# State of Play: Spain and Portugal

SDI services' state of play in autumn 2010

**Lopez-Pellicer, Francisco J.; Florcyk, Aneta J.; Béjar, Rubén; Nogueras-Iso, Javier; Zarazaga-Soria, F. Javier; Muro-Medrano; Pedro R.**

## Abstract

This paper presents a technical overview of the State of Play in autumn 2010 of OGC Web services across Spain and Portugal. This work provides descriptive and statistical analysis about the availability of public geospatial Web services. The information contained in this paper might provide to stakeholders an independent and accurate insight into the current level of implementation of public OGC and INSPIRE network services in both Spain and Portugal. When possible, the information is compared with data obtained in winter 2010. The purpose of this comparison is providing to stakeholders information about trends in the level of deployment of SDI in the Iberian Peninsula. All the data is provided by an advanced crawler that is able to discover access points to public geospatial Web services across the Web. The crawler is able to perform a deep search for services in general purpose search engines, form based applications and standard OGC catalogues.

**KEYWORDS**

OGC services, discovery, SDI, state of play, Spain, Portugal

## 1. Introduction

The number of SDI has been growing for years. Public administrations at all levels are implementing geoportals that offers OGC web services. The increasing numbers of providers and web services difficult the discovery of the appropriate service using the traditional service registries and directories. However, the process of discovering Web service access points is no longer attached to service registries or directories. Recently, search engines, such as Google and Yahoo, have become surprisingly a new major source for automated Web service discovery.

This paper presents an analysis of the number and distribution of OGC web services in Portugal and Spain. The services analyzed have been found using generic search engines (Google, Bing, Yahoo!) complemented with services found through spatial data infrastructure services: geoportals and catalogues. The number of publicly available services in the Web and the amount of data served by them are suitable indicators of the degree of implementation of OGC services in the Europe-INSPIRE Web. Additionally, in the INSPIRE context, stakeholders may require these kind of indirect indicators to choose among possible service implementations, and foster changes in OGC specifications.

This study is focussed in OWS standards that have been indicated as possible implementations of the INSPIRE compliant services:

- Web Map Service (WMS) (and possibly the recent Web Map Tile Service, WMTS) for implementing view services.
- Web Feature Service (WFS) and Web Coverage Service (WCS) for implementing download services.
- Web Processing Service (WMTS) for implementing invoke spatial process services and transformation services.
- Catalogue Service for the Web (CSW) for implementing discovery services.

The information contained in the survey may serve to INSPIRE stakeholders in Spain and Portugal as a complementary technical annex to the official State of Play reports for 2009 that will be available before the end of 2010.

This paper is structured as follows. First, we present related works about the discovery of web services in general and OGC services in particular. Then, the methodology of this survey is described. Next, we present the results. Finally, after a brief discussion, the conclusions are presented.

## 2. Related work

The investigation of Al-Masri et al. [1] found that all the web services that were discovered, 72% were found only in Web search engines, and detected a significant growth rate of services found in search engines. The OGC world has paid little attention to the automated discovery of OWS access points in search engines. However, some recent works shows a growing interest in the automated discovery of OGC web services [2] [3].

Effective automated collection OWS information from search engines goes beyond simple crawling. Our approach uses techniques similar to those employed in deep Web crawlers, such as the described in Ntoulas et al. [4], to increase the number of OWS service references found in search engines. After being crawled, the references are visited, identified, validated and analyzed to collect any additional information.

Nevertheless, basic heuristics derived from the signatures of Web Map Services methods has been exploited to discover map servers by developers. For example, Refractions Research applied this kind of strategy for identifying 695 WMS [5], and Sample et al. [6] described a similar approach to discover 761 WMS.

