A Global Synopsis of OGC Web Services

J. Moules

GeoSeer - science@geoseer.net

January 15, 2019

Summary

This paper seeks to provide a global overview of Open Geospatial Consortium (OGC) Web Services (OWS) usage for Web Map Services (WMS), Web Feature Services (WFS), Web Coverage Services (WCS), and Web Map Tile Services (WMTS) as well as identify patterns in current deployments. We look at the history of OWS deployments, contemporary OWS deployments and their characterisations, and make observations about service and layer discoverability and metadata.

KEYWORDS: resource discovery, Open Geospatial Consortium, metadata, OGC web services

1. Introduction

Since the Web Map Service (WMS) 1.0.0 was released (2000) – providing a web standard for map-images – usage of Open Geospatial Consortium (OGC) Web Services (OWS) has gone from strength to strength. Today these standards include Web Feature Service (WFS) for raw vector data, Web Coverage Service (WCS) for raw raster data, and Web Map Tile Service (WMTS) for pre-rendered base-maps. Recently they have seen an explosion in usage in part as a result of being an explicit requirement of the EU-wide INSPIRE Directive (2007). This paper seeks to provide a global synopsis of OWS usage for WMS, WFS, WCS, and WMTS as well as identify patterns in current deployments, and look at the history of service deployments and the literature.

2. Contemporary Dataset

The contemporary dataset used for this research is from GeoSeer, a free-to-use OWS search engine. GeoSeer crawls ~300 catalogues, and in December 2018 had 141,376 active services in its index. While the exact details of the crawler are beyond the scope of this paper, there are a number of caveats that require highlighting. To be included a service must:

- Be publicly accessible
- Not require authentication
- Have layers
- Be active at the time of the crawl
- Respond with a valid WMS, WFS, WCS, or WMTS GetCapabilities document

A "host" is the hostname of the server (i.e. www.example.com) and may host many "services". A service can serve many layers. We use the term "layer" in this paper to inclusively cover WMS Layers, WFS FeatureTypes, WCS Coverages, and WMTS TileMatrixes.

3. Historical Numbers

There have been numerous attempts to quantify the usage of OWS. Some of these have explicitly sought to measure adoption with the first such review being by Paul Ramsey in 2004 (Reichardt, 2005). Lopez-Pellicer et al. (2011) investigated the status of OWS globally, and later in relation to EU INSPIRE (2012). In 2016, Gui et al. performed a comprehensive look at global services, their performance, and their characteristics.

A number of incidental investigations of OWS deployments have also been conducted, with the number of services found being used as a common metric for comparing the efficacy of OWS focused search engines. Of these, several have numbers that can be pulled from their papers, including GIDB (2006), Li et al.'s 2010 work, GSE (Bone et al., 2014), GeoWeb Crawler (Huang et al., 2016), FGDC (Anthony et al., 2012), and most recently PolarHub (Li, 2017). Finally, there are snapshots from commercial search efforts from MicroImages (2008), SkyLab (2013) and the contemporary Spatineo.

