Open Standards used in Oceanography Research Spatial Data Repositories in Spain

Enrique Wulff

Marine Sciences Institute of Andalusia, Spanish National Research Council (CSIC),
Campus Universitario Rio San Pedro s/n, 11510 Puerto Real (Cádiz), Spain
Email: enrique.wulff@icman.csic.es

ABSTRACT

Spatial data repositories specialised in geographic information systems (GIS) are an extension of map libraries and archives where much of the increase in use and citation for its geo-spatial data comes from making them available through Open Geospatial Consortium (OGC) web services. In this paper, resource needs and teaching perspectives of these open standards will be described and explained as they have become a domain of application in spatial data repositories and, marine data literacy. The research is based on checking ocean research data rules and contexts in Spain within the European obligations as defined by the EU directive INSPIRE. On that premise, Spanish spatial data infrastructures (SDI) are shown integrating OGC Web Services (OWS) with repositories for ocean observation data, a typical kind of big data. The study revealed that the broad European support to the big data open standards (OGC) implementation in the oceanographic community, is conducted with a model suitable for library management systems. However, Spanish participation in European ocean data spaces is limited and a likely explanation is that this question has not been discussed for about a decade. These findings strengthen the links between spatial data repositories and OGC standards, to identify requirements for interoperability work.

Keywords: Spatial data repositories; Marine data literacy; OGC Web services; GIS; SDI; Oceanography; EU Legislation

1. INTRODUCTION

A spatial data repository is a server-based metadata database and a spatial data server, which together are responsible for providing information about stored data and for the delivery of spatial data1. Spatial data management as a technique for organizing and retrieving information by positioning it in a spatial framework gives spatial data repositories peculiar properties. Special attention is required to expand our understanding of spatial data repositories as many applications adopted by libraries require the efficient management of spatial data2. Without them, today’s challenges in big data applications such as those faced by the spatial data infrastructures (SDI) cannot be solved3. The goal of this article is to check the role played by big data open standards, like OGC standards, in the collection, preparation, analysis and visualisation of Spanish spatial ocean data4-6. This has been achieved by considering that only becoming aware of what is going on in the European context, can vision and goals be developed from a view marked by the consideration of only one country7.

Spanish online cartographic resources (Table 1), face a shortage of professional literature in the field of expertise of spatial data repositories8-9. The role of the National Library of Spain (BNE) has been acknowledged as a web 2.0 leader that puts and keeps heritage information online10. But the institution established in order to take the responsibility for finding geospatial (GIS) data for Spain is the National Geographic Institute (IGN-E)11. The resources of the Spatial Data Infrastructure (SDI) of Spain (IDEE, www.idee.es), moreover, are based on the Technical Guidelines for Network Services of the INSPIRE directive (European Commission)12 and the specifications of the Open Geospatial Consortium (OGC).

Of the resources cited in Table 1, the majority are Spatial Data Infrastructures (SDI-IDEE); this paper describes and explains the OGC web services that manage them. Additionally, we have used web crawling engines for OGC service discovery and search engines for further test INSPIRE-compliant published OGC web services (OWS)13-15,16.

2. LITERATURE REVIEW

The major directions to implement a spatial data repository are apparent from the current debate17 on the new roles of librarians18, and can make reference to typical problems at the map library19. Ideally, librarians must provide their users with solutions to those technical difficulties that have been solved in the GIS (Geographic Information System) community, like scales, longitude, latitude, areas of interest with no-well defined boundaries (eg, Mediterranean Sea)20-21. This need recognises the potential that librarians can offer and hence the need to develop their skills in ocean data management22-23 by associating the EU INSPIRE directive coverages with OGC/ISO coverage standards24-26,27.
It must be taken into account that around 60% of the data managed by organisations are geographic data\(^\text{28}\), and that Spain is a regular contributor to the statistics of OWS found by crawlers\(^\text{29}\). Therefore, Spanish digital repositories should be recognised as an important vehicle to assess the impact of this open spatial information\(^\text{30}\). Libraries are concerned with the value of spatial information because the role of Spatial Data Infrastructures (SDI) play in the Digital Government promotes access to public and open data\(^\text{31}\). And the role of GIS librarians profits from their long experience and well-established know-how in the field of metrics information retrieval\(^\text{32}\).

