

# GeoSeer, A Global Search Engine for OGC Services

Jonathan Moules

INSPIRE Workshop, Ispra, July 2019

- GeoSeer
- Statistics
- What Works
- Problems
- Summary/Conclusions

# GeoSeer

# GeoSeer

The spatial data search engine

Search over **1.4 million** distinct spatial **WMS, WFS, WCS, and WMTS** layers hosted on over **190,000 services** from around the world.

#### Search Tips

**British Columbia** or **geology Australia** or **rivers Asia** - Search spatially.

**tree preservation 27700** - Include an EPSG code to prefer results in that projection.

**arbol Spain** - Not all layers are in English, try the official language.

**wcs Germany chart** - Search for a specific service (*wcs, wmts, wfs, wms*). We'd suggest only including one service in a search.

**York - "New York"** - Standard search features are supported too. *Exact phrase searching* (wrap with ") and *exclude* (prefix with -)

**wms (Alps | Himalayas)** - Boolean logic - by default GeoSeer puts an **AND** between all your search terms meaning **all terms are mandatory** in the result. You can use *OR* (with the pipe: |) instead to expand the search results, and brackets **()** for grouping.

Latest blog post: [2019-04-09 - GeoSeer API Goes Live](#)

[Blog RSS Feed](#) 

# Portals as Seeds

Portal Type	Number	Appx. Number of Services
CKAN	194 Instances	10,000*
CSW	151 Instances (481 – June 2019)	21,000
<a href="http://opendata.arcgis.com">http://opendata.arcgis.com</a>	-	800
<a href="http://inspire-geoportal.ec.europa.eu">http://inspire-geoportal.ec.europa.eu</a>	-	13,000+
GeoSeer	-	186,170

Table shows the **approximate** number of live distinct Services that can be found within a certain data-portal or type of data-portal (April 2019)

\* Over a third of these are the US NOAA

+ Over half of these are one German Agency ([www.geoportal.rlp.de](http://www.geoportal.rlp.de))

# GetCapabilities Documents

```

<?xml version="1.0" encoding="UTF-8"?><WMS_Capabilities xmlns:inspire_vs="http://inspire.ec.europa.eu/
schemas/inspire_vs/1.0" xmlns:inspire_common="http://inspire.ec.europa.eu/schemas/common/1.0"
version="1.3.0" updateSequence="82548" xmlns="http://www.opengis.net/wms" xmlns:xlink="http://
www.w3.org/1999/xlink" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://
www.opengis.net/wms http://msmap1.atkinsgeospatial.com:8080/geoserver/schemas/wms/1.3.0/
capabilities_1_3_0.xsd http://inspire.ec.europa.eu/schemas/inspire_vs/1.0 http://inspire.ec.europa.eu/
schemas/inspire_vs/1.0/inspire_vs.xsd">
  <Service>
    <Name>WMS</Name>
    <Title>Scottish Government - Marine Scotland Maps - Web Map Service</Title>
    <Abstract>An OGC compliant Web Map Server used to deliver spatial data to the National Marine Plan
Interactive (NMPi) web mapping application and associated services. If using this service in desktop GIS,
you must ensure you have appropriate licences for any layers.</Abstract>
    <KeywordList>
      <Keyword>WFS</Keyword>
      <Keyword>WMS</Keyword>
      <Keyword>GEOSERVER</Keyword>
      <Keyword>MARINE</Keyword>
      <Keyword>SCOTLAND</Keyword>
      <Keyword>NMP</Keyword>
      <Keyword>NMPI</Keyword>
      <Keyword>NATIONAL</Keyword>
      <Keyword>PLAN</Keyword>
    </KeywordList>
    <OnlineResource xlink:type="simple" xlink:href="http://wms.marine.gov.scot"/>
    <ContactInformation>
      <ContactPersonPrimary>
        <ContactPerson>Marine Planning Team</ContactPerson>
        <ContactOrganization>Marine Scotland</ContactOrganization>
      </ContactPersonPrimary>
      <ContactPosition/>
      <ContactAddress>
        <AddressType>Work</AddressType>
        <Address>1A(S), Victoria Quay</Address>
        <City>Edinburgh</City>
        <StateOrProvince/>
        <PostCode>EH6 6QQ</PostCode>
        <Country>Scotland</Country>
      </ContactAddress>
      <ContactVoiceTelephone>0131 244 7479</ContactVoiceTelephone>
      <ContactFacsimileTelephone>0131 244 7163</ContactFacsimileTelephone>
      <ContactElectronicMailAddress>marinescotland@gov.scot</ContactElectronicMailAddress>
    </ContactInformation>
    <Fees>NONE</Fees>
    <AccessConstraints>This view service is provided for Marine Scotland Maps (maps.marine.gov.scot).
Consult Marine Scotland Information (marine.gov.scot) and metadata for licensing information</