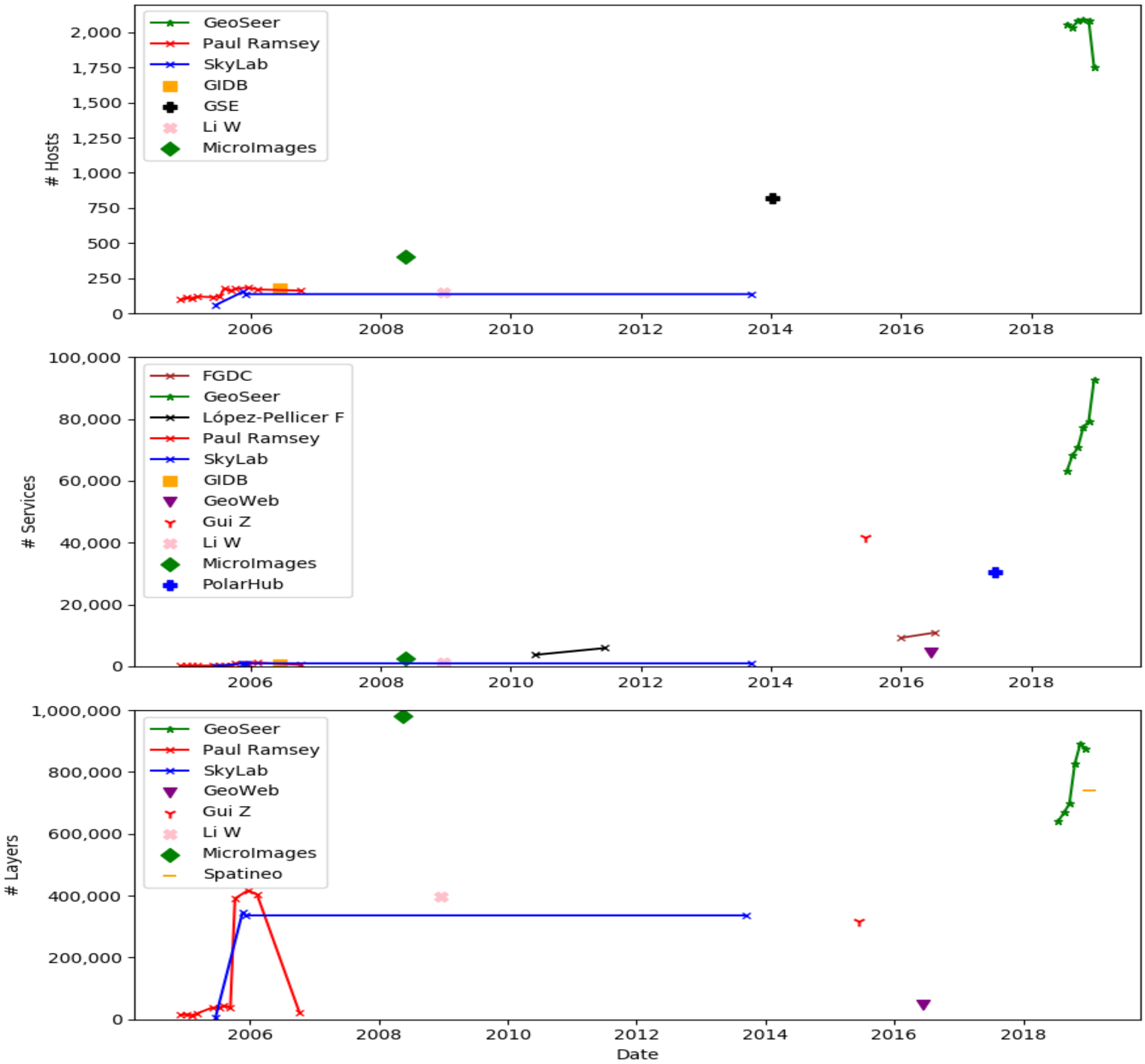

Figure 1: A timeline of WMS host, service, and layer counts.
Data amalgamated from research papers and commercial sources.
Historical Skylab data is from archive.org and duplicate services were removed. 2005-12-07 to 2009-03-12 datasets are identical, so only the first dataset of that sequence was used. The Ramsey data remains available and was also downloaded and processed.

Figure 1 shows all the counts from various sources for WMS' amalgamated into a single graph. While the many differing motivations and methodologies mean the results cannot be compared like-for-like, they do provide an indicator as to the large increase in deployment of WMS', as well as the capabilities of state-of-the-art OWS search engines. The large differences in the average (mean) numbers of layers per service, from ~400 (Ramsey, Skylab, MicroImages) to ~10 today (GeoSeer, Spatineo) are particularly notable, as are the concomitant increases from ~6 services/host to around 50 service/host.