## 3. Methodology

The search of OGC services is done by crawling main search engine results, web pages from geoportals and metadata records when available. The web crawling allows minimizing the number of services that remain hidden by the difficulty to find them. The crawling and analysis of OGC services has the following stages:

- **Crawl the Web.** Identify candidate pages and web service querying search engines. We use an OGC focused crawler able to discover OGC endpoints. This system is similar to the described by [3] but can also query form based systems and query search engines.
- **Fetch and store**. The URL identified by the crawler as candidates are fetched and stored in a long term repository as raw data. This is the base for future retrospective analysis of data.
- **Extract metadata**. The raw data is processed for the discovery of valid OGC documents and possible new URL candidates. The metadata of the capabilities is extracted using a process dependent of their standard specification.
- **Map metadata**. The metadata extracted from each capability is mapped to a RDF data model that homogenizes the different OGC standards. The resulting data is stored in a RDF repository. As the time of writing, the RDF repository stores 3.526.574 statements.
- **Reporting**. The generation of reports is automated. This process is similar to other report generation processes against relational database but this is against a RDF repository.

## 4. Results

The results of presented bellow comes from the analysis of web services discovered between September 27, 2010 and September 30, 2010 by querying the main search engines and crawling the available data in the main geoportals of Portugal and Spain. The search was restricted to the WMS, WFS, WCS, WPS and CSW services. WMTS services, although they belong to a different standard, were considered as WMS services as both are web map services. Other services were detected but no included in this report, such as Sensor Observation services.

### 4.1 Services discovered and its distribution

This study has found 1390 Web services in Spain and 51 Web services in Portugal (see Table 1). That is, there are at least 1441 public Web services in Portugal and Spain that can be found following links from web pages in geoportals, metadata records in catalogue applications and response pages in search engines. The most common service interface found is the WMS, followed by the WFS. These results suggest that the current efforts in public network services in Portugal and Spain are oriented to provide maps for their use in geoviewers and mashups.

| | *Portugal* | *Spain* | *Total* |
|---|---|---|---|
| **WMS** | 34 | 1140 | 1174 |
| **WFS** | 12 | 209 | 221 |
| **WCS** | 1 | 24 | 25 |
| **WPS** | 3 | 3 | 6 |
| **CSW** | 1 | 14 | 15 |
| **Total** | 51 | 1390 | 1441 |

**Table 1. Services found in Portugal and Spain.**

### 4.2 Service specifications

Table 2 shows for each service interface standard the number of web servers per version. The data collected shows that the version 1.3.0 of the WMS standard has not been adopted widely in Portugal and Spain. This fact might suggest that public administration is not investing in the most recent mapping software, and it is more interested in publishing data using mature versions of OGC services. This is not incompatible with the preference for the most recent version of the WFS for the access of data as the version 1.0.0 implementations are designed for GML 2.1.2 while 1.1.0 are designed for GML 3.1.1. The standard of choice for WCS is 1.0.0. The data also shows that there are pre-standard WPS operating in the peninsula. Finally, the last version of the CSW standard is the most common, simplifying the interoperability among CSW catalogues.

| *Services* | | *Services* | | *Services* | | *Services* | | *Services* | |
|---|---|---|---|---|---|---|---|---|---|
| *WMS* | *1174* | *WFS* | *221* | *WCS* | *25* | *WPS* | *6* | *CSW* | *15* |
| *1.3.0* | 143 | *1.1.0* | 202 | *1.1.1* | 4 | *1.0.0* | 3 | *2.0.2* | 9 |
| *1.1.1* | 1011 | *1.0.0* | 17 | *1.1.0* | 1 | *0.4.0* | 3 | *2.0.1* | 2 |
| *1.1.0* | 13 | | | *1.0.0* | 20 | | | *2.0.0* | 2 |
| *1.0.0* | 5 | | | | | | | *1.0.0* | 2 |
| *WMTS* | 2 | | | | | | | | |

**Table 2. Specification of services found.**

### 4.3 Issues found in capabilities that affects service discovery

There are some of issues that affect the discovery and use of the services published by SDI geoportals, organizations, academia and map hosting services (Table 3). As [7] acknowledges, so far, little attention has been paid to fill in adequately the capabilities documents of OGC web services, one of the best sources of service metadata. These are:

- In many capabilities, the keywords are only used for describing the service. The layers, features and coverages seldom have keywords. The absence of keywords difficult automatic processes of indexing and classification based on keywords.
- A half of the services have no rights or fees declaration. When the right or fees are present, their format varies and their content can be considered sometimes ambiguous.
- A quarter of the services have no narrative description of their content or purpose at service or at content level. This difficult the discovery of the service using a free text

search.

