

# Análisis Visual para Datos Abiertos Enlazados vinculados a las Ciencias del Mar

Gustavo Nuñez<sup>1</sup>, Carlos Buckle<sup>1</sup>[0000-0003-0722-0949], Marcos Zárate<sup>1,2</sup>[0000-0001-8851-8602]

<sup>1</sup> Laboratorio de Investigación en Informática, Facultad de Ingeniería, Universidad Nacional de la Patagonia San Juan Bosco (LINVI-UNPSJB), Puerto Madryn, Argentina.

<sup>2</sup> Centro para el Estudio de Sistemas Marinos, Centro Nacional Patagónico, Consejo Nacional de Investigaciones Científicas y Técnicas (CESIMAR-CENPAT-CONICET), Puerto Madryn, Argentina.  
gnunez@ing.unp.edu.ar, cbuckle@unpata.edu.ar,  
zarate@cenpat-conicet.gob.ar

**Resumen:** El propósito de la exploración y visualización de datos (DV) es ofrecer formas de percibir y manipular información, así como extraer e inferir conocimiento. En este breve artículo presentamos avances en representaciones visuales y técnicas de interacción intuitivas basadas en inteligencia artificial. Esto contribuye significativamente a la exploración y comprensión de la información relacionada con las ciencias marinas representada por ontologías y datos enlazados. Esta investigación preliminar llevada adelante desde 2018 entre LINVI-UNPSJB y CESIMAR-CENPAT-CONICET permitirá a los científicos y usuarios no expertos analizar conjuntos de información relacionada con la oceanografía, meteorología y parámetros ambientales, con el fin de promover el conocimiento científico y la innovación productiva en el océano Atlántico Sur utilizando Datos Abiertos Enlazados (LOD por siglas en inglés).

**Palabras clave:** Visualización de Datos · Datos Abiertos Enlazados · Ciencias Marinas.

## 1 Introducción

El propósito de la DV es ofrecer formas de percibir y manipular la información, así como extraer e inferir conocimiento [1,2]. La DV proporciona a los usuarios una manera intuitiva de explorar el contenido, identificar patrones de interés e inferir correlaciones y causalidades, además de brindar un gran aporte a las actividades de construcción de significado. Uno de los enfoques más prometedores para abordar la problemática asociada con la integración y la representación gráfica es almacenar los datos de forma estructurada y reproducirlos mediante gráficos. La Web Semántica (SW) [3] ofrece soluciones a estas necesidades al utilizar LOD [4], en donde las entidades se identifican de forma única y las relaciones entre ellas se especifican explícitamente. LOD es un enfoque potente que permite difundir y consumir datos

científicos de varias disciplinas [5,6,7,8]. Implica publicar, compartir y conectar datos en la Web.

En los últimos años, esta forma de publicar datos ha sido adoptada en un gran número de disciplinas LOD [9]. Esto ha hecho que la visualización y exploración de información sea una tarea crucial para la mayoría de los consumidores LOD. Científicos de datos, expertos del dominio y usuarios no experimentados buscan maneras intuitivas y visuales de interactuar con estos recursos. En el campo de las ciencias marinas, la visualización de datos en disciplinas como Oceanografía, Meteorología y la Biodiversidad enfrentan grandes desafíos, ya que existe un aumento exponencial de su volumen debido al crecimiento de las tecnologías y la multiplicidad de plataformas, además de la demanda de conocimiento para contribuir globalmente a los distintos modelos que buscan combatir el cambio climático [10]. Además, existe una gran diversidad en el tipo de registros que deben mostrarse adecuadamente, las características físicas, químicas, geológicas, meteorológicas y los valores biológicos deben integrarse correctamente, y los productos de análisis/información deben basarse en todos ellos para que el usuario pueda hacer una interpretación correcta [11].

El resto de este documento breve está estructurado de la siguiente manera: la Sección 2 presenta diferentes iniciativas basadas en LOD para las ciencias marinas. La sección 3 presenta una prueba conceptual de una plataforma desarrollada para visualizar información relacionada a las ciencias marinas en el Atlántico Sur. Finalmente, en la sección 4, presentamos algunas conclusiones basadas en experiencias y la planificación de trabajos futuros.

## 2 Antecedentes y trabajos relacionados

En los últimos años se han introducido una gran cantidad de herramientas de visualización de LD, la mayoría provenientes del ámbito académico. Las herramientas de DV en datos enlazados proporcionan representaciones gráficas de un conjunto de datos o partes de él, con el fin de facilitar su análisis y generación de conocimientos a partir de información compleja e interrelacionada en tiempo y espacio. Las técnicas pueden variar según el dominio, el tipo de registro, la tarea que el usuario está tratando de realizar, así como las habilidades del usuario.