4. Results and Discussion

4.1. Deployment patterns

The rest of this paper concentrates on the GeoSeer data which represents 88 Top Level Domains (TLDs) of which 73 are country code TLDs (ccTLD) such as ".uk". Figure 2 shows a breakdown of the top fifteen ccTLDs as sorted by layer count. The USA is under-represented here, as it largely uses 3-letter sponsored TLDs. France, Belgium, and Germany between them evidence 507,931 (53.2%) of the 954,125 layers hosted on ccTLDs. In contrast, France and Germany alone host 64,489 (68.8%) services of the 93,785 ccTLD services. For number of hosts however, the UK comes in third place with 104 hosts, despite being 15th for number of layers. These numbers clearly indicate the different strategies that countries are using in their Spatial Data Infrastructures (SDI). The UK averages (mean) 115 layers per host indicating considerable distribution across many providers. In contrast, Belgium averages 4,384 layers per host, with Ecuador (1,697), France (1,544), Brazil (1,224), and Poland (1,213) rounding off the large centralised deployments. This represents a continuation of the characterisations identified by López-Pellicer et al. (2011). Further evidence of this can be seen in the different profiles of the Figure 3 histograms.

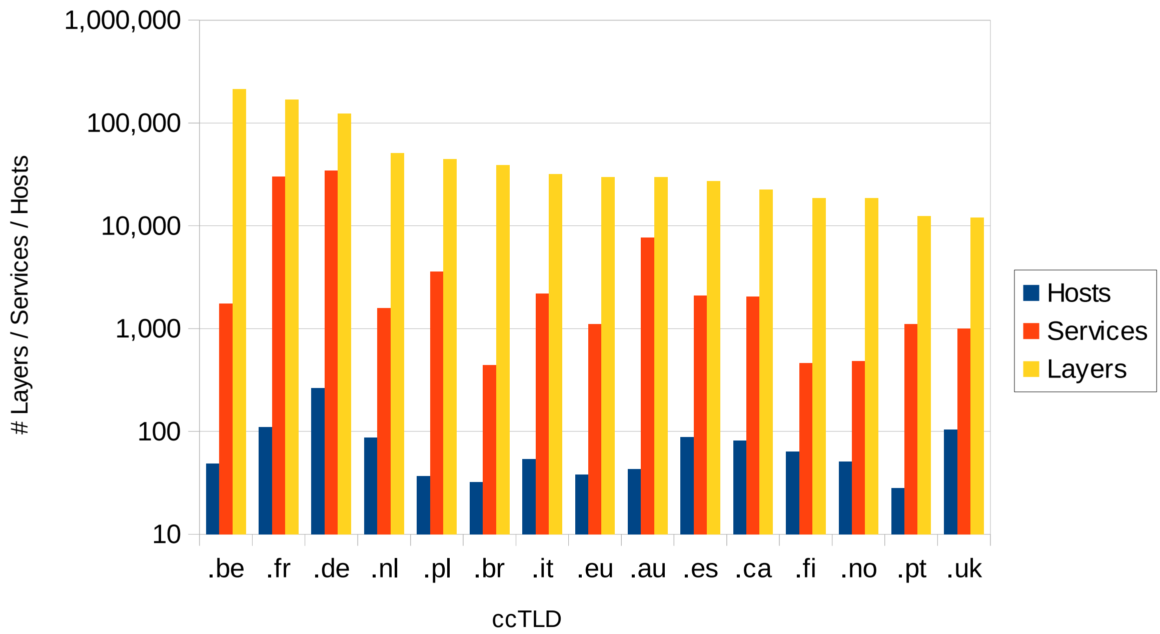

Figure 2: Number of hosts, services, and layers deployed per ccTLD across all service types, for the top 15 ccTLDs by layer-count, on a logarithmic scale.