```

# Postprocessing

- Data cleaning
- De-duplication
- Location Extraction
- Meaningless layer removal
- “Merging” layers (i.e. WMS + WFS)

# What works



# Data Portals

- GetCapabilities - Everything in “one” place
- They’re (kind of) curated
- API’s
  - Discovery Standards
  - Proprietary but reverse engineered

# Discovery Standards are Good

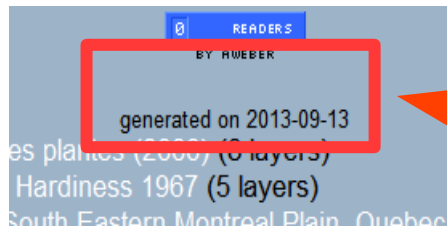
- CSW Standard
- CKAN API
- Large deployments of these

# “GetCapabilities” Document

- Title
- Name
- Abstract
- Metadata
- Keywords
- Bbox
- Etc.

# Problems

# Outdated lists



**S · K · Y · L · A · B**  
Mobilesystems Ltd. *putting the planet into your hand...*

News  
About us  
Products  
Services  
WMS Server List  
Contact  
Imprint

**Latest OGC compliant WMS Servers**

## OGC WMS Server List

Subscribe to our OGC-Services newsletter to be informed about new interesting OGC Web Services.

Name:  Email:

We respect your privacy!

**BlackBerry**  
Alliance Member

generated on 2013-09-13

- AAC - Zones de rusticité des plantes (2000) (8 layers)
- Agriculture Canada - Plant Hardiness 1967 (5 layers)
- Agro-Pedological Atlas of South Eastern Montreal Plain, Quebec (54 layers)
- Areallinformation - Viborg Amt, Denmark (182 layers)
- Areallinformation - Viborg Amt, Denmark (182 layers)
- Atlas agrop?dologique du sud-est de la plaine de Montr?al, Qu?bec (54 layers)
- Atlas of Canada WMS (26 layers)
- Atlas of Canada WMS (26 layers)
- Atlas of Canada WMS (26 layers)
- Atlas of Prague's environment (98 layers)
- Atlas of Prague's environment (99 layers)
- Atlas ??ivoltn??ho prost??ed?? v Praze (99 layers)
- Atlas ??ivoltn??ho prost??ed?? v Praze (98 layers)
- Atlas ??ivoltn??ho prost??ed?? v Praze (99 layers)
- Atlas ??ivoltn??ho prost??ed?? v Praze (99 layers)
- Australian Natural Resources Atlas (339 layers)
- Australian Natural Resources Atlas (339 layers)
- B-Plan Server (1 layer)
- B-Plan Server (1 layer)
- Basic Maps (19 layers)
- Basic Maps (19 layers)
- Basiskarte Sachsen (12 layers)
- Bavarian WMS MAP Server (3 layers)
- Bird Studies Canada WMS/WFS Server (11 layers)
- Bird Studies Canada WMS/WFS Server (11 layers)
- BM (3 layers)
- BM24001T (2 layers)
- BMNG200401 (2 layers)
- BMNG24001TB (2 layers)
- Bulgarian Protected Areas (4 layers)
- Canadian Geographical Names Web Map Service (3 layers)
- Canadian Mineral Deposits (5 layers)
- Canadian Mineral Deposits (5 layers)
- CAROL WMS (1316 layers)
- CAROL WMS (1316 layers)
- CGKN WMS Layers (101 layers)
- CGKN WMS Layers (101 layers)
- Compusuit Limited METAR Weather Mapping service (1 layer)
- Compusuit Limited METAR Weather Mapping service (1 layer)
- CSIR Sensor Web WMS server (15 layers)

