PublicaMundi

Scalable and Reusable Open Geospatial Data

Angelos Tzotsos, IMIS Athena

Scientific & Technical Manager

OSGeo Charter Member

OGC Member

GeoDataCamp 2014, Athens, 30 May 2014

Genesis of the project

- Consortium established on active research, commercial, and personal interactions
- Consortium members are active contributors to Free and Open Source Geospatial Software
- Idea based on practical experience (2010-) from geodata.gov.gr and our interaction with users,publishers and SMEs

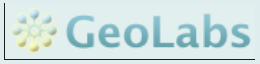
Why PublicaMundi?

- Imago Mundi
  - Image of the World, Babylonian world map
  - Maps, created by cartographers and geographers
- PublicaMundi
  - Image of the World, based on Open Data
  - Maps, created by open knowledge and FOSS

Consortium

Athena IMIS 

Rasdaman 

Geolabs 

GET 

Athena

- Athena Research and Innovation Center in Information, Communication and Knowledge Technologies
- Non-profit, research organization, governed by public law
- Institute for the Management of Information Systems (IMIS)

Rasdaman

- rasdaman GmbH
- R&D centric SME, established in 2003 (MBO in 2010)
- Areas of business: commercial support for rasdaman Array DBMS; consultancy on SDIs & standards
- Geospatial World Innovation Award (2013)

GeoLabs

- GeoLabs SARL
- R&D centric SME focused on FLOSS GIS
- Senegalease Land register (2007)
- 3D module development for Terra Explorer (2008) in IGN 3D Geoportal
- Development of the MapMint SDI using WPS, other OGC Web Services and OASIS

GET

- Geospatial Enabling Technologies LtD
- SME focused on GeoInformatics
- Successful design and implementation of projects for public and private sector related to geospatial data production, management, curation, geospatial applications (desktop, mobile, web)
- One of the first Greek private companies invested in Open Source GIS technology

Experiences and Problems

Reoccurring practical problems in open geospatial data reuse:

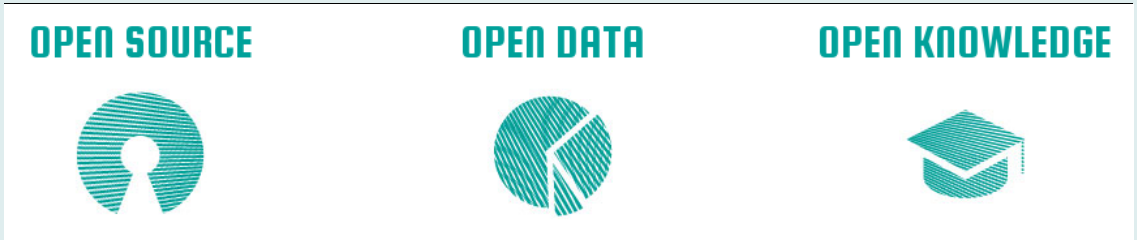
- Data formats/CRS
- Web maps
- Interlink
- Translations
- Publishing
- OGC/INSPIRE documents not for all

Goals

Research and develop methodologies, as well as scalable, reusable tools to facilitate:

- the **publication**
- **discovery**
- and **reuse**

of open geospatial data



Goals

- Open data catalogues fully supporting **publishing, curation and management lifecycle** of geospatial data
- **Interlinking** of geospatial data and **multilinguality** support in a cross-boundary context
- Scalable technologies and services to create and reuse on-demand **maps** from open geospatial data
- **Analytics** to accurately monitor the usage of open geospatial data
- Scalable technologies and reusable data APIs supporting **querying, processing, and analysis** of open geospatial data

Free and Open Source Software (FOSS)

- It's a prerequisite for **Big Data** and **Cloud Computing**
- Everybody uses FOSS today, even without knowing it

CKAN

An abbreviation for “Comprehensive Knowledge Archive Network”

Open Source web platform for **publishing and sharing data** with impressive deployment history:

- EU Open Data Portal
- USA data.gov
- UK data.gov.uk
- Australia data.gov.au
- and many many others...

PublicaMundi and FOSS

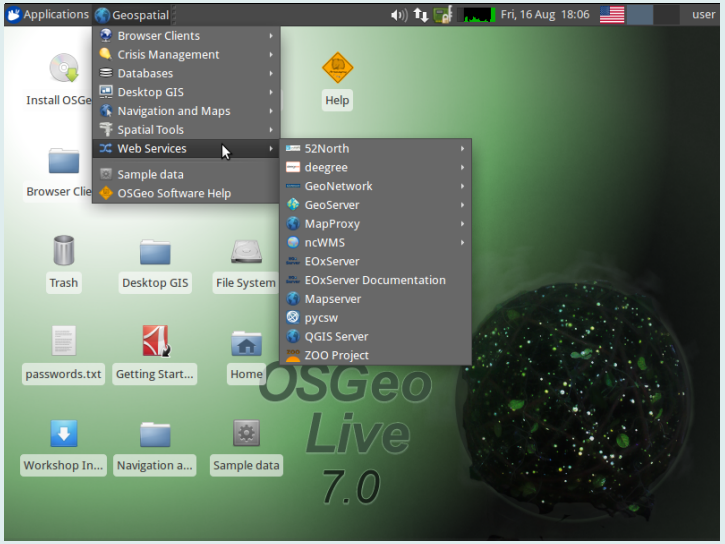
- Based on CKAN open data catalogue
- PublicaMundi will spatially extend CKAN using OGC standards
- Development is based exclusively on the OSGeo stack

Open Source Geospatial Foundation (OSGeo)

Since 2006 A Non Profit Umbrella for:

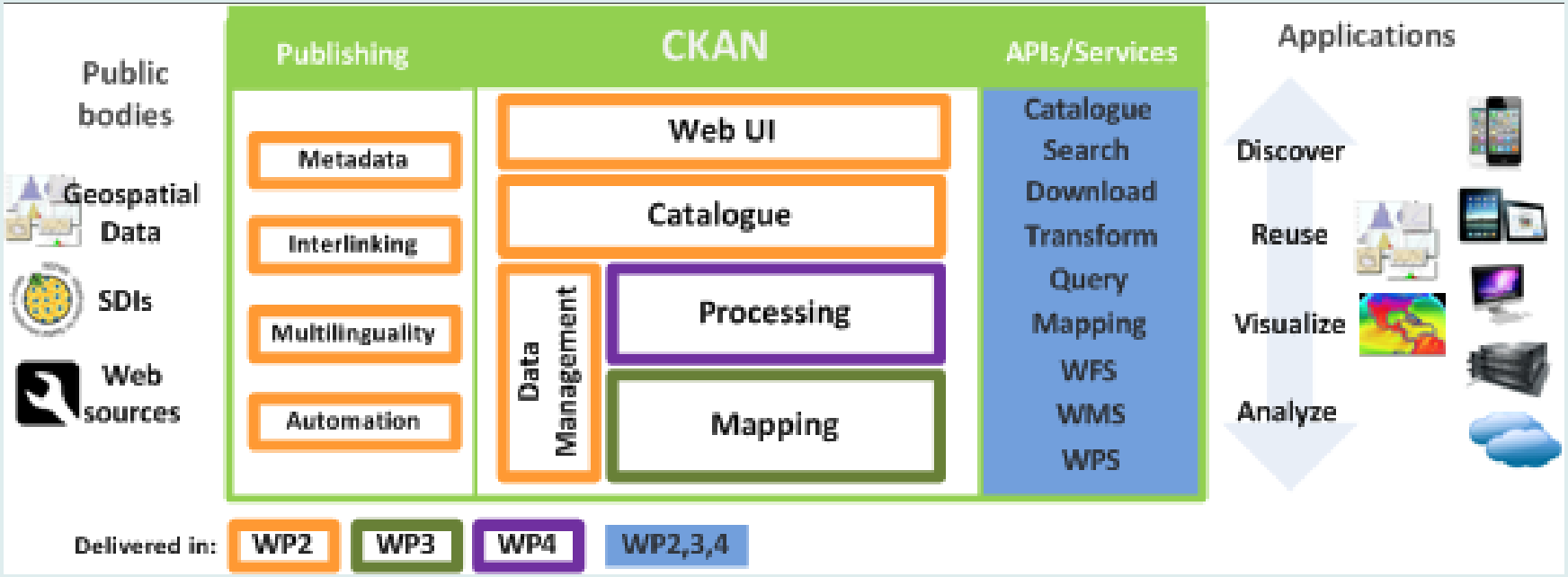
- GeoSpatial Free and Open Source Software
- Education
- Open Data

OSGeoLive



- GNU/Linux distribution
- 60+ Open Source Geospatial Applications
- Sample Datasets
- Consistent Overviews & Quickstarts
- Translations
- Unfortunately Big Data do not fit on a 4GB DVD...

