

# Functional «Semantic Web» Practice in the Marine Research Domain

Marc Portier<sup>1</sup> (marc.portier@vliz.be), Cedric Decruw<sup>1</sup>, (cedric.decruw@vliz.be), Lennert Tyberghein<sup>1</sup> (lennert.tyberghein@vliz.be), Lucy Scott<sup>2</sup> (l.scott@unesco.org), Arno Lambert<sup>3</sup> (a.lambert@unesco.org), Pieter Provoost<sup>3</sup> (p.provoost@unesco.org), Douglas Fils<sup>4</sup> (doug@fils.network), Pier Luigi Buttigieg<sup>5</sup> (pier.buttigieg@awi.de), Bart Vanhoorne<sup>1</sup> (bart.vanhoorne@vliz.be), Simon Claus<sup>6</sup> (simon.claus@vlaanderen.be), Julian Rojas<sup>7</sup> (julianandres.rojasmelendez@ugent.be), Giorgia Lodi<sup>8</sup> (giorgia.lodi@istc.cnr.it), Alessandro Russo<sup>8</sup> (alessandro.russo@istc.cnr.it)

<sup>1</sup>VLIZ, Flanders Marine Institute (Belgium)

<sup>2</sup>IODE (ZA)

<sup>3</sup>IODE (Belgium)

<sup>4</sup>Ronin Institute (USA)

<sup>5</sup>AWI (DE)

<sup>6</sup>Digitaal Vlaanderen (Belgium)

<sup>7</sup>UGent-imec (Belgium)

<sup>8</sup>CNR (Italy)

## Introduction

Independently published documents on distributed web servers delivered almost overnight the compelling browsing experience of «reading a single book», the analogous counterpart of «searching a single database» is still only a vague possibility, despite all efforts to publish structured semantic data on the web. The criticism is that the complexity involved is not balanced quickly enough with tangible result; a sentiment that further cripples motivation and slows progress. To combat that narrative, this talk searches for the opposite: easily achievable (low cost) Linked Open Data Publishing (LOD); and showcasing tools that leverage those efforts to ready, and tangible benefits.

## Keep it simple

The combination of its central policy to “leave no-one behind”, its focus on creating a robust, federated, and scalable backend to “provide robust plumbing that services many fountains”, and its continuous global outreach to co-develop its functionality, has resulted in the basic RDF usage patterns from ODIS (Ocean Data and Information System) becoming the de facto underpinning for new projects in the marine research domain. We mostly applaud its technical simplicity and elegance: (i) by focusing on the simplest thing that could possibly work globally: i.e., basic Schema.org usage in JSON-LD (JavaScript Object Notation for Linked Data); and (ii) narrowing the core task to publishing only. This in stark contrast with the frequent technology-driven approach to RDF publishing, which can showcase impressive technologies and capacities, but adds unnecessary components that, more often than not, increase the cost of publishing and sometimes even hinders semantic readability. Explicitly:

You don't need to set up a triple-store or SPARQL (SPARQL Protocol And Query Language) endpoint to do Linked Open Data Publishing.

Just like the web didn't need working search engines upfront before creating linked HTML (HyperText Markup Language) pages.

You don't need to introduce new property paths or highbrow ontologies to share what you have to say, especially when you are aiming for a maximum understanding by an as wide as possible audience. In day-to-day usage, a property-path that navigates the information shape suffices

to get the answer. The optional type assessments are a luxury that ontologists and developers can engage in outside the executed real-time data-access.

### **Connecting dimensions through common identifiers**

Connections between independently published structured data graphs are created by their referencing shared concept URIs (Uniform Resource Identifiers). Indeed, just as cartographic overlays help to discover relations between independently measured values that share a geo-temporal coincidence, the application of knowledge graph technologies allow us to disclose pathways extended down to the level of data points and over many more dimensions, thus allowing both deeper and broader connections to be made.

In the European marine research domain, many such sets of concept URIs exist. Often directly usable in RDF: persistent identifiers that are (mostly) leading to valid triple-exposing representations. Adopting these is yet another (low cost) “no brainer” to create viable LOD records.

The BODC (British Oceanographic Data Center) NVS (NERC Vocabulary Server) vocabulary for “all things measurable”.

The European Directories maintained by SeaDataNet for Marine organisations (EDMO), Environmental Research Projects (EDMERP), and Ocean Observing Systems (EDIOS) ORCID (Open Researcher and Contributor IDentifier) for individual researchers and ROR (Resource Organisation Registry) for institutes.

The MareGraph project has the ambition to further complete this set with core elements of the taxonomic backbone – (i) WoRMS (World Register of Marine Species) and (ii) the Marine Regions. This work can interoperate with systems like ODIS, augmenting the marine LOD commons, while also innovating around it. Reporting on the work involved here discloses a number of interesting real-life lessons concerning Linked Data publishing.

### **Showcasing benefits**

After making the case for reducing costs we also want to balance even further by introducing a number of equally achievable ways that these efforts can benefit the community. First we will show a useful JavaScript shim script that augments the browser experience when end-users navigate HTML with structured data elements. FAIR-Signposting guidance makes concept URIs discoverable and retrieving their RDF descriptions allows for Wikipedia-inspired previews as well as copy-paste access to the actual concept URIs. The second example will showcase a free and open service component that generates and feeds standard pluggable UI widgets for the selection of concept URIs. These widgets are offered in the ‘web components’ standard for easy integration by developers of data-systems that need to manage these connections. Finally, we present a comprehensive Knowledge Graph Analysis Platform (K-GAP) that brings the potential of available connected Graph Data into the discovering spyglass of the common data scientist.

### **Conclusion**

In short, with a clear focus on the field of the Marine Research domain, we want to drive home the central argument of this quote: “People think RDF is a pain because it is complicated. The truth is even worse. RDF is painfully simplistic, but it allows you to work with real-world data and problems that are horribly complicated. While you can avoid RDF, it is harder to avoid complicated data and complicated computer problems.” – Dan Brickley and Libby Miller (in the foreword of the book “Validating RDF”).