XtraServer/WMS fuer GDI NRW Verbund  
XtraServer/WMS fuer GDI NRW Verbund

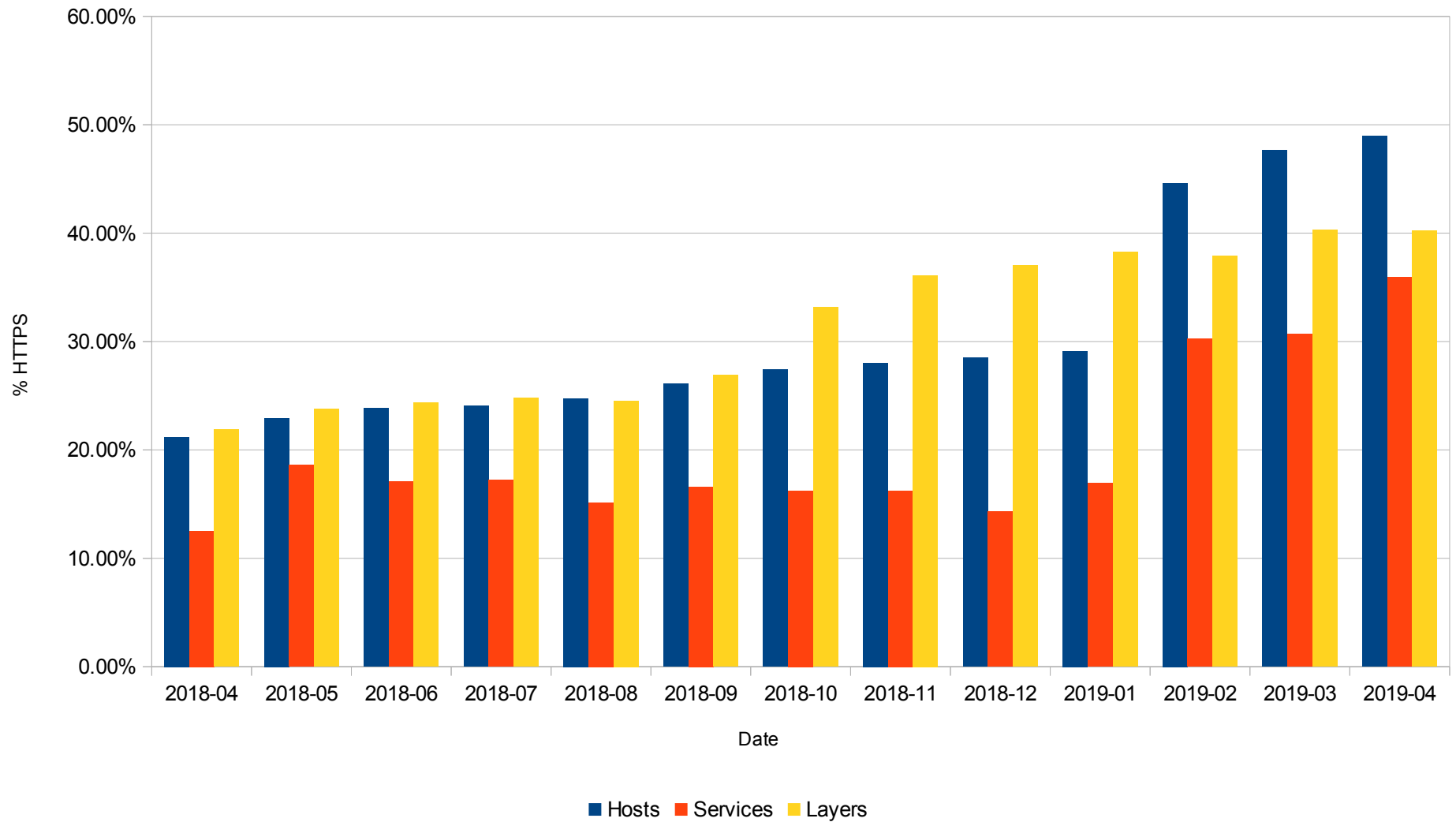
Total Number of Servers: 994  
Total Number of Layers: 339254

Not found 965  
Worked 4

4 of 969 worked! (0.4%)

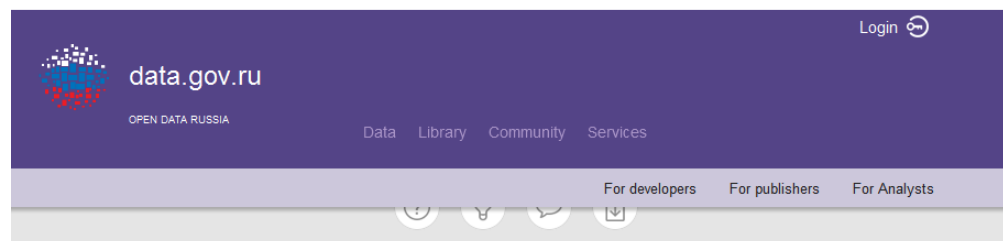
(and 2 of the 4 are the same!)

# HTTP (insecure) vs HTTPS (secure)



# “Registration Required”

- Data Portals
- OGC Services



Services → Developer resources

API

Get the key to the API

Для разработчиков и опытных пользователей, с целью быстрого получения массивов открытых данных в машиночитаемых форматах на Портале открытых данных РФ реализована возможность подачи запросов посредством API.