- 17% of the services has their endpoint address mangled. The most common case is the presence of URLs pointing to an intranet address. This is a serious issue as they might confuse client applications.

| *Issues* | *Capabilities* | |
|---|---|---|
| *Missing keywords* | 832 | 57% |
| *Implicit rights* | 810 | 56% |
| *Implicit fees* | 793 | 55% |
| *Missing abstracts* | 396 | 27% |
| *Endpoint address mangled* | 249 | 17% |
| *No contact point* | 109 | 8% |

**Table 3. Issues found in capabilities.**

Table 4 shows the most common keywords found in capabilities. Many of the terms used for describing a service are redundant (wms, servidor de mapas en red, wcs, wfs), names their extension (andalucia, canarias, barcelona, huelva) or identifies the service provider (idée, igme, ign, grafcan). We can conclude that the textual description of web services cannot be used for the discovery of the best service for a purpose. The narrative description is sparse and often refers to the service and not to their content.

| *Keywords* | *#* | *Keywords* | *#* | *Keywords* | *#* | *Keywords* | *#* | *Keywords* | *#* |
|---|---|---|---|---|---|---|---|---|---|
| Wms | 379 | Barcelona | 42 | Islands | 31 | Wfs | 23 | Ca | 19 |
| Andalucia | 135 | Ign | 42 | Granada | 31 | Catastro | 23 | 2007 | 18 |
| servidor de mapas en red | 131 | Grafcan | 37 | Cadiz | 30 | Map | 22 | Aerial | 18 |
| Cartografia | 78 | Huelva | 37 | Malaga | 30 | Geology | 21 | tenerife | 18 |
| medio ambiente | 73 | Cabildo | 35 | Almeria | 30 | sit_chg | 20 | Suelos | 18 |
| Ortofotos | 63 | Mapas | 34 | Jaen | 29 | Marm | 20 | ordenacion | 17 |
| Arcims | 63 | Geologia | 34 | Cordoba | 28 | Color | 20 | topographic | 16 |
| Idée | 49 | Espana | 33 | Mmamb | 27 | Mapa | 19 | extremadura | 15 |
| Canarias | 46 | Sevilla | 33 | Wcs | 27 | Palma | 19 | 10000 | 14 |
| Igme | 43 | Canary | 31 | Spain | 24 | Ideib | 19 | 2005 | 13 |

**Table 4. Common kewyords.**

### 4.4 Available data

Nowadays, WMS users can choose between 20521 named layers of information about Spain and Portugal (see Table 5). This information is organized in 2928 categories, that is, layers that cannot be requested in a GetMap request. T If we analyze the Pareto diagram of offerings per service (Figure 1). We can conclude that this is distribution is a good example of the 80/20 rule: 80% of the layers, coverages and feature types are served by the 20% of web services. That is, a tiny number of services, and possibly service providers, account for almost all of the layers available in Portugal and Spain. Some of these services correspond to academia provided services which are discussed in the next section. If we don't consider the distribution of layers, the average number of layers offered by each WMS is 25 layers. If we exclude these 20% considering them as outliers, the average number of layers offered by each WMS falls to 5 layers.

