

Open Geoportal Technical Introduction

Chris Barnett
Steve McDonald
Tufts University
October 2013

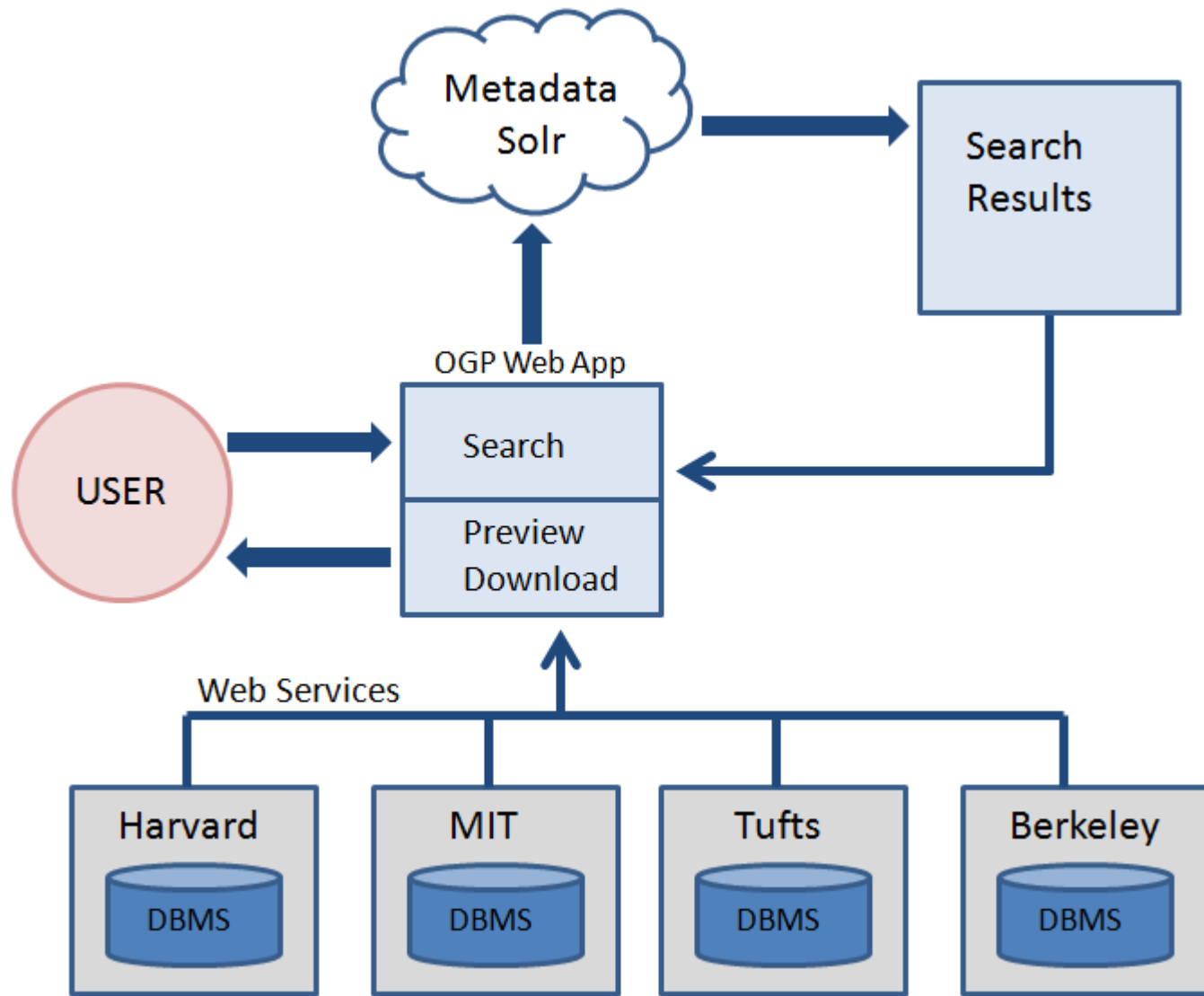
OGP Elements

Solr index

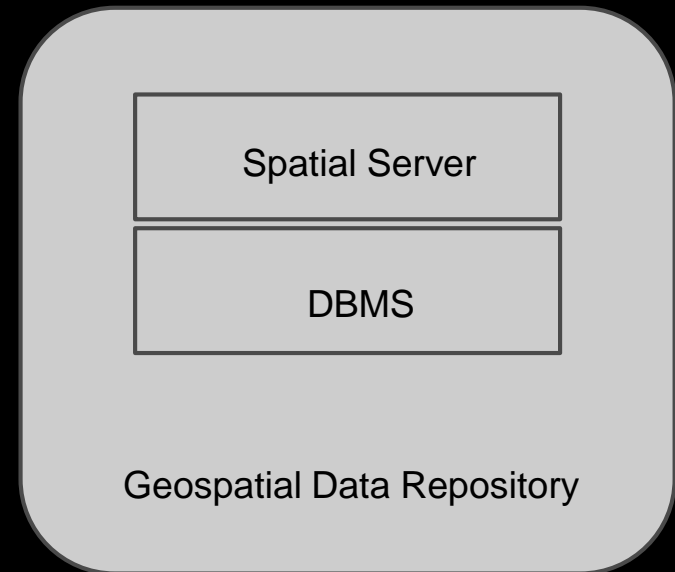
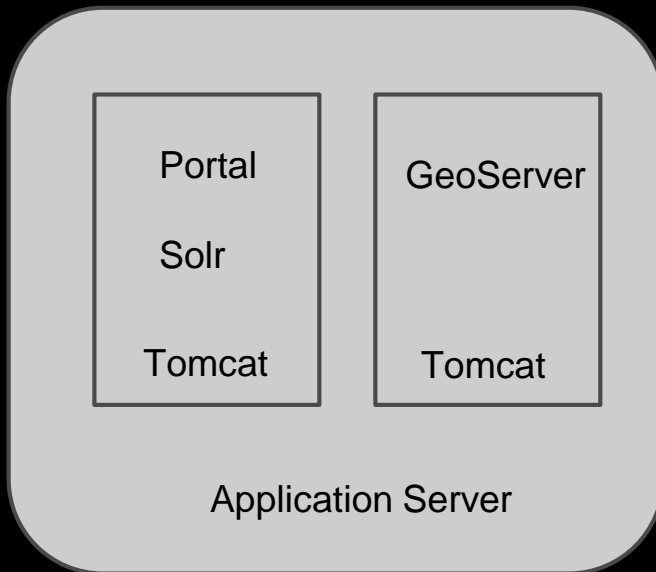
Portal Client