API находится в стадии разработки и поэтому методы могут быть изменены без предупреждения, но с последующим информированием пользователей.

Для работы с API требуется получить личный ключ, который доступен после регистрации на Портале.

Более подробную и детальную информацию вы можете найти в [документации по API](#).

После получения ключа вы можете использовать API. Ключ необходимо прописывать в конце запроса в следующем виде: `?access_token=ваш ключ`. Например, для получения списка датасетов полный запрос будет выглядеть следующим образом: [http://data.gov.ru/api/dataset?access\\_token=ваш ключ](http://data.gov.ru/api/dataset?access_token=ваш ключ)

API предлагает несколько форматов вывода результатов запроса. На данный момент поддерживаются следующие форматы:

- **JSON** (в строке запроса используется название в нижнем регистре «`json`»)
- **XML** (в строке запроса используется название в нижнем регистре «`xml`»)

# Other Portals

- **DKAN**
  - No Discovery API
- **Bespoke Portals**



## De-jure over De-facto

- ArcGIS “Geoportal”
- [inspire-geoportal.ec.europa.eu](http://inspire-geoportal.ec.europa.eu) – CSW
- (Not so much an issue for OGC Services)

# Duplicates / Default Data

- Same data across many services
- One organisation had over 2 million layers!
- Common Default data
  - Spearfish
  - TOPP
  - SF
  - TIGER
  - etc

www.spatineo.com Currently monitoring 107173 Spatial Web services with 2334948 known data offerings around the world

Arctic Ocean Arctic Ocean

Spatineo Directory

Type in names, keywords, locations etc. to find spatial web services

In English

South Pacific Ocean

Country United Kingdom  INSPIRE services only

**Spatineo Ranking Top 3**

1. BGS Bedrock and Superficial geology  
British Geological Survey (BGS)
2. BGS Bedrock and Superficial geology  
British Geological Survey (BGS)
3. Terrafirma PSI legacy data service  
British Geological Survey (BGS)

# Metadata

- No-one bothers!
- Internal metadata is better
  - Less to crawl
  - No need to “glue” things together
  - High Reliability
  - Less bitrot
  - All in “one place”; i.e. adding to QGIS

# Service Metadata

```

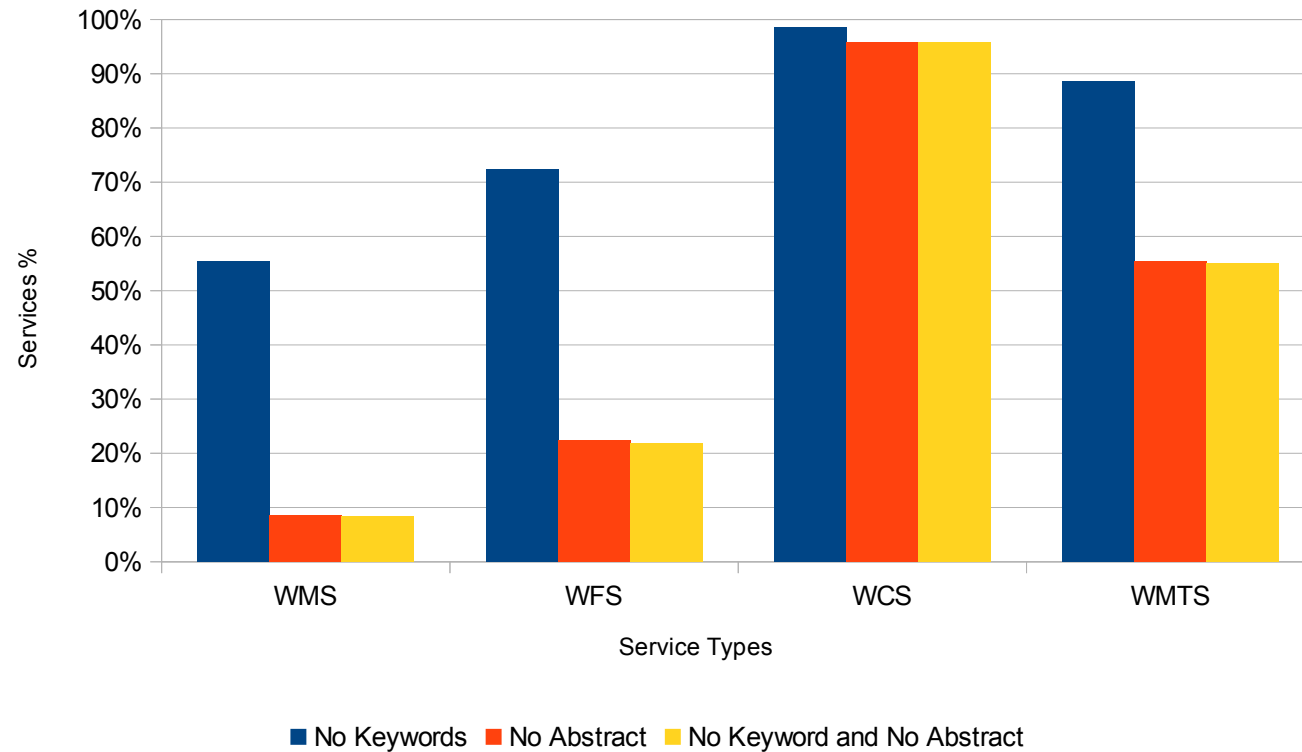
- <Service>
  <Name>WMS</Name>
  <Title>WMS</Title>
  <Abstract>WMS</Abstract>
- <KeywordList>
  <Keyword></Keyword>
</KeywordList>