On the opposite side, the number of coverages and feature types available is less than the services detected. This is due to that this study has detected some WCS and WFS whose capabilities do not contain any information about the feature types and coverages offered by the service.

| | *Portugal* | *Spain* | *Total* |
|---|---|---|---|
| **Layers** | 361 | 23088 | 23449 |
| *Named layers* | 295 | 20226 | 20521 |
| *Categories* | 66 | 2862 | 2928 |
| **Coverages** | 1 | 15 | 16 |
| **Feature types** | 11 | 209 | 220 |

**Table 5. Layers, coverages and feature types available in Portugal and Spain.**

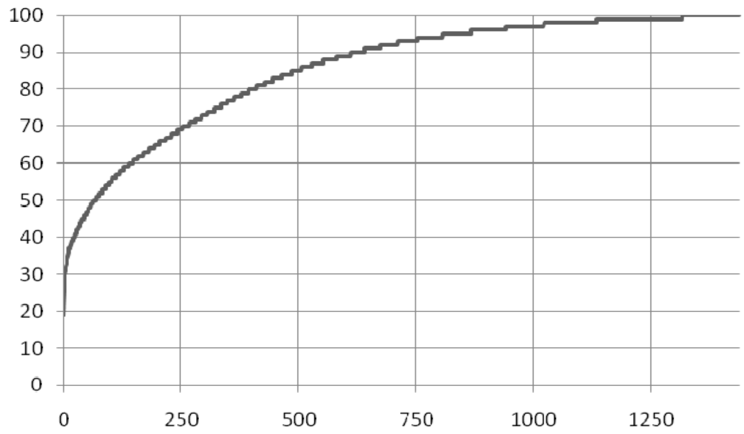


**Figure 1. Pareto diagram of number of layers, feature types and coverages served by each service.**

### 4.5 Service providers

We have identified 6 kinds of service providers (see Table 6).

- **National SDI** geoportals, represented by the geoportal of the IDEE and the SNIG.
- **Regional SDI** geoportal, for example the autonomous communities of La Rioja and Andalucia in Spain.
- **Provincial SDI** geoportal, such as TeIDE, the SDI of Tenerife.
- **Academia** provided services. The most representative in Spain are the initiatives of universities of Barcelona (UPC, UAB) and Girona (UDG) in Spain. The most representative in Portugal is the Universidade do Porto (not included in Table 6).
- **National organization** geoportal, usually a data provider. For example, IGME in Spain provides facilities for displaying, accessing and downloading geophysical data.
- **Map hosting** services for public administrations provided by public, such as Eprinsa, and private companies, such as Geodata. They may be considered as part of the SDI infrastructure of the public administration that uses their services.

Regional SDI and map hosting services for public administrators are the most common, and possibly the most important, agents in Spain. The prevalence of regional SDI in Spain is related with the decentralized administrative structure. The data of Portugal suggest that the most important agent is the national SDI.

| *Web Server* | *Services* | | *Scope* |
|---|---|---|---|
| ogc.larioja.org | 337 | 23.39% | Regional SDI |

| | | | |
|---|---|---|---|
| www.juntadeandalucia.es | 126 | 8.74% | Regional SDI |
| mapserver.eprinsa.es | 78 | 5.41% | Map hosting |
| oslo.geodata.es | 67 | 4.65% | Map hosting |
| atlastenerife.es | 49 | 3.40% | Provincial SDI |
| www.idee.es | 37 | 2.57% | National SDI |
| mapas.igeo.pt | 32 | 2.22% | National SDI |
| cpsv-sig2.upc.es | 31 | 2.15% | Academia |
| mapas.igme.es | 28 | 1.94% | National Org |
| www.ideandalucia.es | 28 | 1.94% | Regional SDI |
| idecan2.grafcan.es | 26 | 1.80% | Regional SDI |
| www.mapasdelapalma.es | 26 | 1.80% | Provincial SDI |
| aligit10.uab.es | 24 | 1.67% | Academia |
| ideib.caib.es | 20 | 1.39% | Regional SDI |
| 161.116.68.1 | 18 | 1.25% | Academia |
| feldespato.igme.es | 16 | 1.11% | National Org |
| www.opengis.uab.es | 16 | 1.11% | Academia |
| wms.sigte.udg.edu | 15 | 1.04% | Academia |
| ideg.xunta.es | 14 | 0.97% | Regional SDI |
| idecan1.grafcan.es | 14 | 0.97% | Regional SDI |