High level architecture



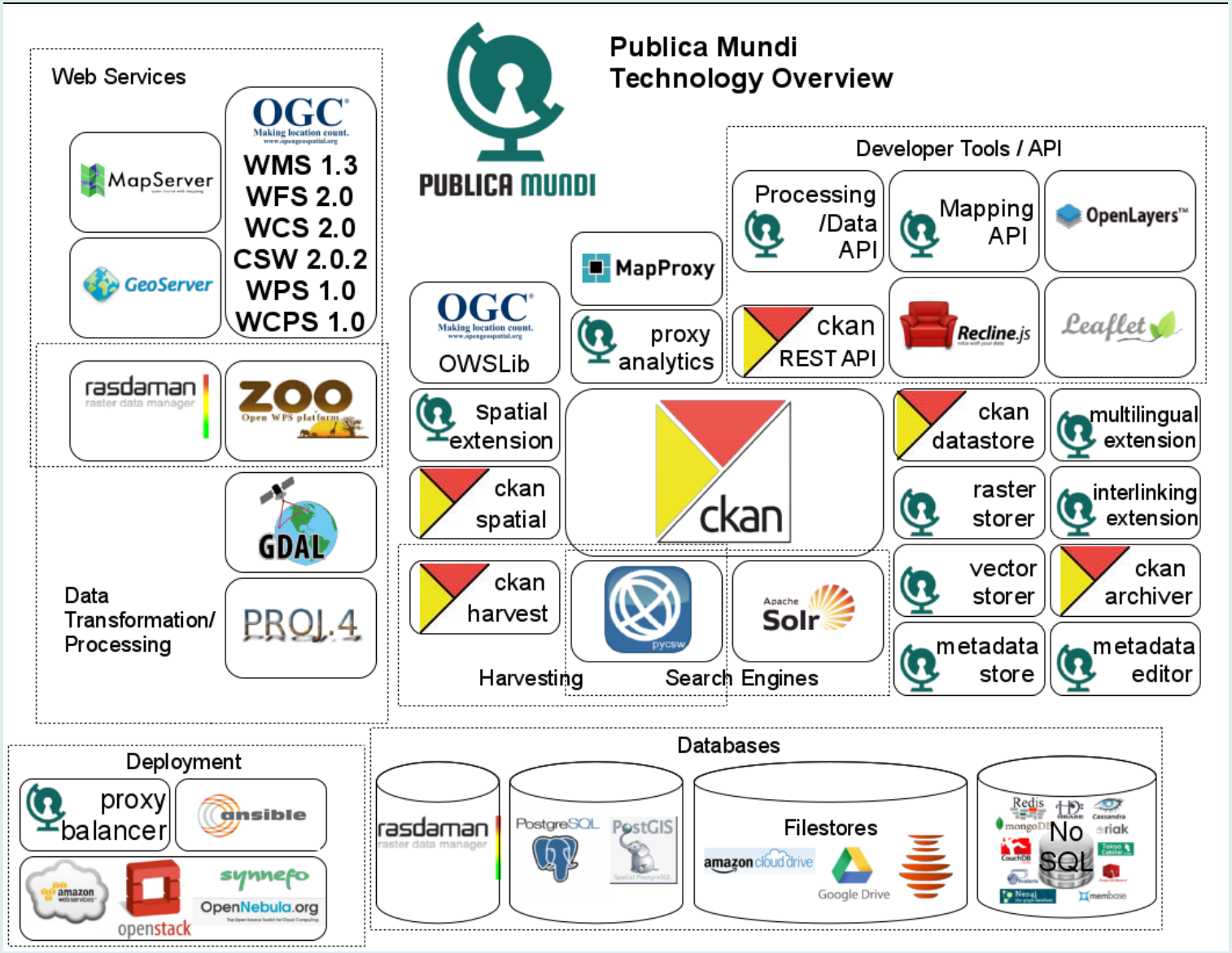
OGC standards and INSPIRE

- Discovery Services
- View Services
- Download Services
- Processing Services

Earth Observation Big Data

- Integration with rasdaman
- WCPS and WPS support

Architecture Overview



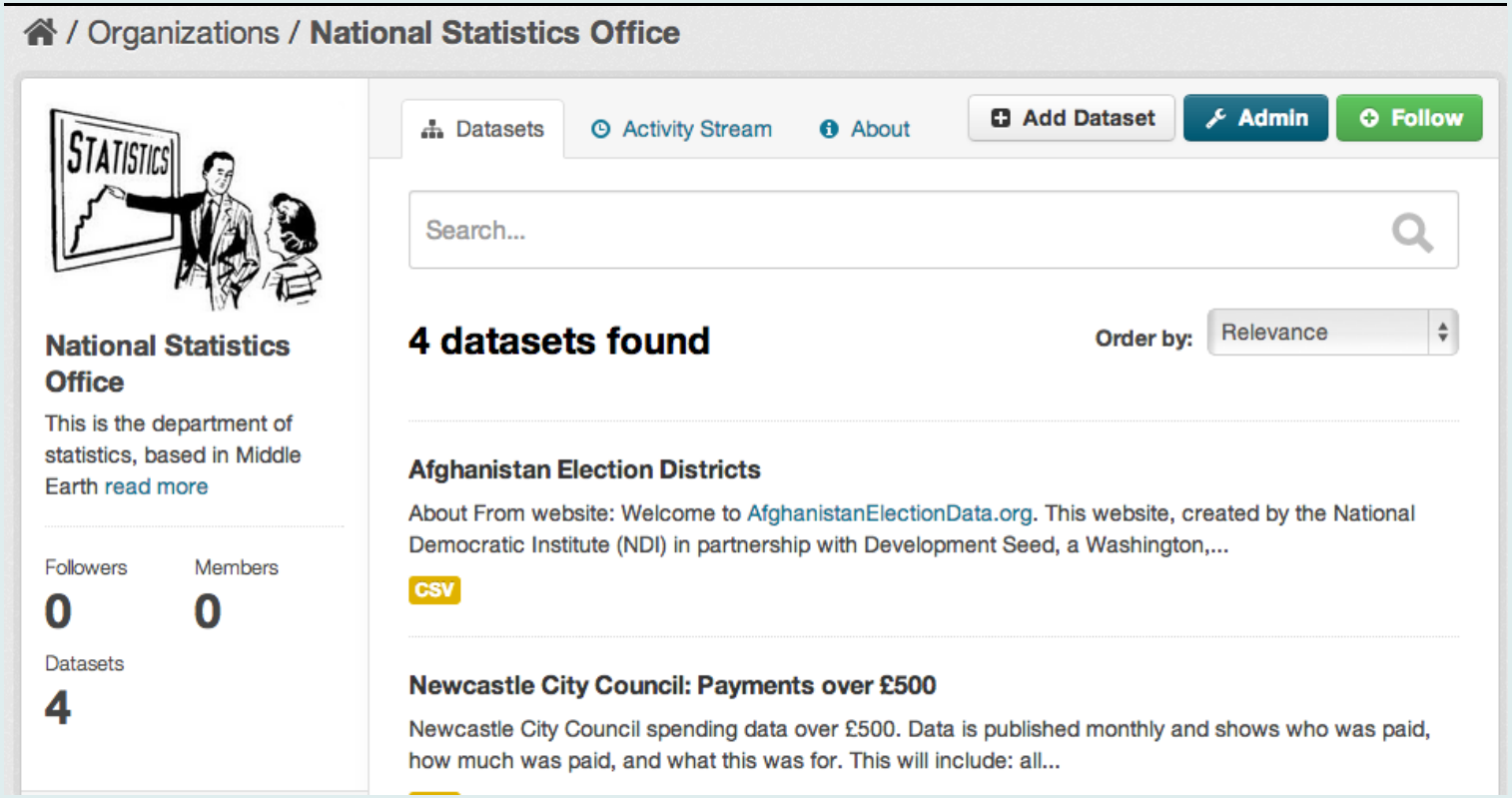
CKAN

- CKAN is a powerful data management system that makes data accessible – by providing tools to streamline publishing, sharing, finding and using data.
- CKAN is aimed at data publishers (national and regional governments, companies and organizations) wanting to make their data open and available.

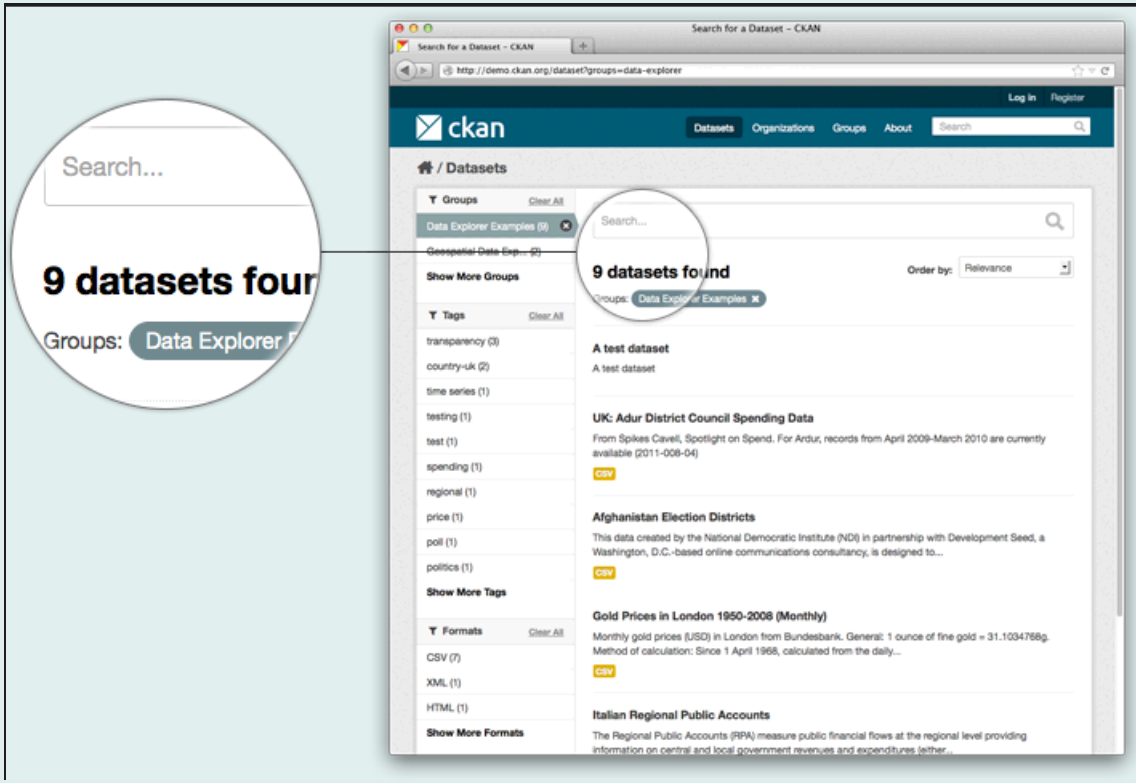
CKAN Features

- Publish and find datasets
- Store and manage data
- Federated nodes
- Harvesting
- Metadata Editing/Management
- APIs and Extensions