```

```

- <Service>
  <Name>WMS</Name>
  <Title>Ifremer Thredds Data Server</Title>
  <Abstract>Scientific Data</Abstract>
- <KeywordList>
  <Keyword>meteorology</Keyword>
  <Keyword>atmosphere</Keyword>
  <Keyword>climate</Keyword>
  <Keyword>ocean</Keyword>
  <Keyword>earth science</Keyword>
</KeywordList>

```



## Metadata at the service level

Excludes default abstracts

# Layer Metadata

```

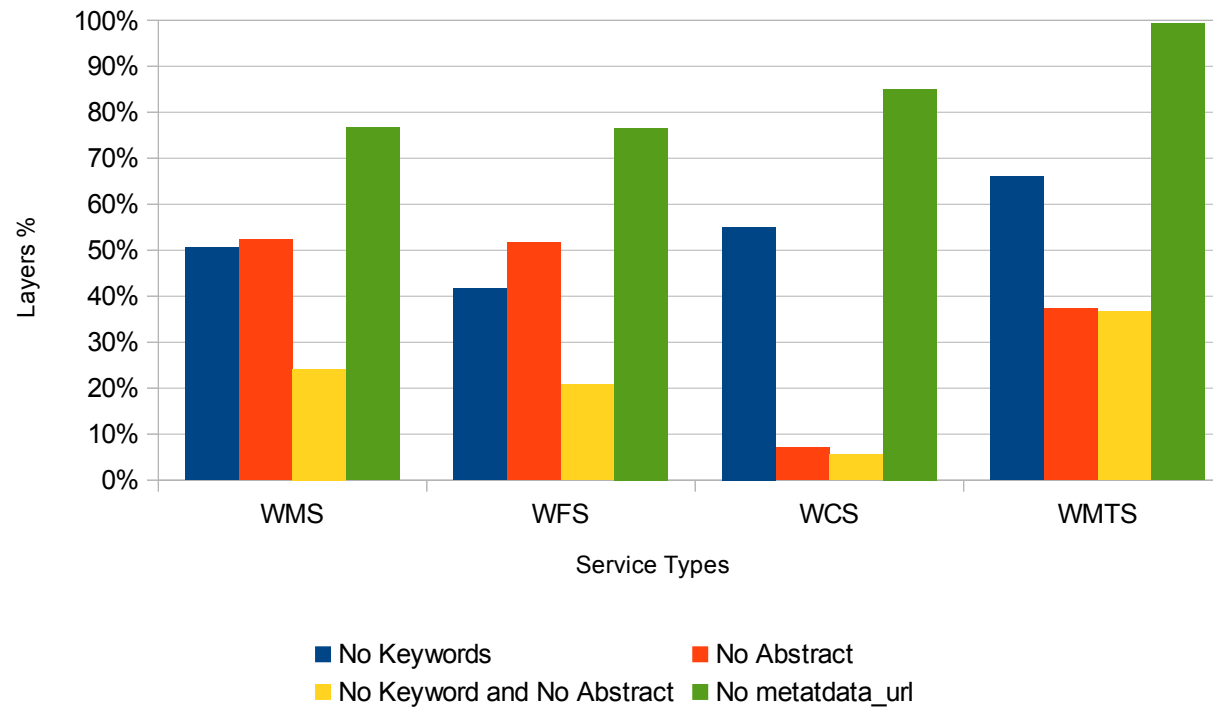
<Name>mass_concentration_of_chlorophyll_a_in_sea_water</Name>
<Title>mass_concentration_of_chlorophyll_a_in_sea_water</Title>
<Abstract>mass_concentration_of_chlorophyll_a_in_sea_water</Abstract>