**Table 6. Top 20 web servers (70% of services under analysis) and their scope**

### 4.6 Regional distribution

Table 7 presents the regional distribution of the web services and the layers that they publish. Two large autonomous communities (Andalucia and Catalonia) and a single province autonomous community (La Rioja) lead the development of web services in Spain. The information is derived from the location of the server, which can be different from the location of the provider, and it is dependant of the precision of the IP geocoder. A good example is the data about Castilla-La Mancha, whose servers have been geolocated in Madrid. The data available for Portugal is not sufficient to be conclusive.

| *Web services* | | *Web services* | | *Named layers* | | *Named layers* | |
|---|---|---|---|---|---|---|---|
| **Portugal** | **51** | **Spain** | **1390** | **Portugal** | **295** | **Spain** | **20226** |
| *Aveiro* | 2 | *Andalucía* | 246 | *Aveiro* | 20 | *Andalucía* | 2134 |
| *Faro* | 32 | *Aragón* | 20 | *Faro* | 125 | *Aragón* | 347 |
| *Lisboa* | 13 | *Asturias* | 13 | *Lisboa* | 150 | *Asturias* | 54 |
| *Undetermined* | 4 | *Canarias* | 93 | | | *Canarias* | 386 |
| | | *Cantabria* | 27 | | | *Cantabria* | 127 |
| | | *Castilla y Leon* | 30 | | | *Castilla y Leon* | 282 |
| | | *Cast.-La Mancha* | 1 | | | *Cast.-La Mancha* | 2 |
| | | *Catalonia* | 215 | | | *Catalonia* | 8970 |
| | | *Com. Valenciana* | 29 | | | *Com.Valenciana* | 463 |
| | | *Extremadura* | 18 | | | *Extremadura* | 119 |
| | | *Galicia* | 24 | | | *Galicia* | 822 |
| | | *Islas Baleares* | 46 | | | *Islas Baleares* | 811 |
| | | *La Rioja* | 349 | | | *La Rioja* | 3014 |
| | | *Madrid* | 164 | | | *Madrid* | 985 |
| | | *Murcia* | 17 | | | *Murcia* | 267 |
| | | *Navarra* | 4 | | | *Navarra* | 250 |
| | | *Pais Vasco* | 94 | | | *Pais Vasco* | 1110 |
| | | | | | | *Undetermined* | 91 |

**Table 7. Distribution of web services and named layers.**

## 5. Discussion

The available State of Play reports are dated in 2007. The State of Play reports of 2009 will be published before the end of 2010. In the State of Play of Spain were explicitly identified 91 services: 20 in the IDEE, 64 in the IDEC (Catalonia) and 7 in the SITNA (Navarra). The State of Play of Portugal indentified only 8 services. The future versions of the State of Play of Spain will include updated numbers. For example, the State of Play of Spain should contain as many services as the IDEE account for Spain: 954 WMS, 218 WFS, 19 WCS, 6 WPS and 15 CSW. This account is updated by reporting new services to the IDEE, and, hence, the directory could be incomplete.

We think that an automated process for the discovery of services can provide a continuous and up-to-date stream of information about the status of network services. For example, the crawler discovered 178 services that were not notified to the IDEE directory. A crawler can also play an important role in the elaboration of State of Play documents. Using a crawler, the State of Play of network services could become a live document instead of being part of document issued each 2 years with 1 year of delay.