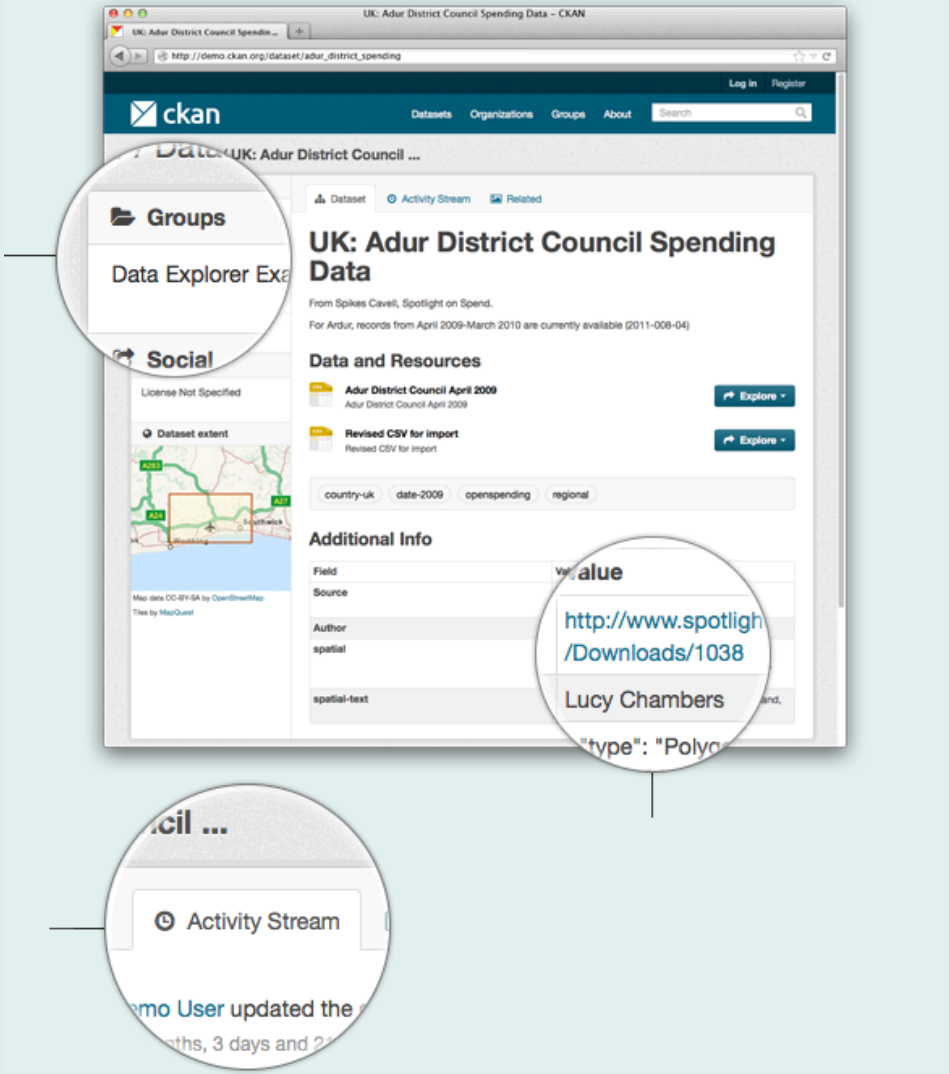
Publish Data



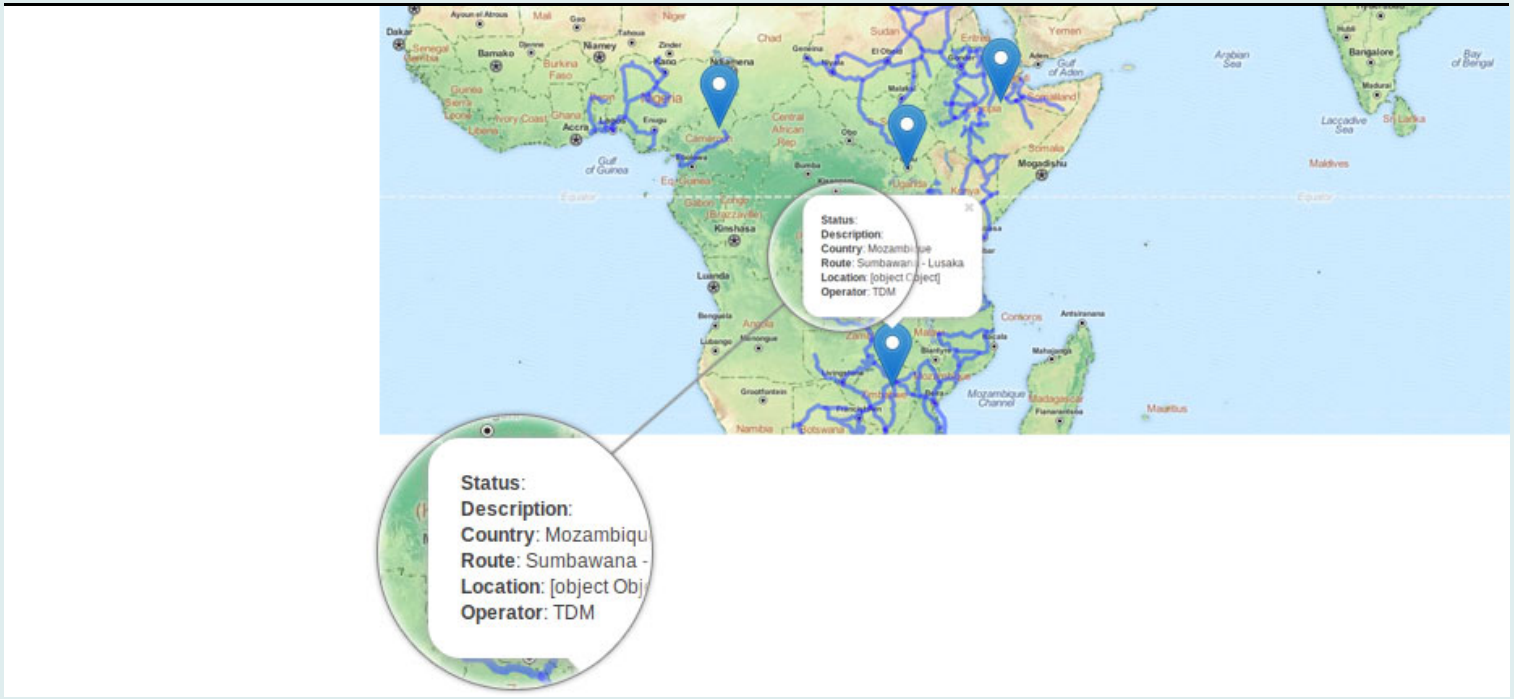
Search and Discovery



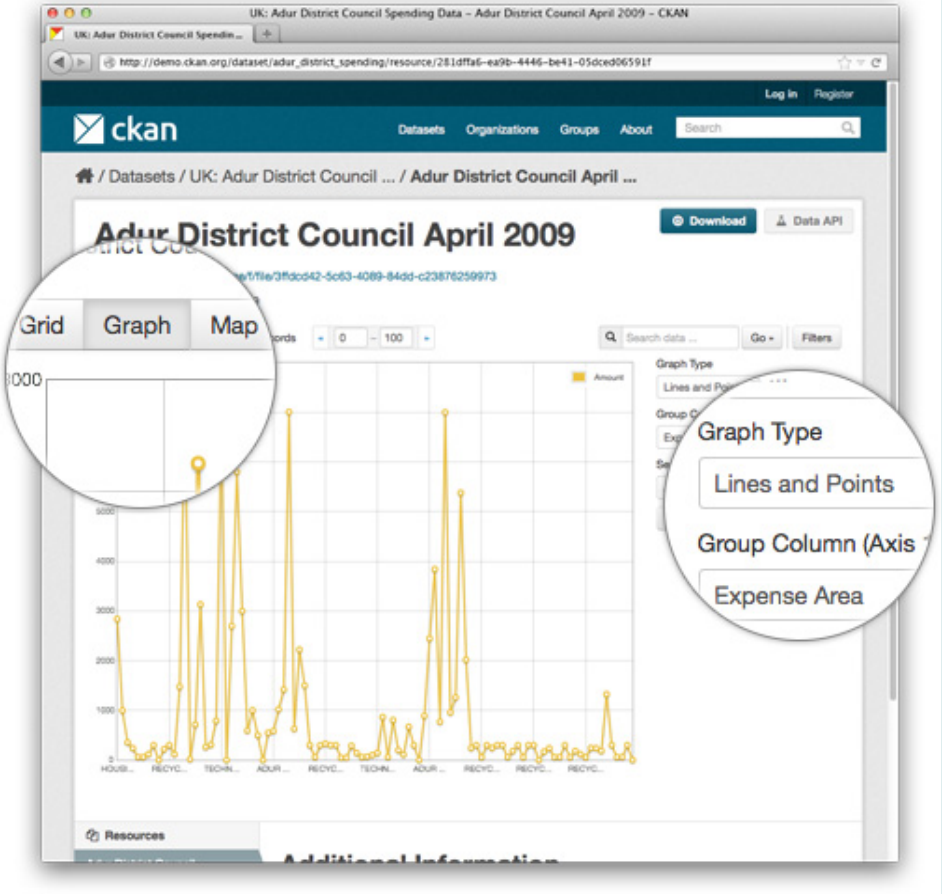
Metadata


UK: Adur District Council Spending Data – CKAN
http://demo.ckan.org/dataset/adur_district_spending
ckan
Datasets
Organizations
Groups
About
UK: Adur District Council ...
Groups
Data Explorer Exa
Social
Dataset
Activity Stream
Related
UK: Adur District Council Spending Data
From Spikes Cavell, Spotlight on Spend.
For Ardur, records from April 2009-March 2010 are currently available (2011-008-04)
Data and Resources
Adur District Council April 2009
Adur District Council April 2009
Explore
Revised CSV for import
Revised CSV for import
Explore
country-uk
date-2009
openspending
regional
Additional Info
Field
Value
Source
Author
spatial
spatial-text
http://www.spotligh
/Downloads/1038
Lucy Chambers
"type": "Polyg
Dataset extent
License Not Specified
Activity Stream
mo User updated the
ths, 3 days and

Geospatial


Status:
Description:
Country: Mozambiqu
Route: Sumbawana -
Location: [object Obj
Operator: TDM

Visualization



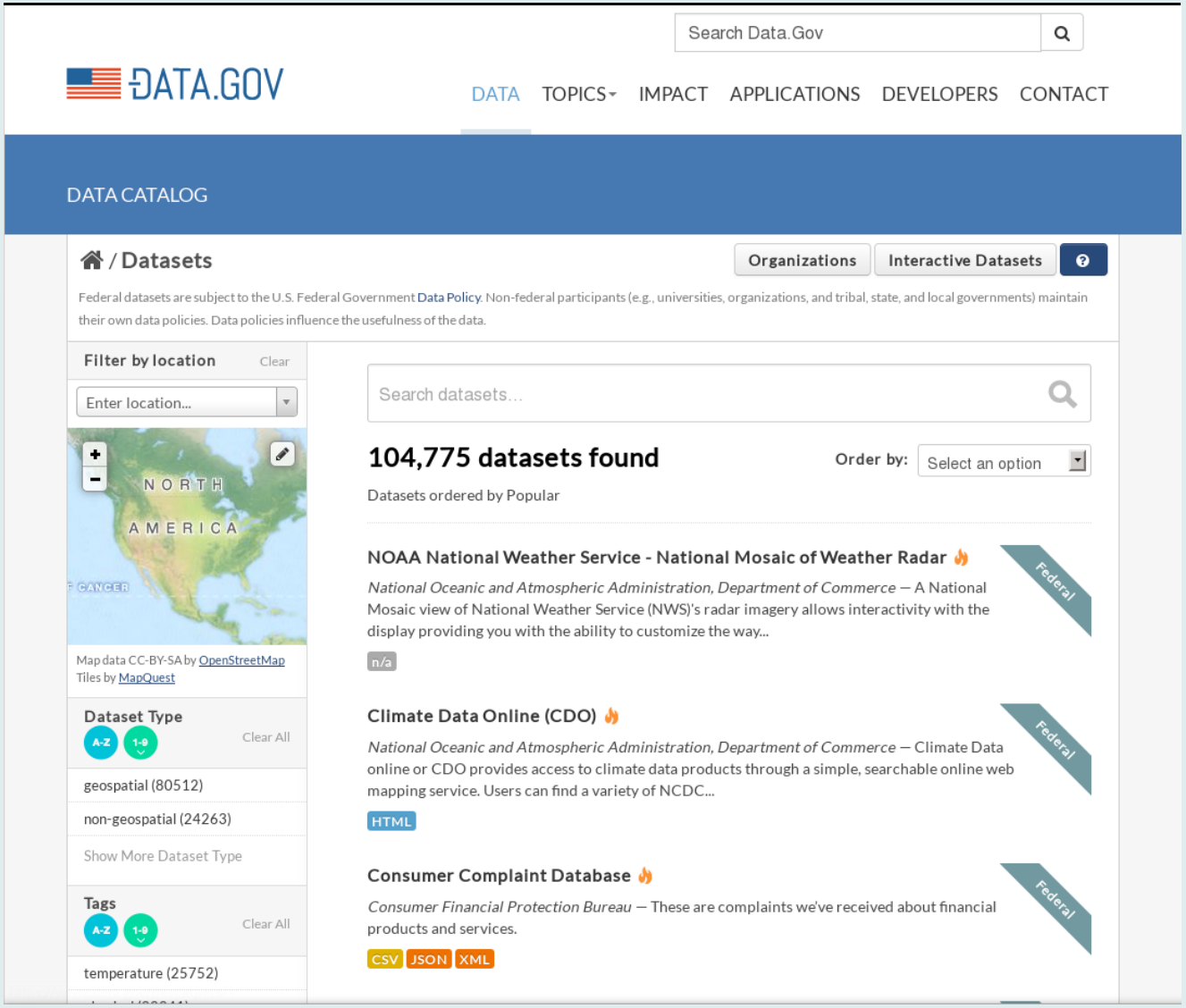
ckanext-spatial - Geo plugins for CKAN

- A spatial field on the default CKAN dataset schema, that uses PostGIS as the backend and allows to perform spatial queries and to display the dataset extent on the frontend
- Harvesters to import geospatial metadata into CKAN from other sources in ISO 19139 format and others
- Commands to support the CSW standard using pycsw
- Plugins to preview spatial formats such as GeoJSON