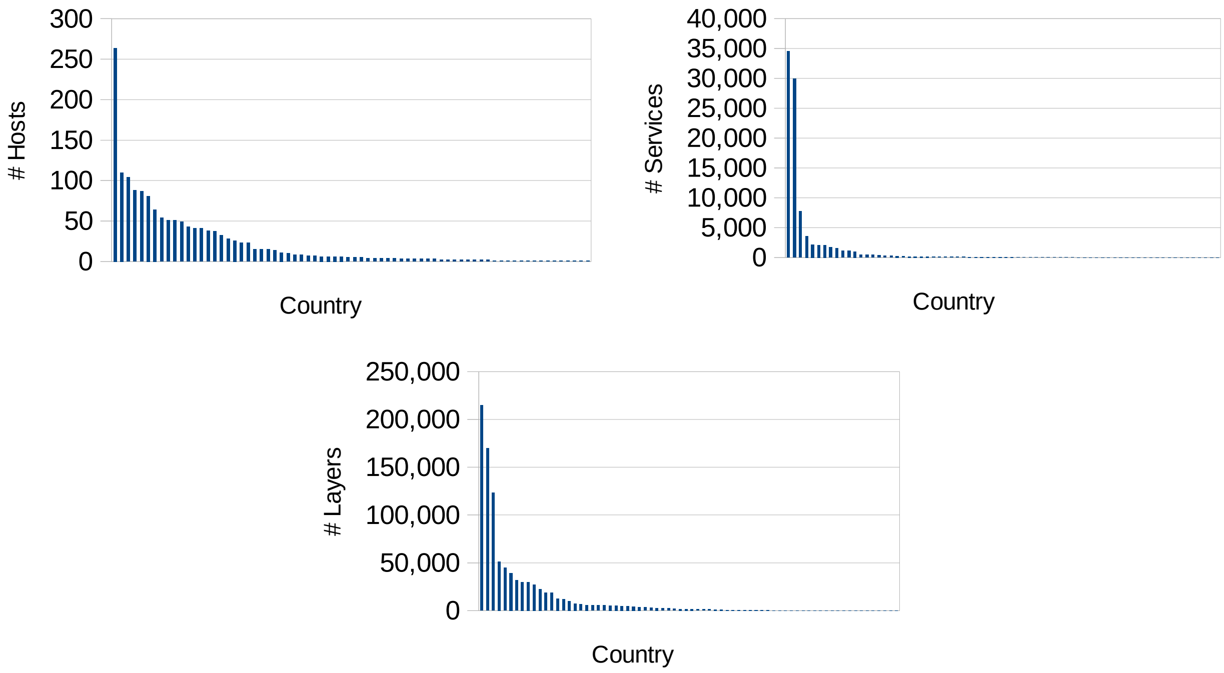

Figure 3: Histograms of number of hosts, services, and layers per ccTLD

This pattern of centralisation is also evidenced in the availability of hosts, services, and layers via unencrypted (HTTP) versus encrypted (HTTPS) protocols. Figure 4 shows that over the past 9 months of 2018 there has been a dramatic uptick in the percentage of layers available via HTTPS, going from 29.0% in March to 61.4% in December. There is a smaller increase across hosts, from 27.2% to 39.5% over the same period. Curiously, at the service level the usage of HTTPS peaked in May at 22.8%, and now rests at just 16.6%. This further highlights the large number of services and layers that just a few hosts control. This general trend of increasing encryption matches the overall web, albeit the web as a whole has a higher rate at 70% HTTPS (Mozilla, 2019).

