obtenidos por un grupo de UNR. Además, se explorará el uso de los modelos de lenguaje grande (LLM) para la generación de descriptores de datasets a partir de los metadatos disponibles a través de la IDE, el RDA y la ontología GeoDCAT.