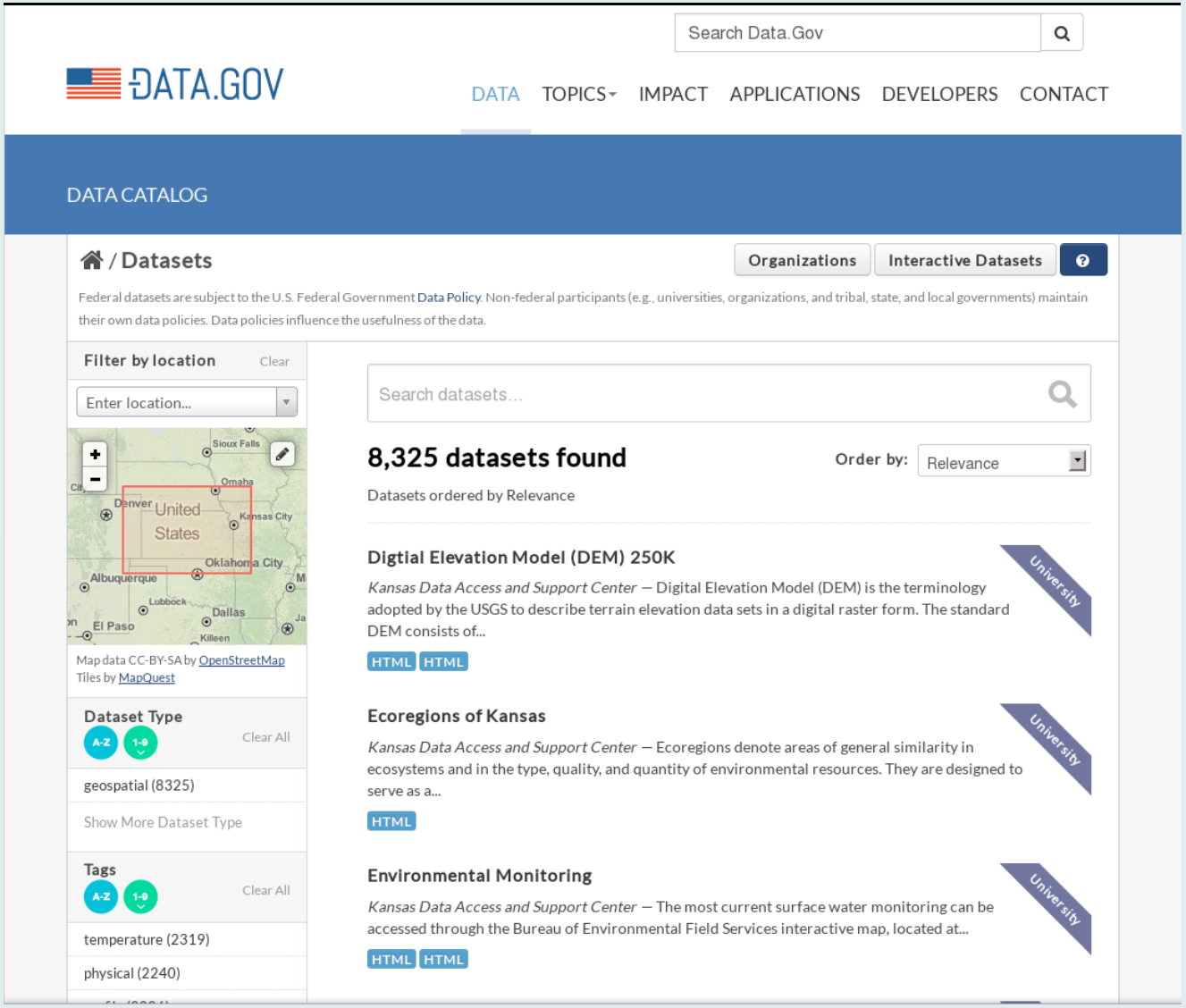
ckanext-spatial Features

- Spatial Search
- Spatial Harvesters
- CSW interface
- WMS Preview

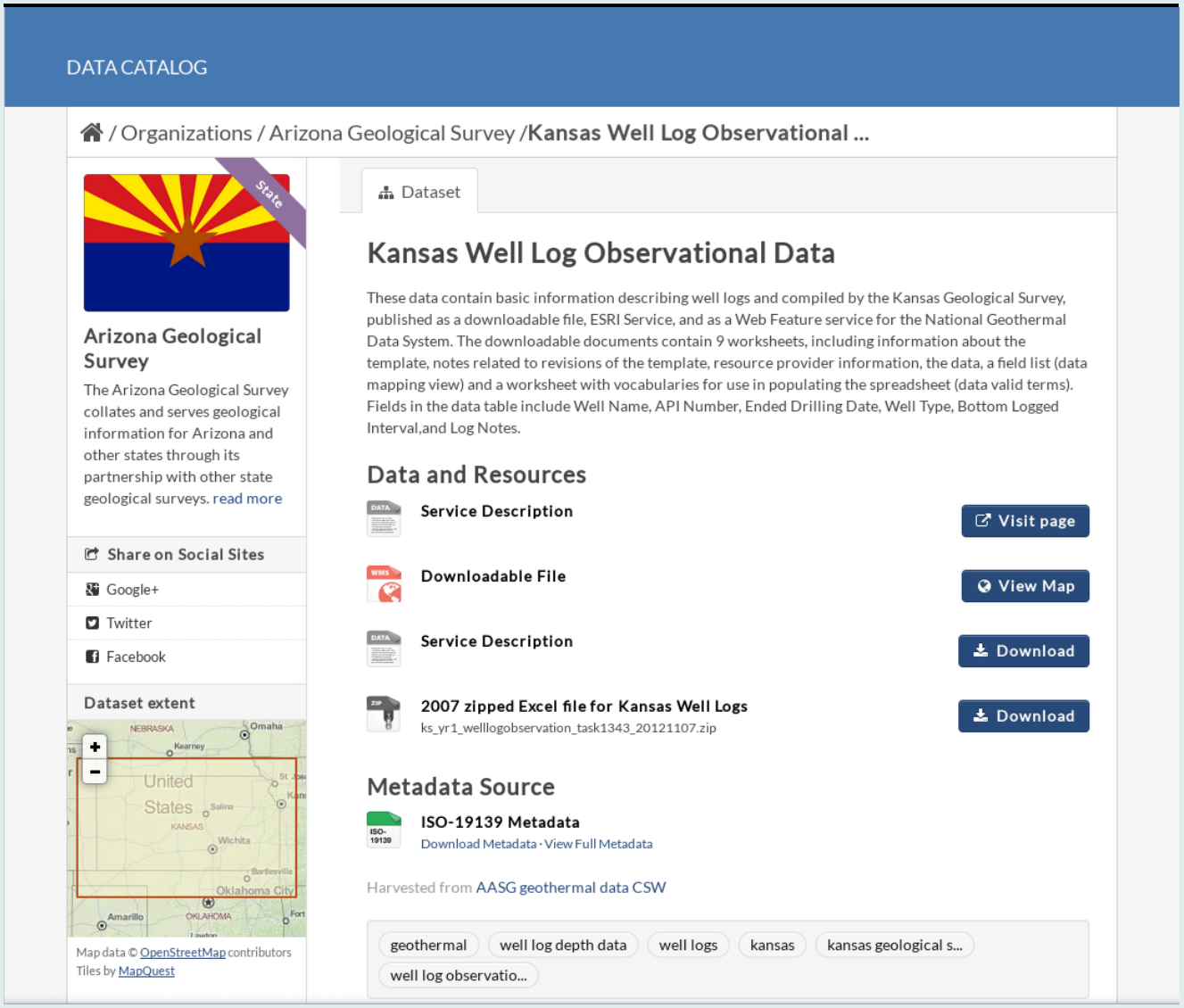
Spatial Search



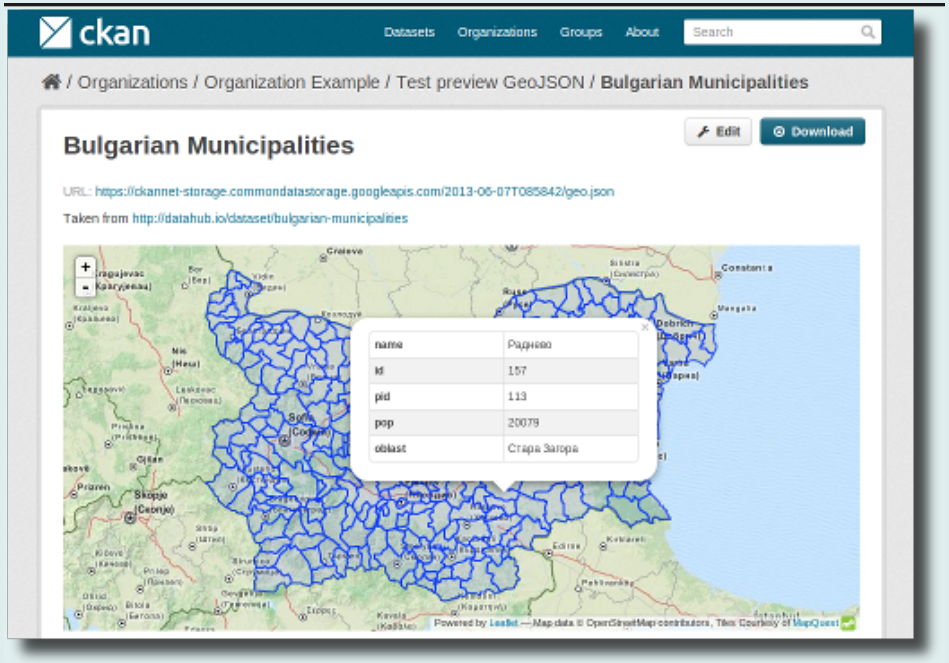
Spatial Search



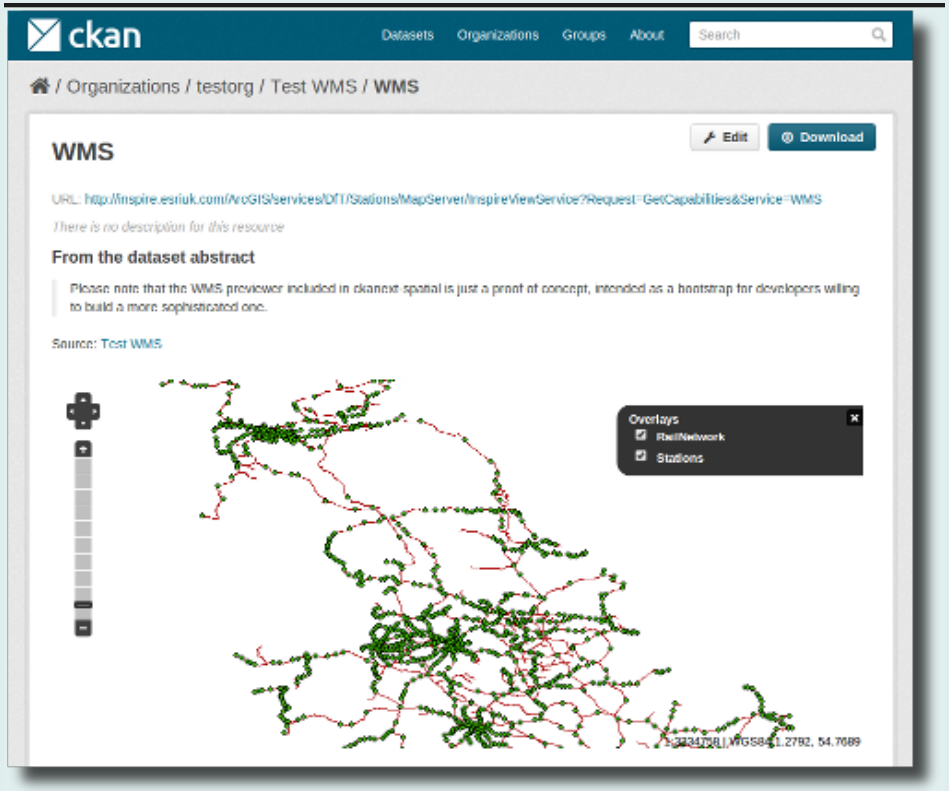
Spatial Datasets



Spatial Datasets Preview



Spatial Datasets Preview

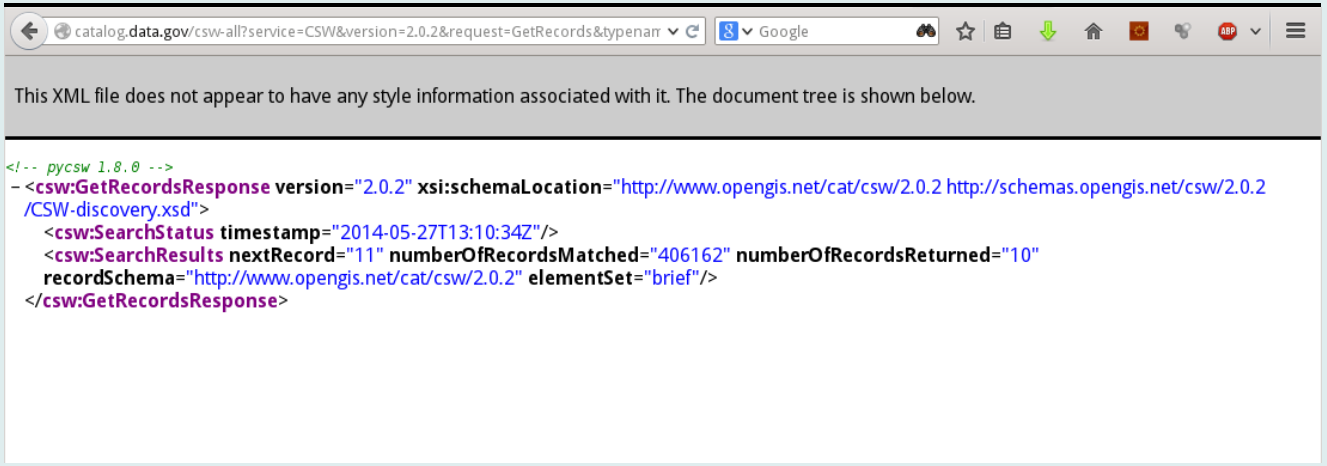


ISO 19115 Metadata

This XML file does not appear to have any style information associated with it. The document tree is shown below.


```
- <gmd:MD_Metadata xsi:schemaLocation="http://www.isotc211.org/2005/gmd http://schemas.opengis.net/csw/2.0.2/profiles/apiso/1.0.0
  /apiso.xsd">
  - <gmd:fileIdentifier>
      <gco:CharacterString>caad517515720208ad5e01bce404cca7</gco:CharacterString>
    </gmd:fileIdentifier>
  - <gmd:language>
      <gco:CharacterString>eng</gco:CharacterString>
    </gmd:language>
  - <gmd:characterSet>
      <gmd:MD_CharacterSetCode codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/Codelist
      /gmxCodelists.xml#MD_CharacterSetCode" codeListValue="utf8">UTF-8</gmd:MD_CharacterSetCode>
    </gmd:characterSet>
  - <gmd:hierarchyLevel>
      <gmd:MD_ScopeCode codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/Codelist
      /gmxCodelists.xml#MD_ScopeCode" codeListValue="Dataset">Dataset</gmd:MD_ScopeCode>
    </gmd:hierarchyLevel>
  - <gmd:hierarchyLevelName>
      <gco:CharacterString>Dataset</gco:CharacterString>
    </gmd:hierarchyLevelName>
  - <gmd:contact>
    - <gmd:CI_ResponsibleParty>
      - <gmd:individualName>
          <gco:CharacterString>Christy Caudill</gco:CharacterString>
        </gmd:individualName>
      - <gmd:organisationName>
          <gco:CharacterString>AZGS</gco:CharacterString>
        </gmd:organisationName>
      - <gmd:contactInfo>
        - <gmd:CI_Contact>
          - <gmd:phone>
            - <gmd:CI_Telephone>
              - <gmd:voice>
                  <gco:CharacterString>520-209-4165</gco:CharacterString>
                </gmd:voice>
              </gmd:CI_Telephone>
            </gmd:phone>
          - <gmd:address>
            - <gmd:CI_Address>
              - <gmd:deliveryPoint>
                  <gco:CharacterString>416 W Congress St</gco:CharacterString>
                </gmd:deliveryPoint>
              - <gmd:city>
                  <gco:CharacterString>Tucson</gco:CharacterString>
```

CSW Interface



pycsw

- pycsw is a OGC CSW server implementation written in Python
- pycsw is an Open Source project released under the MIT license

What is Metadata?

Metadata is often described as “data about data”, or the who, what, where, and when.

In the geospatial world, for each dataset we maintain, we should record information about the data such as:

- general description
- location
- usage restrictions
- projection
- technical contact
- time period
- date created
- date modified
- version

Metadata Standards

- Dublin Core: established a core/common group of 15 metadata elements
- FGDC CSDGM: approved by the U.S. Federal Geographic Data Committee originally in 1994 and composed of Sections, Compound Elements, Data Elements