Son varias las iniciativas que se llevan a cabo en el contexto argentino para publicar datos de ciencias marinas como LD, entre ellos podemos destacar: [12] que presenta la publicación de metadatos de campañas oceanográficas como LD. OceanGraph [13] define un prototipo de gráfico de conocimiento oceanográfico para gestionar información de expediciones, publicaciones científicas y variables ambientales, mientras que en [14] se propone la explotación de OceanGraph con ejemplos concretos de posibles usos por especialistas.

A nivel internacional también existen iniciativas para la publicación de información marina de datos científicos como LD, entre los principales podemos mencionar GeoLink [15], un proyecto financiado por la iniciativa EarthCube, que ha aprovechado los principios de LOD para crear una base de datos que permite a los usuarios realizar consultas en algunos de los repositorios de geociencias más destacados de los Estados Unidos. El conjunto de datos de GeoLink incluye información tan diversa como escalas

en puertos realizadas por cruceros oceanográficos, metadatos de muestras físicas, financiación de proyectos de investigación y personal, y autoría de informes técnicos. Los datos han sido publicados de acuerdo con las mejores prácticas para LOD [16], y están disponibles públicamente a través de un endpoint SPARQL<sup>1</sup> que actualmente contiene más de 45 millones de tripletas RDF.

### 3 Visualización LD en Ciencias Marinas

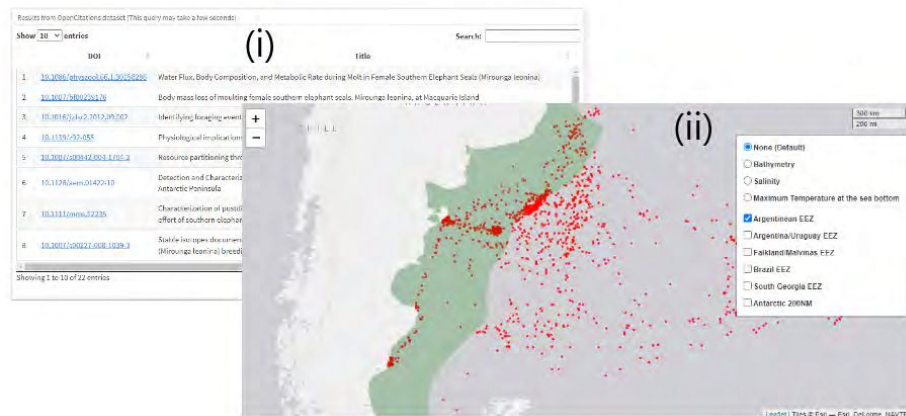
En el contexto de las ciencias marinas, la exploración visual es un enfoque prometedor para explorar y analizar datos y comprender mejor la dinámica de los complejos procesos oceánicos. Aunque la publicación de datos como LD tiene varios casos de éxito [15,17], la visualización sigue siendo un problema porque es una tarea que difiere del clásico DV, principalmente debido a las características de LD. Los usos de vocabularios comunes (dominios cruzados) para la descripción de los registros, o el uso de propiedades tipificadas para capturar las relaciones entre los recursos dentro un conjunto o entre diferentes conjuntos, difieren de las formas tradicionales de visualización que son incapaces de captar las complejas relaciones posibles. Para las pruebas descritas a continuación, utilizamos información pública sobre especies marinas y variables capturadas en el Atlántico Sur a través de un endpoint SPARQL cuya URL es <http://linkeddata.cenpat-conicet.gob.ar/snorql/>. La metodología utilizada para la creación y publicación se detalla en [17].

#### 3.1 Casos de estudio

Nuestro enfoque está puesto en el front-end basado en la Web, más precisamente en herramientas de consultas y visualización. Hemos desarrollado una prueba de concepto para la visualización interactiva de información oceanográfica, ambiental y de biodiversidad marina. La plataforma permite la representación y visualización de mapas interactivos con trayectorias de buques oceanográficos, y la recuperación de esquemas gráficos de la relación entre variables ambientales y especies. Para ello, se llevó a cabo una selección de herramientas de código abierto compatibles con la visualización de tipos específicos de información, por ejemplo, datos geoespaciales, distribución de especies, trazabilidad y registros relacionados con el medio ambiente. La Figura 1 muestra dos visualizaciones utilizadas para interpretar información sobre una especie específica, en este caso Mirounga Leonina (elefante marino del sur). El mapa muestra la información de los viajes realizados por varios individuos durante sus viajes de alimentación en el mar, superponiéndose adicionalmente capas con información ambiental y especial. La otra visualización muestra información bibliográfica asociada a la especie.

---

<sup>1</sup> <http://data.geolink.org/sparql>



**Fig. 1.** Visualizaciones utilizadas para relacionar: (i) especies marinas con información bibliográfica (ii) información geoespacial de especies con variables ambientales y regiones marinas.