There is a gap in the domain covered by the paper, because Open Geospatial Consortium (OGC) standards for the ocean regions spatial data repositories are scarcely discussed\(^\text{33}\). This article searches to fill the gap through the study of the GIS data produced in Spain on marine and ocean regions in European seas; it searches to determine in what way they are downloadable and shareable via OGC compliant web services.

### 3. SCOPE AND AIM OF THIS STUDY

Our analysis contributes to a more accurate assessment of the value of ocean research data\(^\text{34-35}\), and to the literature developing the research lifecycle model for library services. The aim of this paper is to document for library purpose, developments in OGC Spanish oceanographic services as a way of exploring Spatial Data Infrastructures (SDI). And that, following the paradigm of the library that separates the resources from their metadata, thus considering OGC services and ISO metadata as a corollary of SDIs.

### 4. METHODOLOGY

This study used data sets retrieved in May, 2020, from a list of online cartographic resources with an important number of spatial data repositories (the Spatial Data Infrastructures (SDI)) (Table 1). It should be noted that the list was established by the map library of the University of Valencia (Spain), a specialised library in maps that depends organically on the Humanities Library. The data covered the Spanish ocean OGC Web Services (OWS) available in these spatial data repositories, with particular attention on those also present in the directory of the OWS of Spain elaborated by the National Geographic Institute (IGN-E). Additionally, the official INSPIRE Geoportal (CODSI) was used in search for how

<table>
<thead>
<tr>
<th>Spatial data repositories</th>
<th>Spatial data repositories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spanish National Centre of Geographical Information</td>
<td>National Library of Spain. Spanish Digital Library</td>
</tr>
<tr>
<td>DARA. Documents and Archives of Aragon</td>
<td>Virtual Library of Defence</td>
</tr>
<tr>
<td>ideAge. Spatial Data System of the General Administration of the State</td>
<td>Virtual Library of Bibliographical Heritage</td>
</tr>
<tr>
<td>SIGNA. National Geographic Information System</td>
<td>Spanish Royal Academy of History (Madrid). Digital library</td>
</tr>
<tr>
<td>SIGPAC. Geographical Information of Agricultural Plots System</td>
<td>PARES. Website for Spanish Archives</td>
</tr>
<tr>
<td>IDEARAGON. Infrastructure of spatial data of Aragon</td>
<td>General Directorate for Cadastre. Map Servers</td>
</tr>
<tr>
<td>IDEC. Infrastructure of spatial data of Catalonia</td>
<td>Hispana. Spanish aggregator for Europeana</td>
</tr>
<tr>
<td>IDEIB. Infrastructures of spatial data of the Balearic islands</td>
<td>Cartography Institute of Catalonia</td>
</tr>
<tr>
<td>Infrastructure of spatial data of Andalusia</td>
<td>Spanish National Cartography Institute</td>
</tr>
<tr>
<td>Infrastructure of spatial data of Castilla-La Mancha</td>
<td>Spanish Technic and Geomining Institute</td>
</tr>
<tr>
<td>Infrastructure of spatial data of Spain</td>
<td>Ministry of Agriculture and Fishing, Food and Environment, Cartography and GIS</td>
</tr>
<tr>
<td>SIOSE. Information System on Occupation of the Ground in Spain</td>
<td>Ministry of Development</td>
</tr>
</tbody>
</table>

![Figure 1. OGC services summary view.](image)
many spatial data repositories adopted the OGC standards in compliance with the legal European environment (INSPIRE). After IGN-E, the total amount of different websites from where official geospatial data could be downloaded was 110; and CODSI reported 245 OWS and 222 datasets.