```

```

- <FeatureType>
  <Name>geonode:eurasia_lsib9</Name>
  <Title> Eurasia LSIB Lines Detailed 2018May30</Title>
- <Abstract>
  This is the detailed version of the detailed Large Scale International Boundaries (LSIB) dataset. Th
  be updated as needed and is current as of May 30, 2018.
</Abstract>
- <ows:Keywords>
  <ows:Keyword>LSIB</ows:Keyword>
</ows:Keywords>
<DefaultCRS>urn:ogc:def:crs:EPSG::4326</DefaultCRS>
- <ows:WGS84BoundingBox>
  <ows:LowerCorner>-8.87743501799997 -9.50374889599993</ows:LowerCorner>
  <ows:UpperCorner>141.022151 70.0918450000001</ows:UpperCorner>
</ows:WGS84BoundingBox>
<MetadataURL xlink:href="http://3.80.149.152/catalogue/csw?outputschema=http%3A%2F%2Fw
<MetadataURL xlink:href="http://3.80.149.152/catalogue/csw?outputschema=http%3A%2F%2Fgc
<MetadataURL xlink:href="http://3.80.149.152/catalogue/csw?outputschema=http%3A%2F%2Fw
<MetadataURL xlink:href="http://3.80.149.152/catalogue/csw?outputschema=urn%3A%2F%2Fw
<MetadataURL xlink:href="http://3.80.149.152/catalogue/csw?outputschema=http%3A%2F%2Fw
<MetadataURL xlink:href="http://3.80.149.152/catalogue/csw?outputschema=http%3A%2F%2Fw
<MetadataURL xlink:href="http://3.80.149.152/showmetadata/xsl/83"/>
</FeatureType>