#### 4 Conclusiones y trabajos futuros

De experiencias anteriores, podemos concluir que resulta necesario desarrollar sistemas que sean capaces de gestionar visualmente la información para usos integrales y secundarios, tanto de los colectivos participantes como de usuarios externos que requieren información. Los resultados de esta investigación preliminar constituyen un aporte sustancial, no solo para las ciencias del mar, sino también como aporte metodológico a visualizaciones científicas usando LD.

Como trabajo futuro, se plantea la necesidad de formalizar la prueba de concepto. Para ello es necesario profundizar en los siguientes aspectos: a) Estudio e investigación de visualizaciones científicas típicas de las ciencias marinas. b) Desarrollar una plataforma de base de datos en línea para la visualización basada en modelos de predicción, con el fin de proporcionar herramientas visuales analíticas y permitir consultas y análisis interactivos de diferentes capas de información. c) Ampliar la plataforma o escalar los resultados a otros espacios marinos, en particular a las Áreas Geográficas Prioritarias (AGP) de la iniciativa Pampa Azul.

#### Referencias

1. Jeffrey Heer y Ben Shneiderman. Interactive dynamics for visual analysis. *Communications of the ACM*, 55(4):45–54, 2012.
2. Stratos Idreos, Olga Papaemmanouil, y Surajit Chaudhuri. Overview of data exploration techniques. En *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, páginas 277–281, 2015.
3. Tim Berners-Lee, James Hendler, Ora Lassila, et al. The semantic web. *Scientific american*, 284(5):28–37, 2001.

4. Christian Bizer, Tom Heath, y Tim Berners-Lee. Linked data: The story so far. En *Semantic services, interoperability and web applications: emerging concepts*, páginas 205–227. IGI Global, 2011.
5. Richard K Lomotey y Ralph Deters. Terms extraction from unstructured data silos. En *System of Systems Engineering (SoSE), 2013 8th International Conference on*, páginas 19–24. IEEE, 2013.
6. Syed Ahmad Chan Bukhari, Mate Levente Nagy, Paolo Ciccarese, Michael Krauthammer, y Christopher JO Baker. icyrus: A semantic framework for biomedical image discovery. En *SWAT4LS*, páginas 13–22, 2015.
7. Syed Ahmad Chan Bukhari. *Semantic enrichment and similarity approximation for biomedical sequence images*. PhD thesis, University of New Brunswick (Canada), 2017.
8. Roderic D.M. Page. Ozymandias: a biodiversity knowledge graph. *PeerJ*, 7:e6739, April 2019.
9. The open linked data cloud. <https://lod-cloud.net/>, 2021. [Online; accessed 3-May-2021].
10. Tanu Malik y Ian Foster. Addressing data access needs of the long-tail distribution of geoscientists. En *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, páginas 5348–5351. IEEE, 2012.
11. Alex Hardisty y Dave Roberts. A decadal view of biodiversity informatics: challenges and priorities. *BMC ecology*, 13(1):16, 2013.
12. Marcos Zárate, Pablo Rosales, Pablo Fillottrani, Claudio Delrieux, y Mirtha Lewis. Oceanographic data management: Towards the publishing of pampa azul oceanographic campaigns as linked data. En *Proceedings of the 12th Alberto Mendelzon International Workshop on Foundations of Data Management (AMW 2018)*, 2018.
13. Marcos Zárate, Pablo Rosales, Germán Braun, Mirtha Lewis, Pablo Rubén Fillottrani, and Claudio Delrieux. Oceangraph: Some initial steps toward an oceanographic knowledge graph. In Boris Villazón-Terrazas and Yusniel Hidalgo-Delgado, editors, *Knowledge Graphs and Semantic Web*, páginas 33–40, Cham, 2019. Springer International Publishing.
14. Marcos Zárate, Carlos Buckle, Renato Mazzanti, Mirtha Lewis, Pablo Fillottrani, y Claudio Delrieux. Harmonizing big data with a knowledge graph: Oceangraph kg uses case. En Enzo Rucci, Marcelo Naiouf, Franco Chichizola, and Laura De Giusti, editors, *Cloud Computing, Big Data & Emerging Topics*, páginas 81–92, Cham, 2020. Springer International Publishing.
15. Michelle Cheatham, Adila Krisnadi, Reihaneh Amini, Pascal Hitzler, Krzysztof Janowicz, Adam Shepherd, Tom Narock, Matt Jones, y Peng Ji. The geolink knowledge graph. *Big Earth Data*, 2018.
16. Krzysztof Janowicz, Pascal Hitzler, Benjamin Adams, Dave Kolas, and Charles Vardeman. Five stars of Linked Data vocabulary use. *Semantic Web*, 5(3):173–176, 2014.
17. Marcos Zárate, Germán Braun, Mirtha Lewis, and Pablo Fillottrani. Observational/hydrographic data of the south atlantic ocean published as lod. *Semantic Web*, 13(2):133–145, 2022.