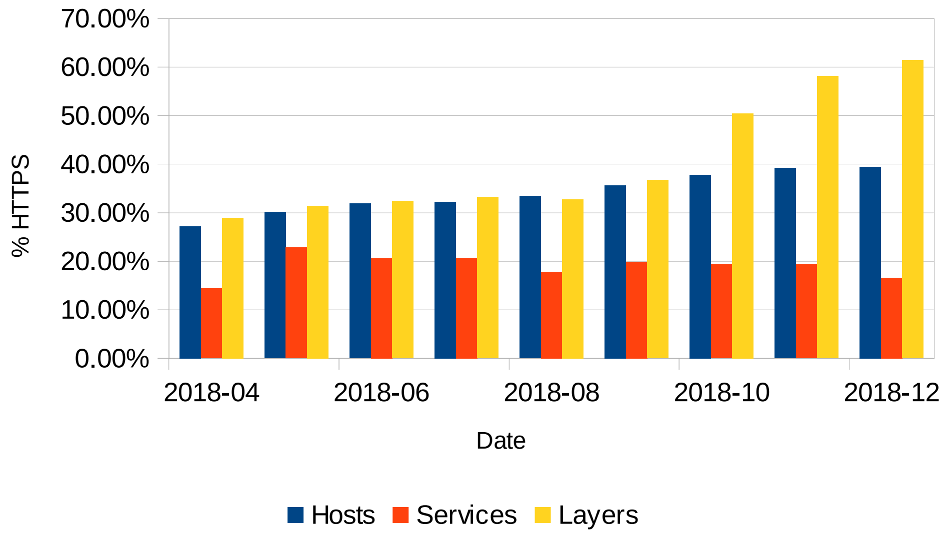

Figure 4: Percentage of hosts, services, and layers that are available via the encrypted HTTPS protocol

4.2. Discoverability and metadata

One of the issues repeatedly highlighted in the literature is discoverability, and this goes hand-in-hand with the problem of understanding what the service returns. Metadata solves these problems, with services needing meaningful abstracts and keywords, while at the layer level abstracts, keywords, and links to external metadata are all useful.

In Figure 5 we see that at the service level, most services don't have keywords set (64.66%), considerably hampering discoverability. WCS in particular is notable with only 1% of WCS services having any keywords. Abstracts were provided for most WMS and WFS services, while most WMTS and almost all WCS services have neither keywords nor abstract.

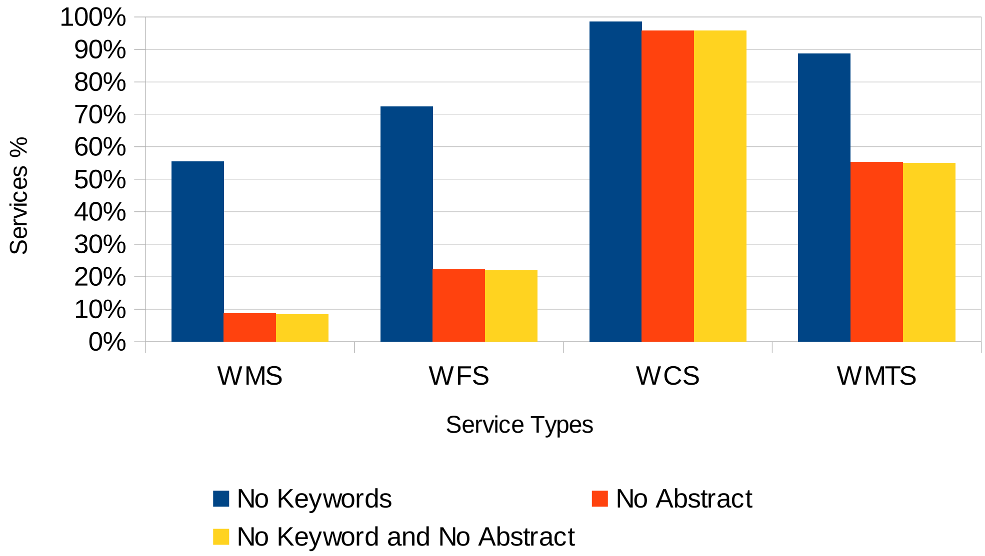

Figure 5: Absence of various types of metadata per service. Default keywords and abstracts were stripped before analysis.

Per-layer metadata tells a similar story in Figure 6. While metadata is not important for all layers – for example a county's fire station location data is unlikely to need it – such information can still improve discoverability rather than solely relying on a hopefully-meaningful layer title. For WMS, which represents the lions share of available layers, 24% do not have either a keyword or abstract, severely limiting their discoverability. WCS however tells a different story, with only 7% of layers not having an abstract, this likely follows from the fact that WCS tends to be used almost entirely for scientific data where such information is particularly important. That said, the use of the metadata_url field is extremely low across all services, missing from 77% of all layers.