A comparison of results using the different discovery methods (crawling, portals, search engines) has been considered an interesting methodology. The analysis shed lights on how the implementation of OWS means the opportunity for the library to associate OGC/ISO standards to the electronic object and its metadata.

5. RESULTS AND DISCUSSION

This article discusses the implementation and development of OGC standards for Spanish oceanography. Viewed through this prism, the objective is to analyse the spatial data repositories and to encourage further involvement in marine data literacy programs. Accordingly, some technical solutions are proposed to fill the gaps in these local repositories for ocean observational data. This is done with a view on the international framework of the European law that guarantees the implications of this study for other countries.

5.1 OGC in Marine Use for Spain

Not many users of IDEE (the Spanish NSDI, Geographic Information National Infrastructure) use the infrastructure for spatial information in Europe (INSPIRE). Whereas IDEE involves over 10,492 resources, INSPIRE in Spain amounts to over 465 resources. Thus, much more OGC services are available, but not all are INSPIRE compliant services. As for the INSPIRE data that can be downloaded, in Spain only 20 % uses a Creative Commons license or any other type license. Implementation costs of a single OGC web service can be estimated in more or less 4,400 € and its maintenance in 550 €/year[3]. The reasons behind these trends are related with a lack of licensed data and with the costs (in the range of 13,000€ to 100,000€ for a geoportal, and of 37,000€ to 270,000€ for a SDI node).

5.1.1 OGC Web Services

Although the most common protocol for delivering data over the internet is the thematic real-time environmental distributed data services (THREDDS), big data providers are implementing OGC standards which facilitate the availability of a very significant amount of data. Core OGC standards are the catalogue service web (CSW), which will enable access to metadata resources, web map services (WMS), which allow requests for maps across the web and web feature services (WFS), which allow requests for geographic features across the web. Some of the other OGC standards are web coverage services (WCS), which allow requests for gridded data across the web, web map tile service (WMTS), to dynamically produce and supply maps and, web processing service (WPS) to perform online spatial analysis. These defined OGC standards are classified in compliance with the INSPIRE monitoring guidelines into one of the following categories (Fig. 1): Discovery services; view services; download services; transformation services.

The scheme is based upon the idea of making data discovery at first instance, and then proceed with the visual screening to highlight trends in the data caused. To support delivery of these data to users via dedicated download services or to put emphasis on spatial analysis to facilitate advice through transformation services is proposed as a final, dual approach.

5.1.2 Coverage of Ocean OGC Services

Coverage of OGC web services is focused on the use of different discovery methods (crawling, portals, search engines). A web crawler is a computer program that uses a list of web pages as seeds, identifies in these seeds descriptors and computes them into the right indexing tools to be subsequently provided to users as to make tiles maps. Geoportal servers are open source and free technologies that can be accessed to discover OGC services. Spatial search engines enable the discovery of both online geoprocessing services and online geodata sources. The tools that assist in discovering ocean OWS include web directories, web crawlers, geoportal servers, and search engines.

The question about how to search for resources has been posed in support of an active website or listing (a web directory) with links to all the different OGC web services across the web (for example, ref. 15, p. 910). Of course, there are administrative issues around who can afford with the creation and maintenance of a comprehensive catalog of OGC services. Consequently, there is no exhaustive list of all OGC Web Services (OWS), to automatically locate distributed GIS resources. In spite of that, Skylab Mobilesystem uses a form of Web crawling that covers WMS servers; the number of WMSs found is 994 (http://www. skylab-mobilesystems.com/en/wms_serverlist.html). There are two other WMS crawlers that depend on the Google application programming interfaces (APIs): refractions research (RR) and the Geographic Information Database™ (GIDB). But not all their functionalities seem to be there. While the OWS crawler developed by the NSF-funded PolarHub project, as of 2017 appears to be discontinued.

Listing available online WMS services in the oceanographic domain makes it possible to identify 15 WMS Servers from the marine data literacy area (marinedataliteracy. org/hdf_nc/wms/wms.htm#MDL WMS). Incidentally, it exists a WMS list from MyOcean, the European Commission project now continued by the European Union’s Earth Observation Programme Copernicus.