- ISO 19115: International Standards Organization's TC211 committee created this in 2003 and is composed of more than 400 "Core", "Mandatory", and "Optional" elements
- ISO 19139: The XML implementation schema for ISO 19115 specifying the metadata record format

OGC CSW Specification

The Open Geospatial Consortium (OGC) OpenGIS Catalogue Service Implementation Specification, currently at version 2.0.2, is a standard for discovering and retrieving spatial data and metadata.

Catalogue Services for the Web (CSW) is the HTTP protocol binding of the Catalogue Service Implementation Specificaton that allows for publishing and searching of metadata.

CSW Operations

- **GetCapabilities** (mandatory) - allow clients to retrieve information describing the service instance
- **DescribeRecord** (mandatory) - allows a client to discover elements of the information model supported by the target catalogue service
- **GetRecords** (mandatory) - get metadata records
- **GetRecordById** (optional) - get metadata records by ID
- **GetDomain** (optional) - obtain runtime information about the range of values of a metadata record element or request parameter
- **Harvest** (optional) - references the data to be inserted or updated in the catalog
- **Transaction** (optional) - defines an interface for creating, modifying and deleting catalogue records

Example Requests

- GetCapabilities
- DescribeRecord
- GetRecords
- GetRecordById
- GetDomain

pycsw

- pycsw fully implements the OpenGIS Catalogue Service Implementation Specification [Catalogue Service for the Web]
- pycsw allows for the publishing and discovery of geospatial metadata

pycsw

The project is certified OGC Compliant, and is an OGC Reference Implementation

pycsw is currently under OSGeo Incubation

Features

- Harvesting support for WMS, WFS, WCS, WPS, WAF, CSW, SOS
- Implements ISO Metadata Application Profile 1.0.0
- Implements FGDC CSDGM Application Profile for CSW 2.0
- Implements INSPIRE Discovery Services 3.0
- Supports ISO, Dublin Core, DIF, FGDC and Atom metadata models
- Standalone of embedded deployment (CGI or WSGI)
- Transactional capabilities (CSW-T)
- Flexible repository configuration (SQLite, PostgreSQL, PostGIS, MySQL)
- Federated catalogue distributed searching

More features...

- Simple configuration
- Extensible plugin architecture (profiles, repositories/backends)
- Seamless integration with Python environments (e.g. GeoNode, Open Data Catalog)
- Includes commandline utility to administer the metadata repository
- Implements the Search/Retrieval via URL (SRU) search protocol
- Implements OpenSearch
- Realtime XML Schema validation

Standards Support

- OGC CSW 2.0.2
- OGC Filter 1.1.0
- OGC OWS Common 1.0.0
- OGC OpenSearch Geo/Time
- OGC GML 3.1.1
- OGC SFSQL 1.2.1
- Dublin Core 1.1
- SOAP 1.2
- ISO 19115 2003
- ISO 19139 2007
- ISO 19119 2005
- NASA DIF 9.7
- FGDC CSDGM 1998
- SRU 1.1
- A9 OpenSearch 1.1

ZOO Project WPS

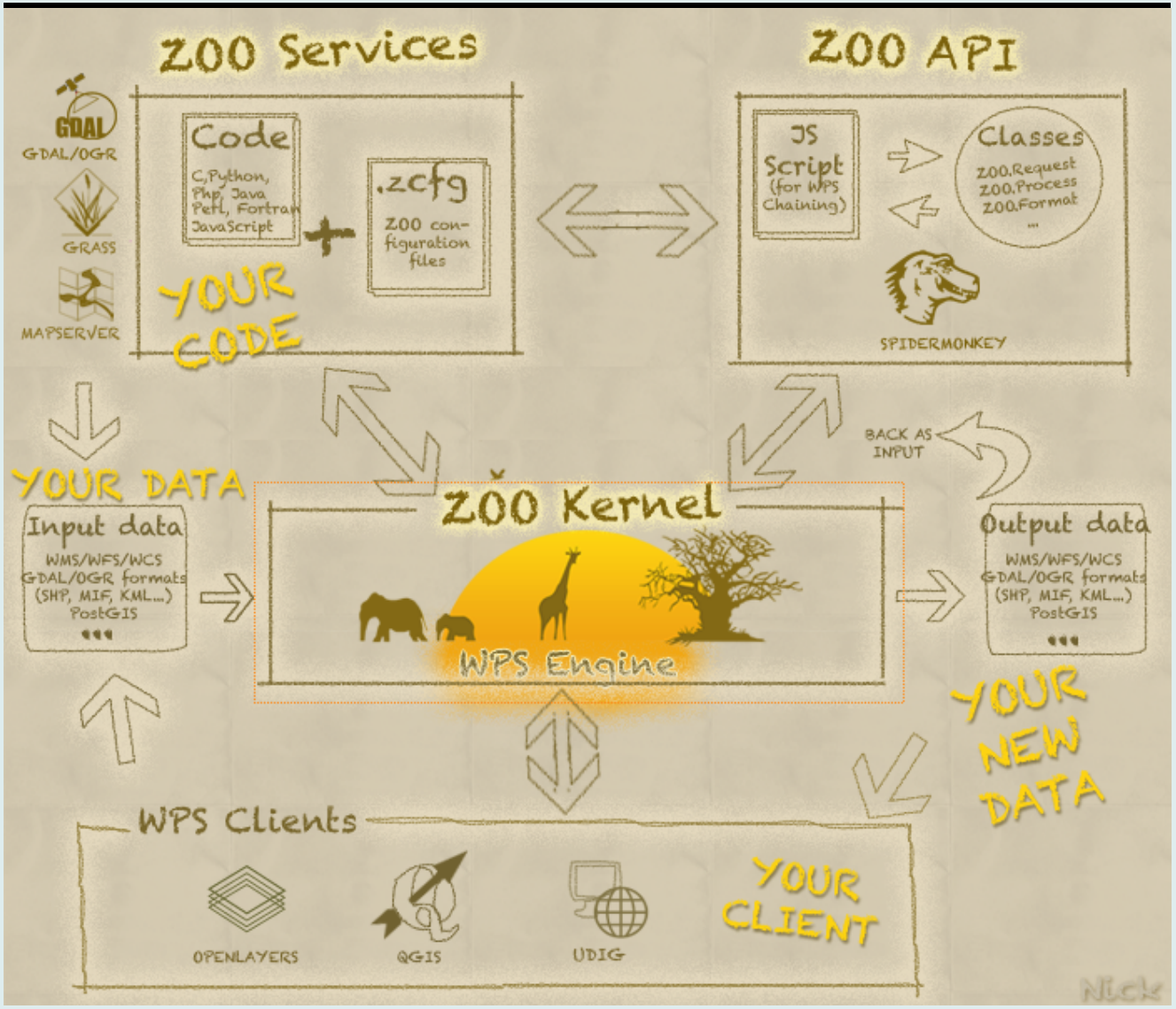
- ZOO is a WPS (Web Processing Service) open source project released under a MIT/X-11 style license
- It provides an OGC WPS compliant developer-friendly framework to create and chain WPS Web services