## 6. Conclusion

This paper has presented the state of play of network services in Portugal and Spain in autumn of 2010. With an OGC focused crawler, we have discovered 51 web services in Portugal and 1390 services in Spain. The most common services in Portugal and Spain, as expected, were WMS services. The analysis also revealed quality problems in the capabilities, such as the lack of adequate descriptions, and interoperability problems, such as URL pointing to intranet addresses. The total of layers, coverages and different feature types offered by these services is over 23500. The analysis of the data suggests that few services offer many them. Excluding these outliers, the average of different offerings (e.g. layers) per service is 5. We have also analyzed the location of services. The decentralization of the location of services in Spain is notable. The data about Portugal is not conclusive. Regional SDI and map hosting services for public administrators are the most common, and possibly the most important, agents in Spain. The regional public administrations of Catalonia, La Rioja, and Andalucia are the major service providers in Spain. The prevalence of regional SDI in Spain is related with the decentralized administrative structure. The data of Portugal suggest that the most important agent is the national SDI.

A large part of the content of this report has been generated automatically from data harvested by a web crawling engine. As this document shows, automated crawling and analysis processes can ease the elaboration of technical annexes of State of Play reports about network services, and the maintenance of large directories and catalogues of network services.

## 8. Ackowledgements

This work has been partially supported by Spanish Government by the projects TIN2009-10971 and “España Virtual” ref CENIT 2008-1030 (through the National Center of Geographic Information (CNIG) and GeoSpatiumLab), the Aragón Government (project PI075/08), GeoSpatiumLab S.L. and Zeta Amaltea S.L. The work of Aneta Jadwiga Florczyk has been partially supported by a grant (ref. AP2007-03275) from the Spanish government.

## 9. References

[1] Al-Masri, E., Mahmoud, Q.H.: Investigating web services on the world wide web, in Proceedings of the 17th international conference on World Wide Web, ACM, New York, NY, USA, pp. 795-804 (2008).

[2] Lopez-Pellicer, F.J., Béjar, R., Florczyk, A.J., Muro-Medrano, P.R., and Zarazaga-Soria, F.J.: State of Play of OGC Web Services across the Web”, INSPIRE 2010 conference, 22-25 June, Kraków, Poland (2010).

[3] Wenwen L., Chaowei Y., Chongjun Y.: An active crawler for discovering geospatial Web services and their distribution pattern - A case study of OGC Web Map Service, International Journal of Geographical Information Science, Vol. 24 Iss: 8, pp. 1127-1147 (2010).

[4] Ntoulas, A., Zerfos, P., Cho, J.: Downloading Textual Hidden Web Content by Keyword Queries, in 'Proceedings of the Fifth ACM+IEEE Joint Conference on Digital Libraries (JCDL 2005)' (2005)

[5] Refractions Research, OGC Services Survey, white paper, Refractions Research, Victoria, Canada, (2006)

[6] Sample, J.T., Ladner, R., Shulman, L., Ioup, E., Petry, F., Warner, E., Shaw, K. and McCreedy, F. (2006), “Enhancing the US Navy's GIDB Portal with Web Services”, IEEE Internet Computing, Vol. 10 Iss: 5, pp. 53-60 (2006).

[7] Capdevila, J.: Member State Report: España, 2009, (2010). Available in http://www.idee.es/SeguimientoINSPIRE/2009/100720_Report_INSPIRE_2009_Spain_v2.pdf

## 1 CONTACT

**LOPEZ-PELLICER, Francisco J.**
fjlopez@unizar.es
Universidad de Zaragoza
Departamento de Informática e Ingeniería de Sistemas

**FLORCYK , Aneta J.**
florcyk@unizar.es
Universidad de Zaragoza
Departamento de Informática e Ingeniería de Sistema

**BEJAR, Rubén**
rbejar@unizar.es
Universidad de Zaragoza
Departamento de Informática e Ingeniería de Sistemas

**NOGUERAS-ISO, Javier**
jnog@unizar.es
Universidad de Zaragoza
Departamento de Informática e Ingeniería de Sistemas

**ZARAZAGA-SORIA, Fº Javier**
javy@unizar.es
Universidad de Zaragoza
Departamento de Informática e Ingeniería de Sistemas

**MURO-MEDRANO, Pedro R.**
jnog@unizar.es
Universidad de Zaragoza
Departamento de Informática e Ingeniería de Sistemas