To get a list of map services that expose OGC based interfaces (like WMS or, WFS), just for oceanic and marine data, there are several online portals. For example, the NOAA’s Web Mapping Portal (nowcoast.noaa.gov), the NASA Earth Observations Web Mapping Service (neo sci.gsfc.nasa. gov/about/wms.php), and the Integrated Marine Observing System (IMOS) (catalogue imos.aodn.org.au/geonetwork/ srv/eng/main.home), But these sites don’t have a taxonomy/ classification originally intended for the discovery of OGC Web Services (OWS). This is also the case of the Spanish national online catalogue of open data (datos.gob.es). So therefore, searching for OWS services in those catalogues is without guarantee.
For the moment, the best engine specialised in the search for web services implementing standards published by OGC (Open Geospatial Consortium) is Spatineo (spatineo.com). Also Geoseer (www.geoseer.net) is a search engine for OGC-compliant (WMS, WFS, WVS, WMTS) data services, where the sources used as seeds are CKAN (open source data) and CSW portals. Table 2 shows a comparison between the Spatineo Crawler and the Geoseer spatial data engine, concerning ocean data services in Spain.

Table 2. Performance comparison of WMS/WFS discovery methods

<table>
<thead>
<tr>
<th></th>
<th>Spatineo</th>
<th>Geoseer</th>
</tr>
</thead>
<tbody>
<tr>
<td>WMS Services</td>
<td>817</td>
<td>13</td>
</tr>
<tr>
<td>WMTS Services</td>
<td>18</td>
<td>nil</td>
</tr>
<tr>
<td>WFS Services</td>
<td>34</td>
<td>6</td>
</tr>
</tbody>
</table>

Figure 2. Spain as a data provider in the world.

Figure 3. Spain by percentage of spatial data services availability in Europe. In March 2020, the Spanish Institute of Oceanography (IEO) continues to rank as one of the top place to work with spatial data (see on the right side column). (Copyright Spatineo Inc.)

5.2 Spanish Ocean OGC Web Services

OGC Web Services (OWS) services in Spain are published and shared on the base of cooperation and agreement among different actors, national, regional and local governments, universities, companies, and also individual citizens. These services for data access, data display and data processing available are showed in the directory of the OGC Web Services (OWS) of Spain (www.idee.es/en_GB/web/guest/directorio-de- servicing). This directory was elaborated by the Spanish NSDI (IDEE), whose geoportal was opened on December 2003, funded by using the budget of the National Geographic Institute (IGN-E). By 2018, the Spanish NSDI was composed of 90 nodes, but only less than 30% conducted open geospatial initiatives; a total number of 2200 WMS OGC services are available from Spain (Fig. 2).

Indeed, for several years (between 2010 and 2016) there has been a defined decrease in the number of reported geospatial datasets from more than 5000 to less than 300 datasets. The explanation of this fact is not particularly clear because there is no Spanish representative in the OGC Meteorology and Oceanography Domain Working Group (OGCMetOceanDWG). And that, in spite of the strong international support for “free and open data exchange”, that has been used with considerable success in the implementation of OGC Web Services (OWS).

There is a need for change in the management of long-term ocean data whilst complying with the requirements of the EU INSPIRE directive. Indeed, focusing on geospatial data and its reuse, in a way, makes it feasible to choose the library
as a combined data and information archival and publication platform. And that because library-oriented technical products and solutions (desktop/server software, libraries, plugins, online services) have lots of suggests for activities which help people to share INSPIRE data, metadata and services. As a matter of course, the employ of digital object identifiers (DOI) to implant the OGC geo-processing services, the use of portable systems to manage the onboard data, its applicability to mobile ad hoc networks are all well-defined chapters inside the annotated bibliographies available in recent years from the information and reference services.

The complexity and variety of contexts that have emerged in the national-states of the European Union for the sake of public authority and implementation of the INSPIRE directive are shown in Fig. 3.