ZOO Overview

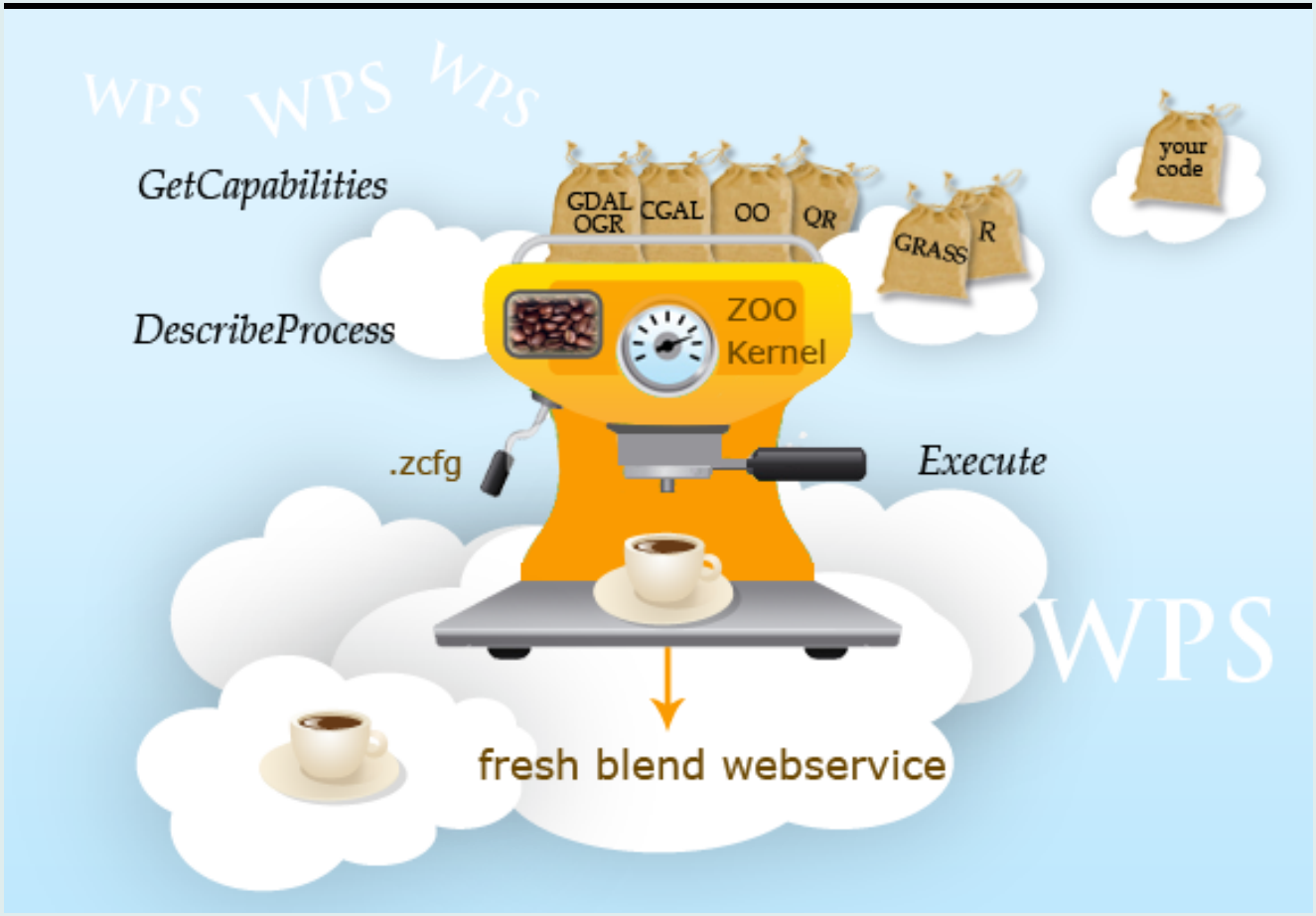
ZOO is made of three parts:

- ZOO Kernel: A powerful server-side C Kernel which makes it possible to manage and chain Web services coded in different programming languages
- ZOO Services: A growing suite of example Web services based on various Open Source libraries
- ZOO API: A server-side JavaScript API able to call and chain the ZOO Services, which makes the development and chaining processes easier

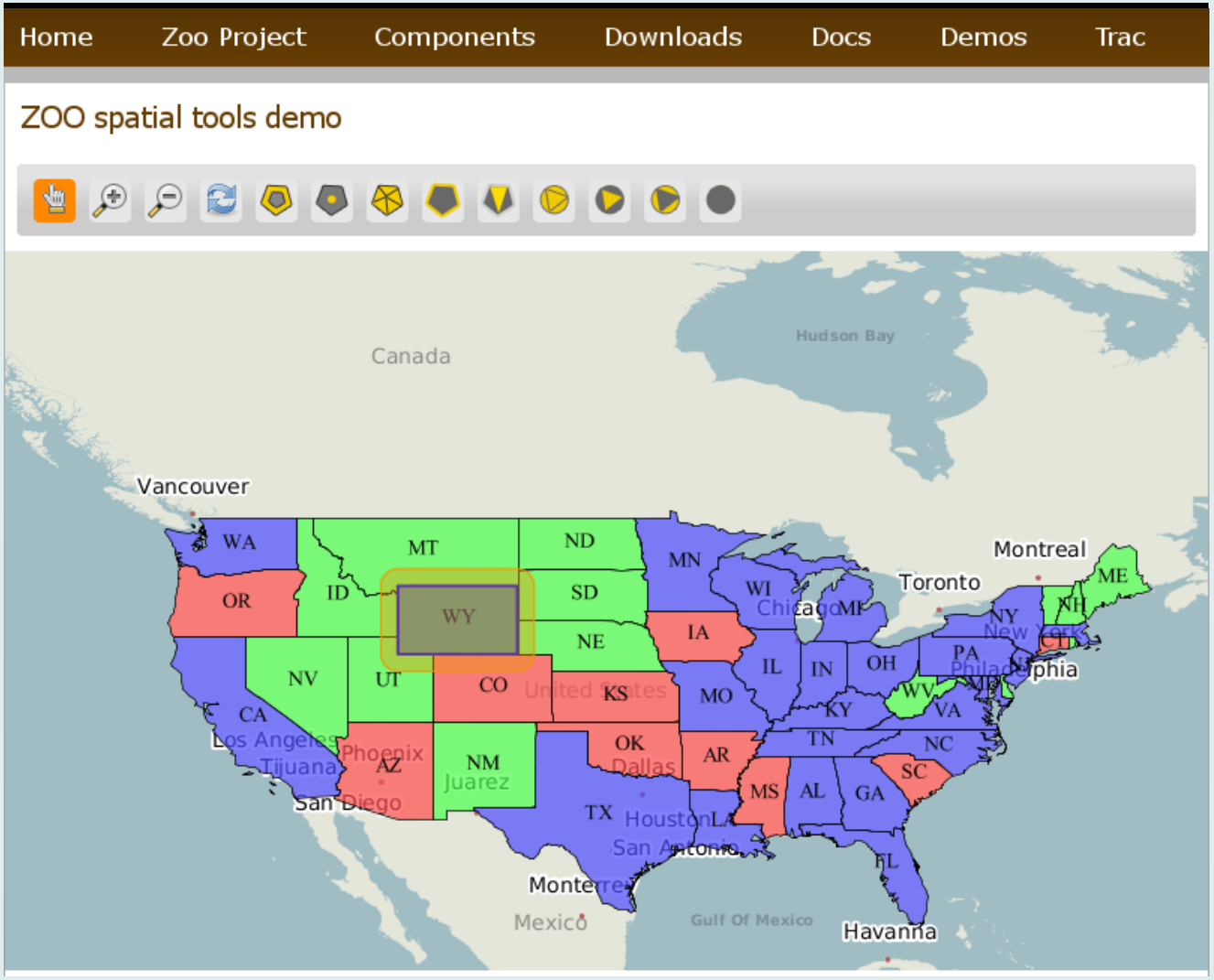
ZOO Overview



ZOO Kernel


WPS WPS WPS
GetCapabilities
DescribeProcess
.zcfg
GDAL OGR
CGAL
OO
QR
GRASS
R
your code
ZOO Kernel
Execute
WPS
fresh blend webservice

ZOO Demos


Home
Zoo Project
Components
Downloads
Docs
Demos
Trac
ZOO spatial tools demo
Canada
Hudson Bay
Vancouver
WA
OR
CA
ID
NV
MT
WY
UT
AZ
CO
NM
ND
SD
NE
KS
OK
TX
MN
IA
MO
AR
LA
WI
IL
MS
MI
IN
KY
TN
AL
OH
WV
GA
FL
SC
NC
VA
PA
NY
ME
NH
Montreal
Toronto
Chicago
Los Angeles
Tijuana
San Diego
Phoenix
Juarez
Dallas
Houston
San Antonio
Monterrey
Mexico
Gulf Of Mexico
Havanna

Rasdaman

- Rasdaman ("raster data manager") is a domain-neutral Array Database System: it extends standard relational database systems with the ability to store and retrieve multi-dimensional raster data (arrays) of unlimited size through an SQL-style query language.
- It provides reference implementation of OGC WCS and WCPS interfaces
- Rasdaman embeds itself smoothly into PostgreSQL
- The Petascope component of rasdaman provides service interfaces based on the OGC WCS, WCPS, WCS-T, and WPS

Rasdaman features

- Rasdaman makes it easy to search in large, multi-dimensional raster data
- RASQL language
- Tiling policies
- Parallel server processing
- OGC interfaces

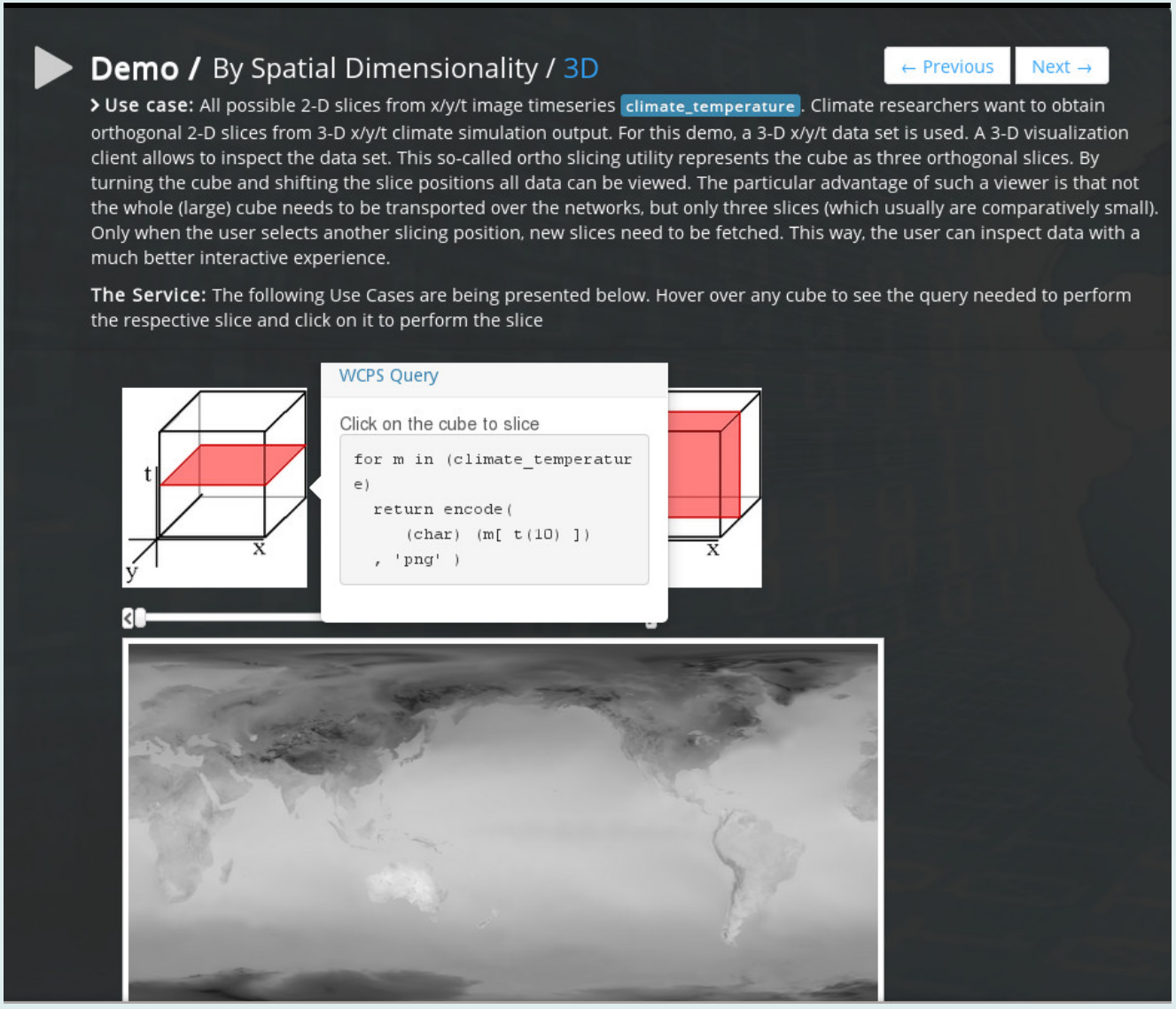
RASQL

The rasdaman query language, rasql, offers raster processing formulated through expressions over raster operations in the style of SQL.