```



Metadata at the layer level

# Layer Metadata - Abstracts

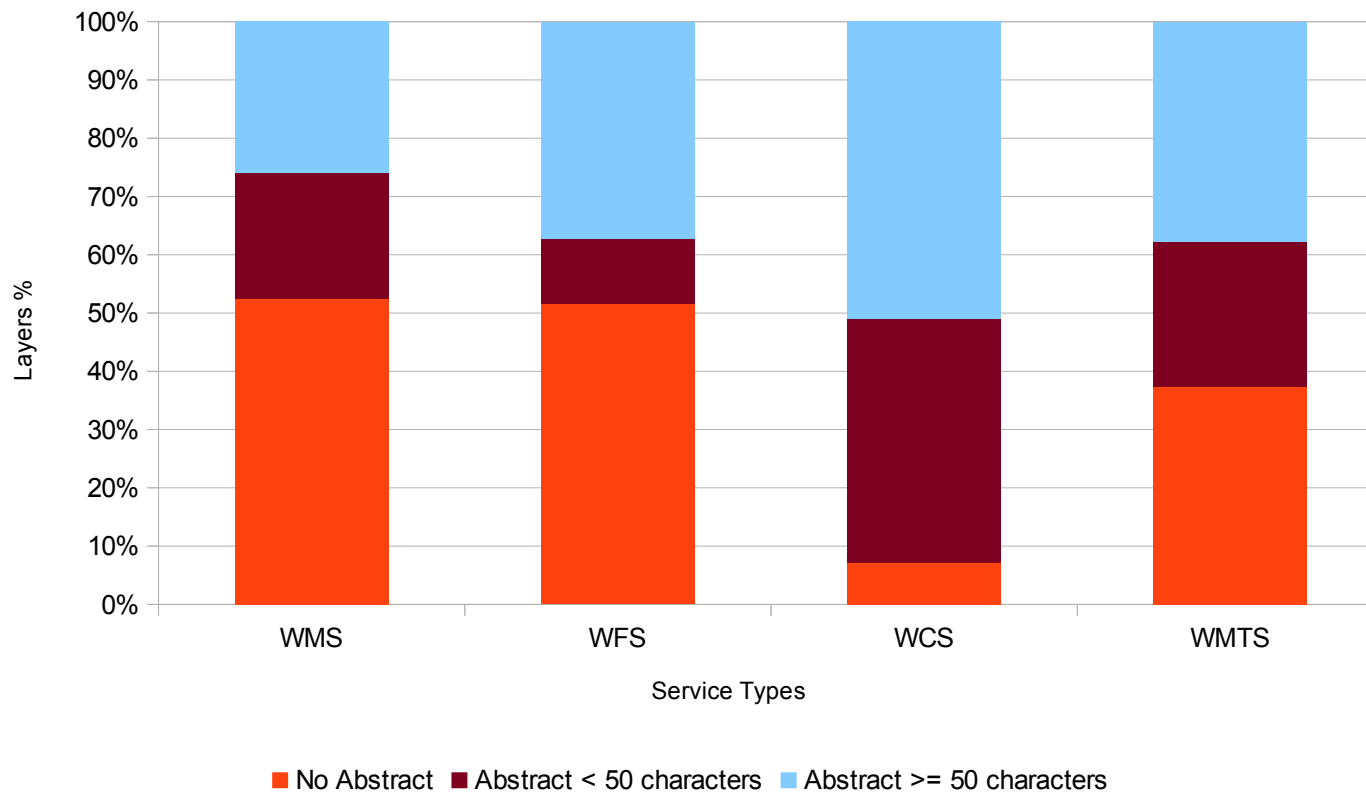


Table showing Layer percentages with meaningful abstracts, grouped by Service Type

# “Meaningless” layers

- Title
  - No Title
  - Title is a random string of gibberish
- No Abstract or Keywords
- 237,382 “meaningless” layers removed for June 2019

# Poorly Configured Servers

- Declared GetCapabilities\_url is wrong
  - Missing scheme
  - Localhost,
  - Otherwise invalid

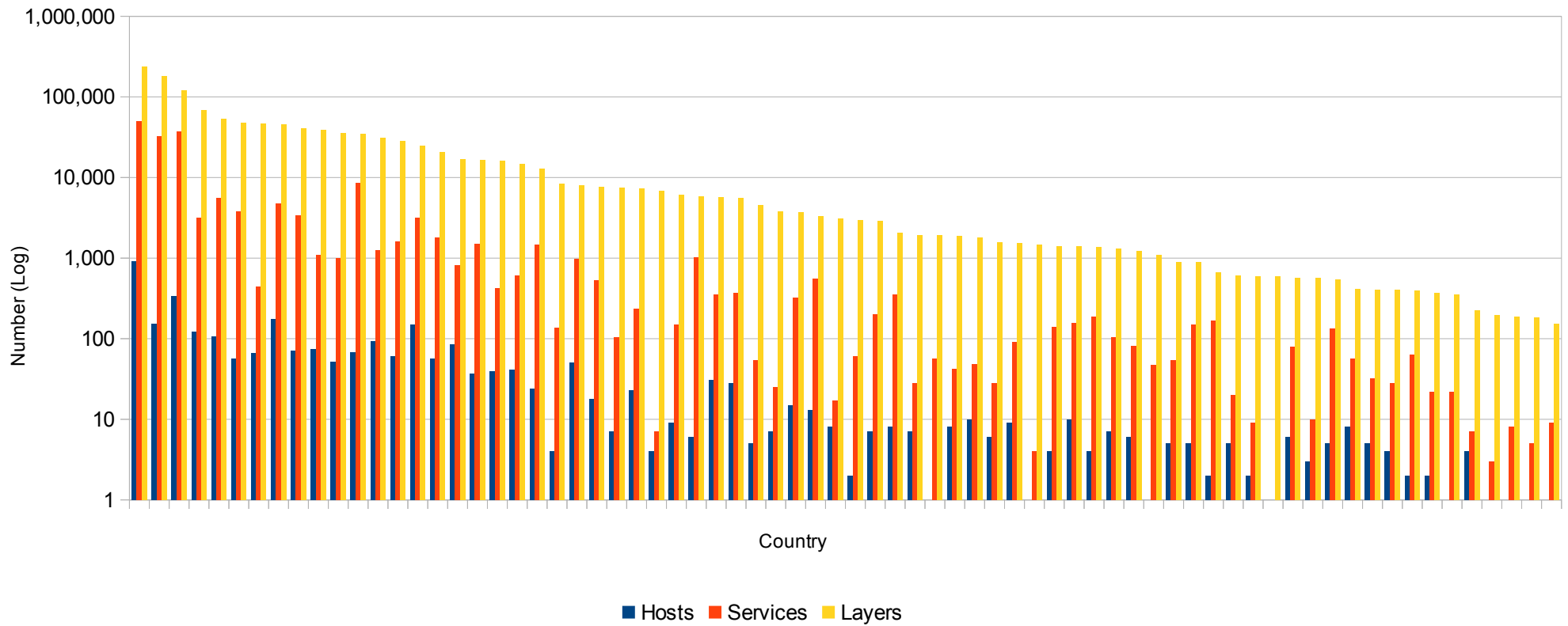
731 services (~0.3%)
- Bad XML / Exceptions
- No Fee/Access Constraint info provided (~50%)
- INSPIRE – Of 28,078 services that mention “inspire”, 6,403 (~22%) don't have an ExtendedCapabilities