Nevertheless, it must be taken into account that in Spain much of the metadata are not INSPIRE-compliant. And that, as EU Member State (MS), Spain is mostly concerned with the sustainability of the already available Dublin Core (DC) metadata elements (cdr.eionet.europa.eu/es/eu/inspire/monitoring/enxnsdiw/). Otherwise, the working language is Spanish, although the IGN-Spain (the Spanish National Geographic Institute) has created an INSPIRE discovery service in English (contenido.ign.es/cs/inspire/srv/eng/main.home). The INSPIRE Geoportal (inspire-geoportal.ec.europa.eu), has also detailed the number of INSPIRE metadata records (81) available from Spain.

So, when considering the two major oceanographic data spaces in Europe (EMODnet, Marine Copernicus), the results have been scarce as measured against the question of the implementation of OGC Services in Spain. Table 3 summarises the specific marine institutions providing OGC datasets and coming from Spain in two examined major data spaces.

Datasets as shown in Table 3, are made available through WMS services. They enable access to public geographic layers by using Geographical Information Systems (GIS) (ArcGIS, QGis) and through web portals that provide access to remote layers. In EMODnet Data Portal, a search was made over 578 data sets, services and maps; while in Marine Copernicus, the search was made over 671 data sets, services and maps. CDI/Thredds interoperability in the SeaDataNet (a third relevant EU ocean data space) framework is based on a OGC CS-W General Catalogue Interface, but it is not easily attainable and was discarded by this study. Great differences are observed in IEO vs CSIC and PE results (Table 3), given that IEO has a Thredds data server (TDS) that provide more metadata and data access by using OGC WMS and WCS protocols than those TDS serving CSIC and PE.

### Table 3. Datasets made available as OGC web services (OWS) by Spanish Oceanographic Institutions to serve major EU data spaces

<table>
<thead>
<tr>
<th></th>
<th>IEO</th>
<th>CSIC</th>
<th>Harbour authority (PE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMODnet</td>
<td>57</td>
<td>7</td>
<td>nil</td>
</tr>
<tr>
<td>Marine Copernicus</td>
<td>nil</td>
<td>nil</td>
<td>7</td>
</tr>
</tbody>
</table>

6. **Conclusions and Future Work**

Collect, prepare, analyse, and visualise Spanish spatial ocean data through big data open standards, like OGC standards, have based the foregoing discussion and, the following conclusions are derived:

- This article showed spatial data repositories progress in the oceanographic community is associated with big data open standards (OGC). There is broad European support to solve significant technical issues in the OGC standards implementation, in a style that have much in common with library management tools.

- European legal obligations with technical aspects (metadata compliance with INSPIRE directive) resulted in a limited participation of the Spanish oceanographic data in the European data spaces. Spanish spatial data repositories have not discussed this question for about a decade.

- Spanish authorities preferred the development of interoperability platforms based on OGC standards from geopORTals and as part of regional spatial data infrastructures (SDI). These geospatial services and data sets can also be found in Spanish spatial data repositories, by using crawlers and search engines.

Future work include the incorporation of other OGC web service interfaces in the spatial data repositories, the increase of demand of training courses on big data management and OGC web services, and the improvement of skills related to managing geographic information systems (GIS) from mobile technology ocean mapping applications in libraries.

### References


doi: 10.1371/journal.pone.0092590.


ACKNOWLEDGEMENTS

The author acknowledges anonymous reviewers for their careful reading of the initial version and their insightful comments and helpful suggestions. A previous version of this article was a poster presented in September 2016 at the INSPIRE Conference in Barcelona.

CONTRIBUTOR

Mr Enrique Wulff, M.Sc. (Biomedicine) is a Librarian serving for the past 10 years at the Spanish Council for Scientific Research (CSIC) in Cádiz (Spain). The key focus area are data repositories, marine information and european law. In this study, he proposed the idea, applied the methods and he wrote the manuscript.