Consider the following query: "The difference of red and green channel from all images from collection LandsatImages where somewhere in the red channel intensity exceeds 127"

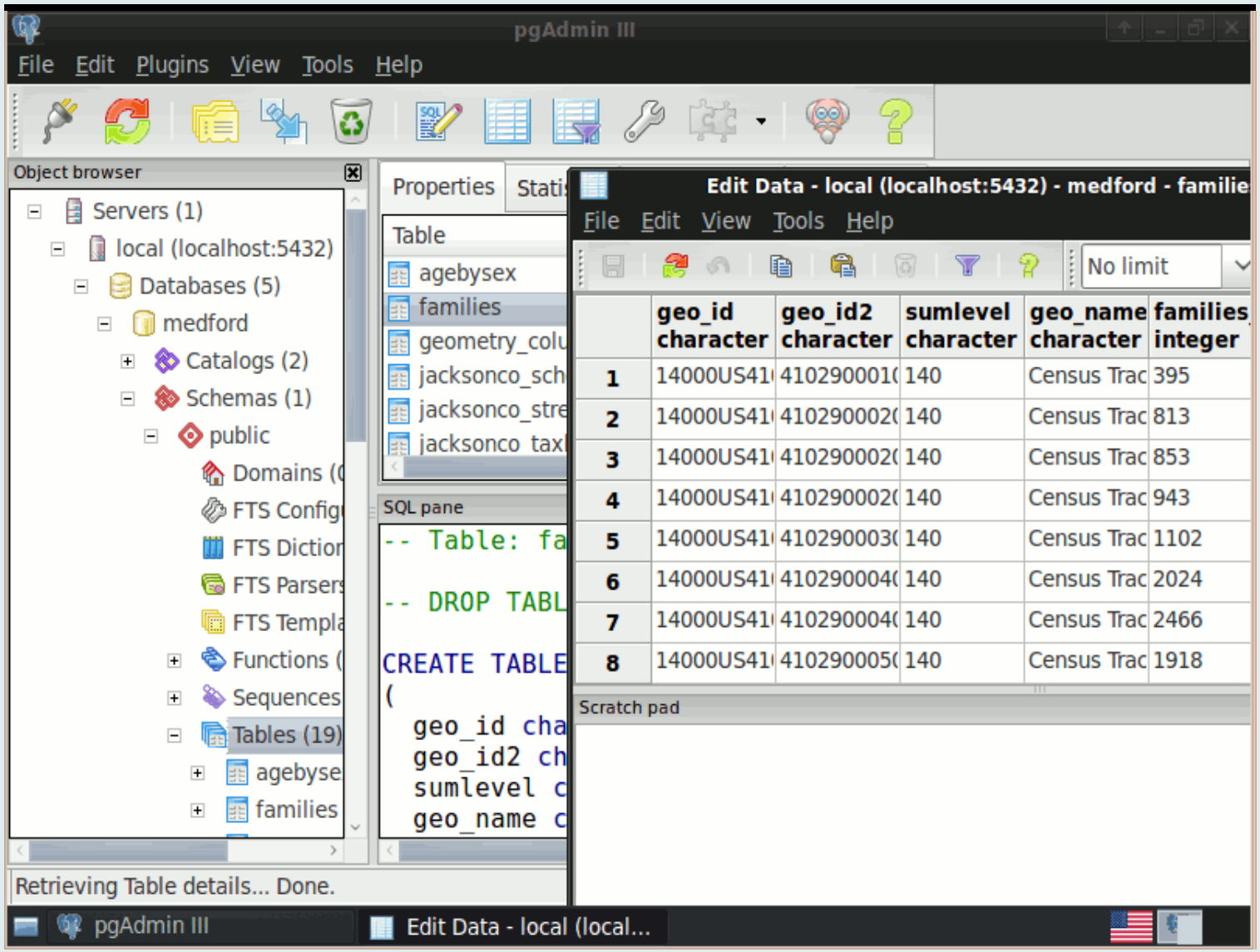
```
select ls.red - ls.green
from LandsatImages as ls
where max_cells( ls.red ) > 127
```