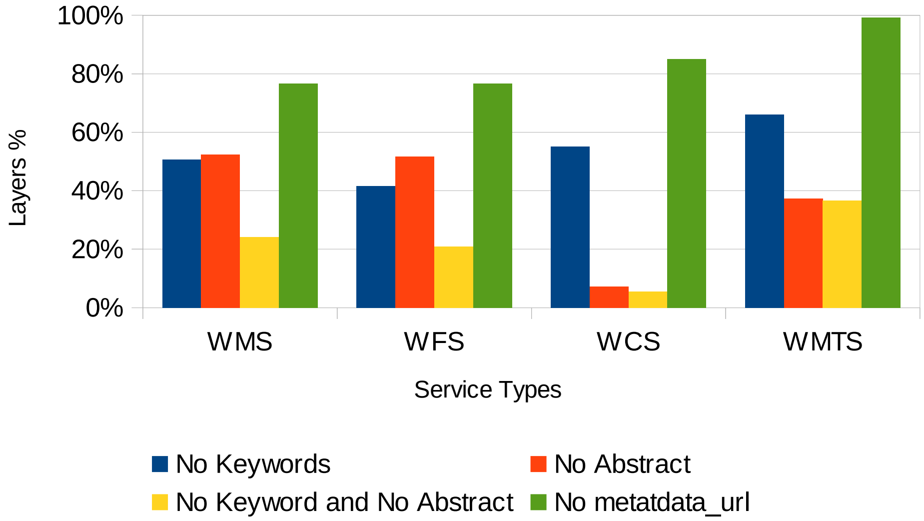

Figure 6: Absence of various types of metadata per layer. Default keywords were stripped before analysis.

As Figure 7 shows, even layers that do have abstracts don't necessarily have useful ones. We split abstracts into “short” (< 50 characters), and “long” (>= 50 characters), and while most WCS layers do have abstracts, a large percentage of them (42%) are short. WMS layers had fewest long abstracts, at only 26% of layers.

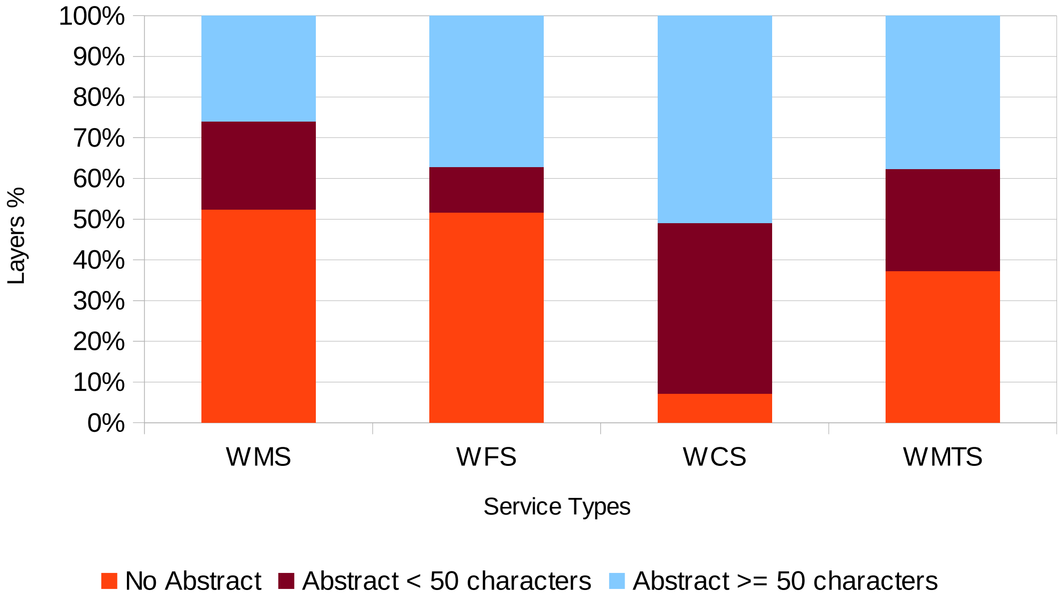

Figure 7: Length of the layer abstract field per service type.