# Other things

- Unreliable Services
  - Offline during the crawl = not in the index
- Language barrier

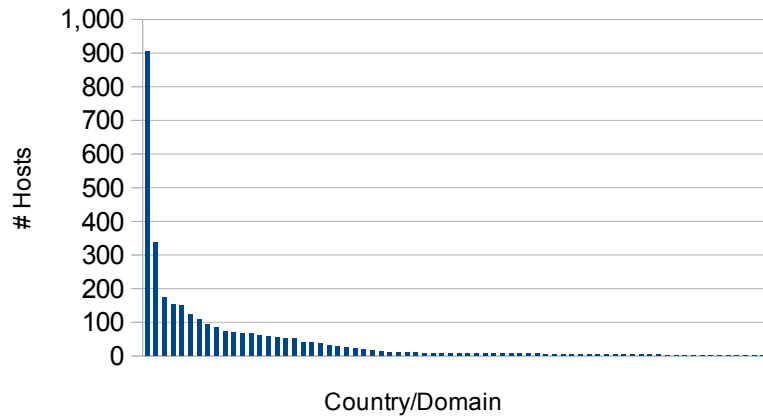
# Statistics



Number of Hosts, Services, and Layers per Country (URL domain).  
Highlights the different national strategies in deployment.

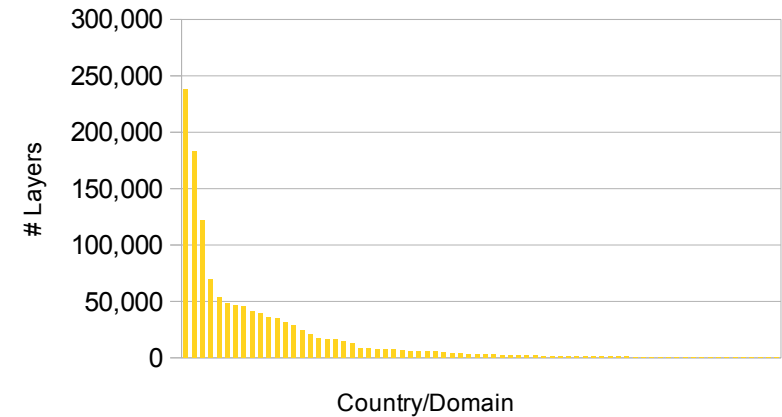
.us, .gov, .edu merged  
 .scot, .wales, .uk merged  
 .brussels, .br merged  
 .int, .eu, .com, .net, .org Not included

# Histograms

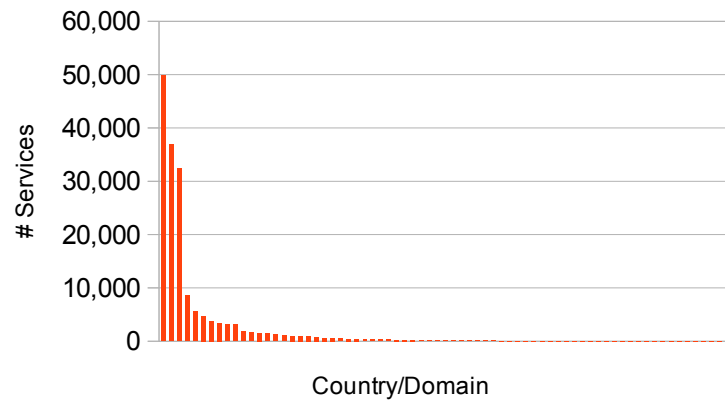


Country	# Hosts
.us	905
.de	337
.ca	174
.fr	152
.uk	150
.nl	123
.es	107
.fi	94
.se	84
.no	74

Country	# Services
.us	49800
.de	36940
.fr	32361
.au	8577
.es	5570
.ca	4717
.pl	3753
.it	3405
.nl	3163
.uk	3150



Country	# Layers
.us	238,126
.fr	183,148
.de	121,749
.nl	69,225
.es	53,548
.pl	48,041
.ar	46,265
.ca	45,523
.it	40,872
.no	39,114



# Summary/Conclusions

- Service providers don't make it easy!
- A UUID would help with de-duplication
- Can't rely on People/Organisations creating metadata

# Questions?

OGC Search Engine:  
[www.geoseer.net](http://www.geoseer.net)

Contact:  
[jonathan.moules@geoseer.net](mailto:jonathan.moules@geoseer.net)