Rasdaman Demo

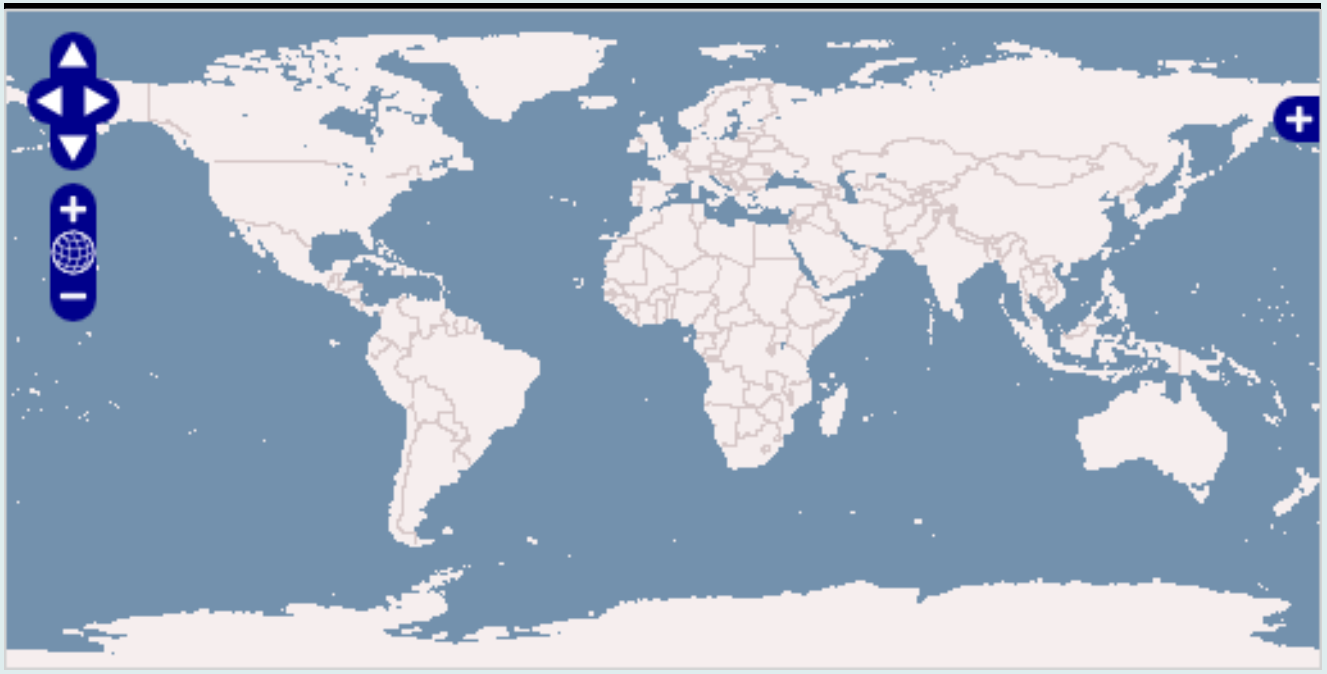


Other Geospatial Technologies Involved

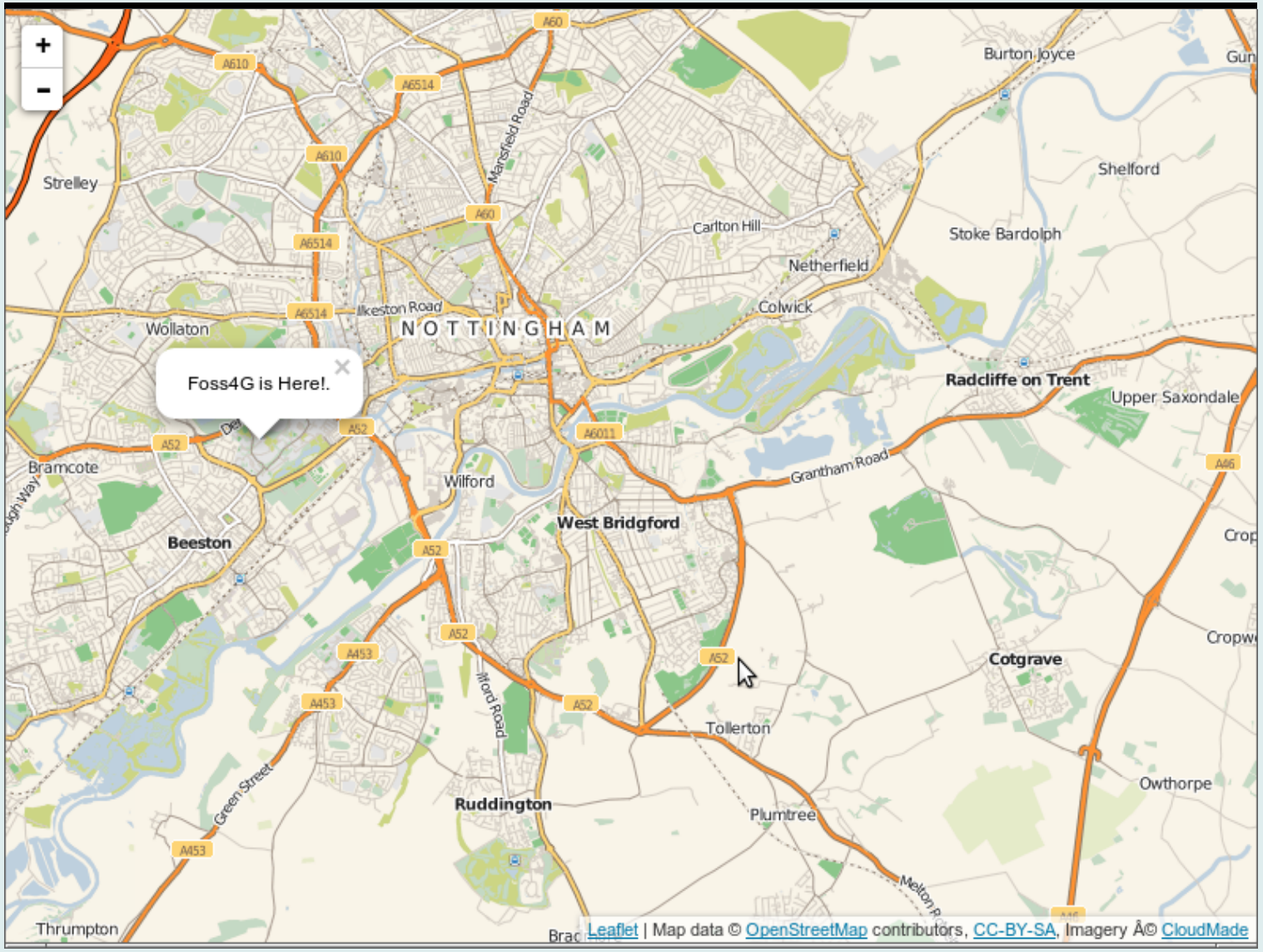
PostGIS - Spatial Database



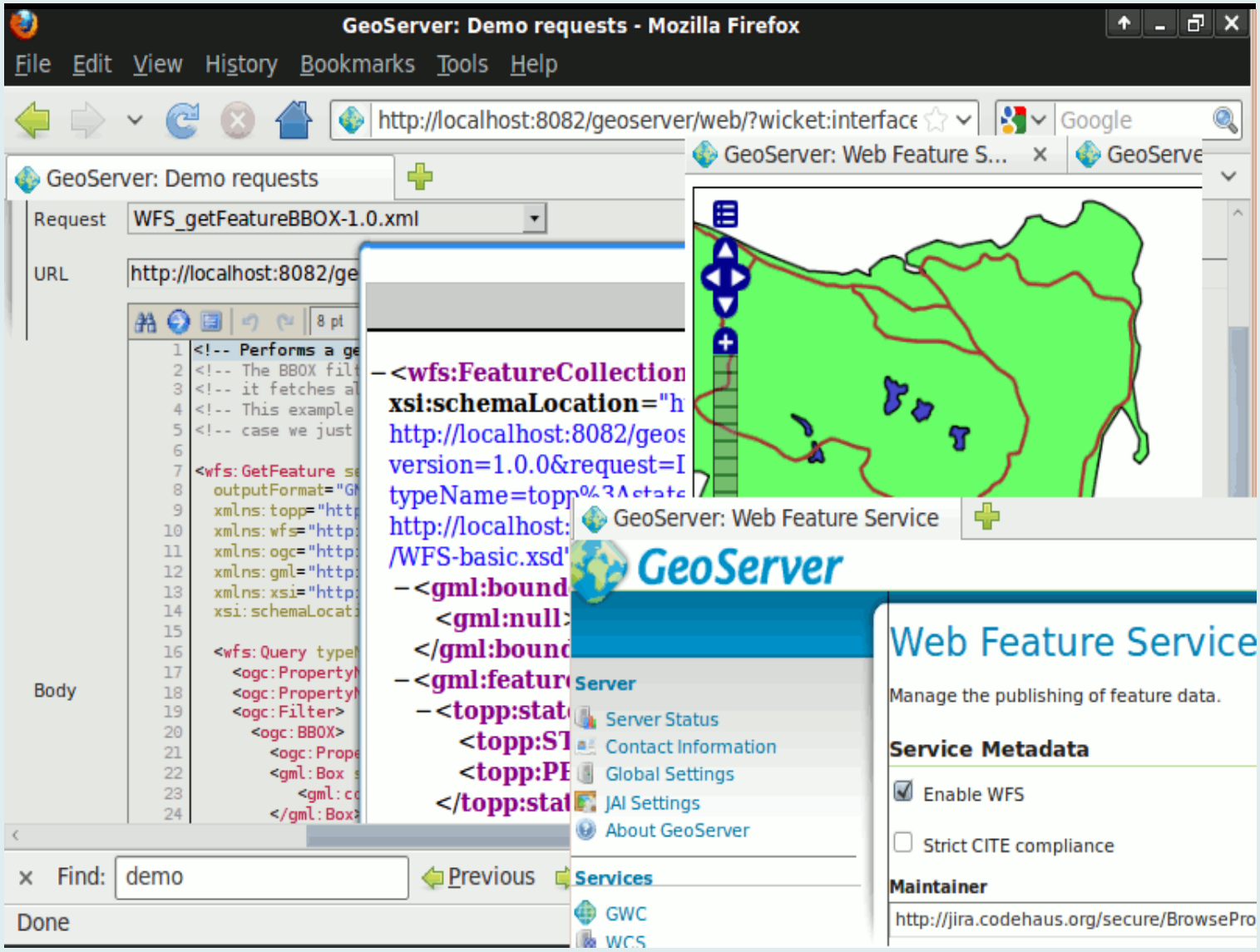
OpenLayers – Browser Mapping Library



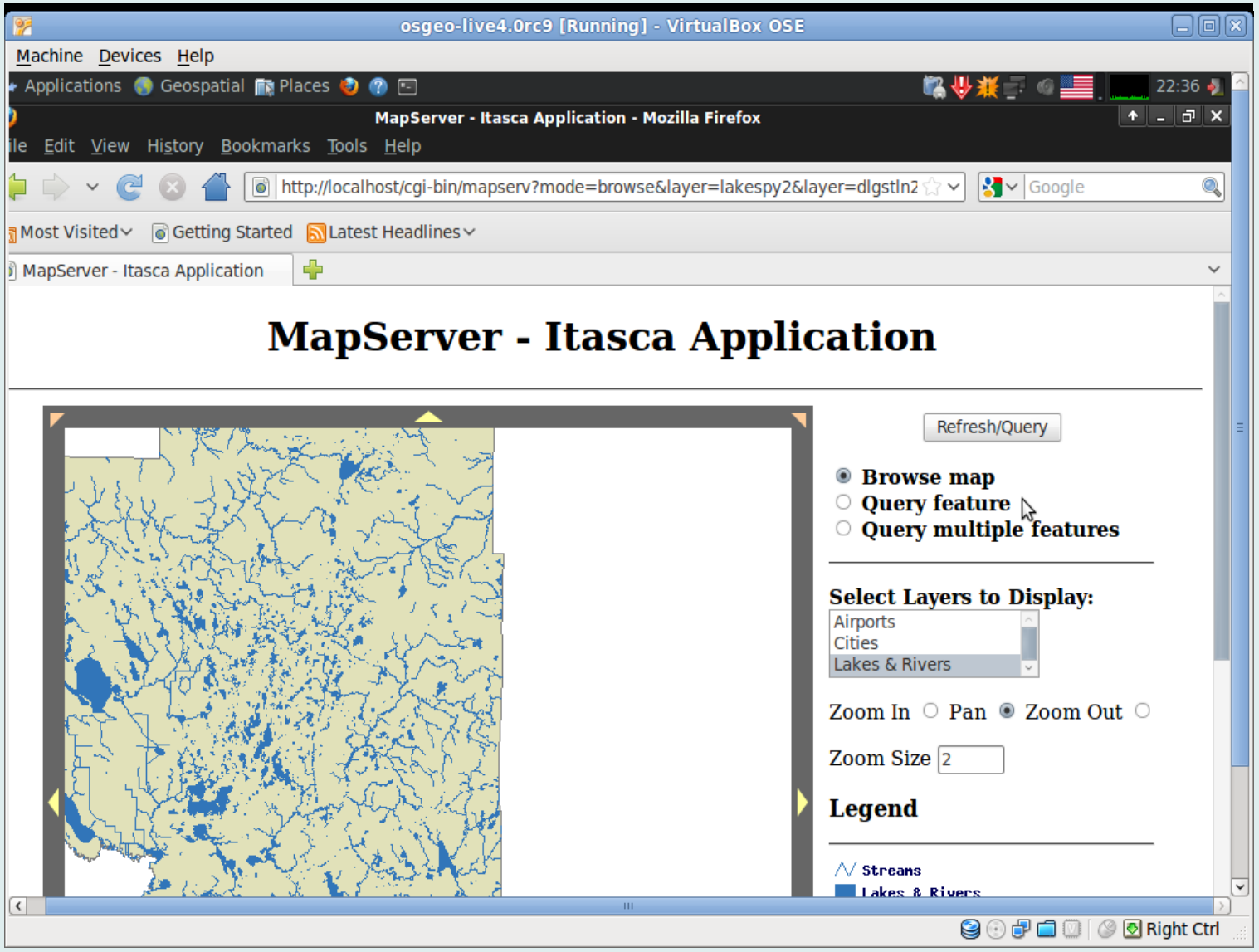
Leaflet – Mobile Friendly Interactive Maps



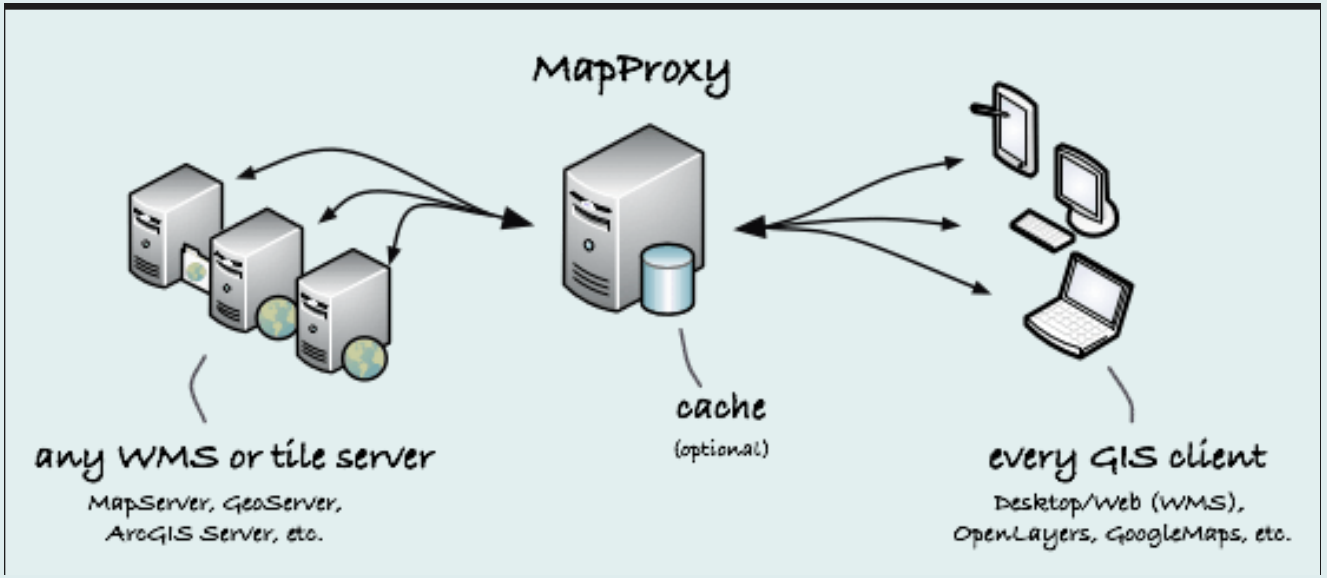
GeoServer – Web Services



MapServer – Web Services



MapProxy – Proxy WMS & tile services



GDAL/OGR – Geospatial Data Translation Tools

GDAL/OGR and PROJ.4

http://www.gdal.org

http://proj.osgeo.org

GDAL for Raster Data - The Geospatial Data Abstraction Library

Arc/Info ASCII Grid, **Arc/Info** Binary Grid (.adf), AIRSAR Polarimetric, Microsoft Windows Device Independent Bitmap (.bmp), BSB Nautical Chart Format (.kap), VTP Binary Terrain Format (.bt), CEOS (Spot for instance), First Generation USGS DOQ (.doq), DODS / OPeNDAP, New Labelled USGS DOQ (.doq), Military Elevation Data (.dt0, .dt1), **ERMapper** Compressed Wavelets (.ecw), **ESRI** .hdr Labelled, ENVI .hdr Labelled Raster, Envisat Image Product (.n1), EOSAT FAST Format, FITS (.fits), Graphics Interchange Format (**.gif**), GMT Compatible netCDF, **GRASS Rasters**, Golden Software ASCII Grid, Golden Software Binary Grid, Golden Software **Surfer** 7 Binary Grid, TIFF / **GeoTIFF** (.tif), GXF - Grid eXchange File, Hierarchical Data Format Release 4 (**HDF4**), Hierarchical Data Format Release 5 (HDF5), **Erdas Imagine** (.img), Vexcel MFF2, **Idrisi** Raster, Image Display and Analysis (WinDisp), ILWIS Raster Map (.mpr,.mpl), Japanese DEM (.mem), **JPEG** JFIF (.jpg), **JPEG2000** (JPEG2000, JP2KAK, JP2ECW, JP2MrSID), NOAA Polar Orbiter Level 1b Data Set (**AVHRR**), Erdas 7.x .LAN and .GIS, Daylon Leveller Heightfield, In Memory Raster, Vexcel MFF, Multi-resolution Seamless Image Database, Meteosat Second Generation, NDF, NITF, **NetCDF**, OGDI Bridge, PCI .aux Labelled, PCI Geomatics Database File, Portable Network Graphics (**.png**), PCRaster (.map), Netpbm (.ppm,.pgm), Swedish Grid RIK (.rik), RadarSat2 XML (product.xml), **ArcSDE** Raster, USGS SDTS DEM (*CATD.DDF), Raster Matrix Format (*.rsw, .mtw), SAR CEOS, **SGI** Image Format, USGS ASCII DEM (.dem), OGC Web Coverage Server, X11 Pixmap (.xpm)

OGR for Vector Data - Simple Feature Library

Arc/Info Binary Coverage, Comma Separated Value (**.csv**), DODS/OPeNDAP, DWG, **DXF**, ESRI Personal GeoDatabase, ESRI ArcSDE, **ESRI Shapefile**, FMEObjects Gateway, GML, **GMT Mapping**, **GRASS Vectors**, INTERLIS, **Google Earth KML**, Mapinfo File, Microstation DGN, Spatial **MySQL**, OGDI Vectors, ODBC generic database access layer, **Oracle** Spatial, PostgreSQL **PostGIS**, S-57 (ENC), SDTS, SQLite, UK .NTF, U.S. Census TIGER/Line, VRT - Virtual Datasource, Informix DataBlade

MetaCRS - Coordinate Reference System Transformations

Thank you for your attention!

Questions?

http://publicamundi.eu