5. Conclusions

Historical data clearly shows a massive recent increase in the available of data via the OWS standards, while contemporary data highlights the ever increasing dominance of just a few hosts, and the different national SDI strategies used to provide data via these standards.
Layer and service discoverability remain a considerable problem, and while a number of solutions have been proposed over the years, almost all require current, filled-in metadata, which real-world data indicates is provided only about half the time.

6. Acknowledgements

With thanks to JK Jolly for assistance with the editing of this paper.

References

Anthony M and Nebert D (2012). *Monitoring the performance and reliability of geospatial web services: Service Status Checker (SSC) system overview.* In Proceedings of the Global Spatial Data Infrastructure World Conference, Québec, QC, Canada.

Bone C, Ager A, Bunzel K, Tierney L (2014). *A geospatial search engine for discovering multi-format geospatial data across the web.* International Journal of Digital Earth, 9, 1–16.

Gui Z, Cao J, Liu X, ChengX, Wu H (2016). *Global-Scale Resource Survey and Performance Monitoring of Public OGC Web Map Services.* ISPRS International Journal of Geo-Information 5: 88.

Huang C and Chang H (2016). GeoWeb crawler: *An extensible and scalable Web crawling framework for discovering geospatial Web resources.* ISPRS International Journal of Geo-Information 5 (8),136.

INSPIRE E. Directive (2007). *Directive 2007/2/EC of the European Parliament.* Official Journal of the European Union.

Mozilla (2019). *LetsEncrypt Statistics* https://letsencrypt.org/stats/ [Accessed 15 January 2019].

Li W, Yang C, Yang C (2010). *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 24(8), 1127–1147.

Li W (2017). *Lowering the Barriers for Accessing Distributed Geospatial Big Data to Advance Spatial Data Science: The PolarHub Solution*, Annals of the American Association of Geographers, 108(3), 773-793.

López-Pellicer F, Béjar R, Florczyk A, Muro-Medrano P, Zarazaga-Soria F (2010). *State of Play of OGC Web Services across the Web*. In INSPIRE Conference 2010. Krakow, Poland.

López-Pellicer F, Béjar R, Rentería-Agualimpia W, Florczyk A, Muro-Medrano P, Zaragoza-Soria F (2011). *Status of INSPIRE inspired OGC Web Services*. INSPIRE Conference 2011. Edinburgh, Scotland.

López-Pellicer F, Rentería-Agualimpia W, Nogueras-Iso J, Zarazaga-Soria F, and Muro-Medrano P (2012). *Towards an Active Directory of Geospatial Web Services.* In Bridging the Geographic Information Sciences, (63–79).

MicroImages (2008). *MicroImages' WMS and ArcIMS Catalogs* http://www.microimages.com/documentation/TechGuides/74MIwebCatalogs.pdf [Accessed 15 January 2019].

Open Geospatial Consortium Inc. (2000). *OpenGIS Web Map Server Interface Implementation Specification*

Reichardt M (2005). *GSDI depends on widespread adoption of OGC standards.* Proceedings of the FIG working week and GSDI8. Cairo, Egypt.

Sample J, 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.* Internet Computing, 10, 53–6.

Skylab (2013). *Skylab WMS List*. http://www.skylab-mobilesystems.com/en/wms_serverlist.html [Accessed 15 January 2019].

Biography

J. Moules has a BSc. in Internet Information Systems and an MSc. in GIS. A GIS contractor with over 10 years of GIS experience, he is the creator of the GeoSeer OWS search engine. His research interests include OGC Web Services, Search Engines, Service Discoverability, and Web-GIS.