Ingest/Harvest

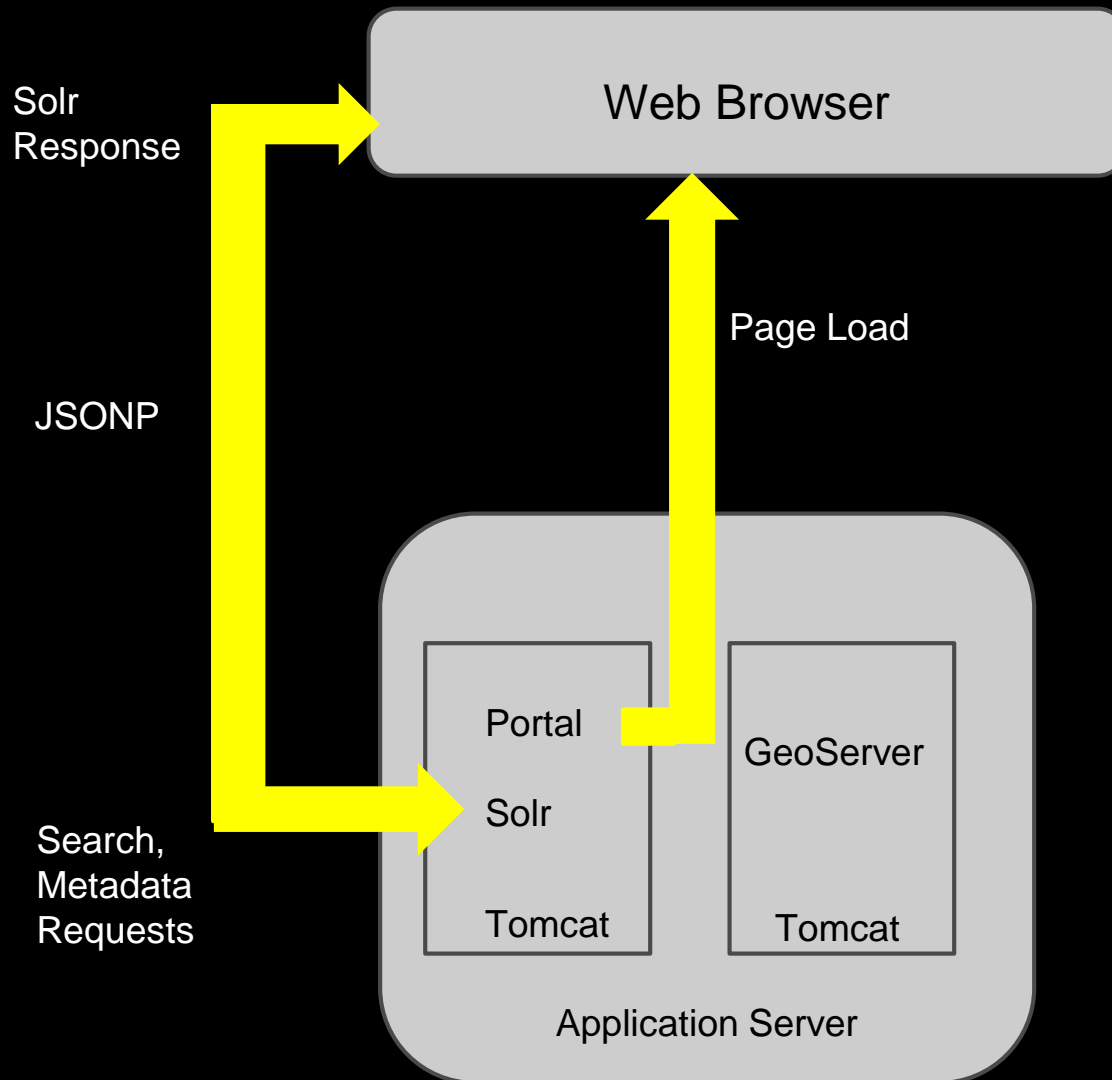
Geospatial data repository



Tufts Production Configuration



Discovery Processing



FGDC Fields

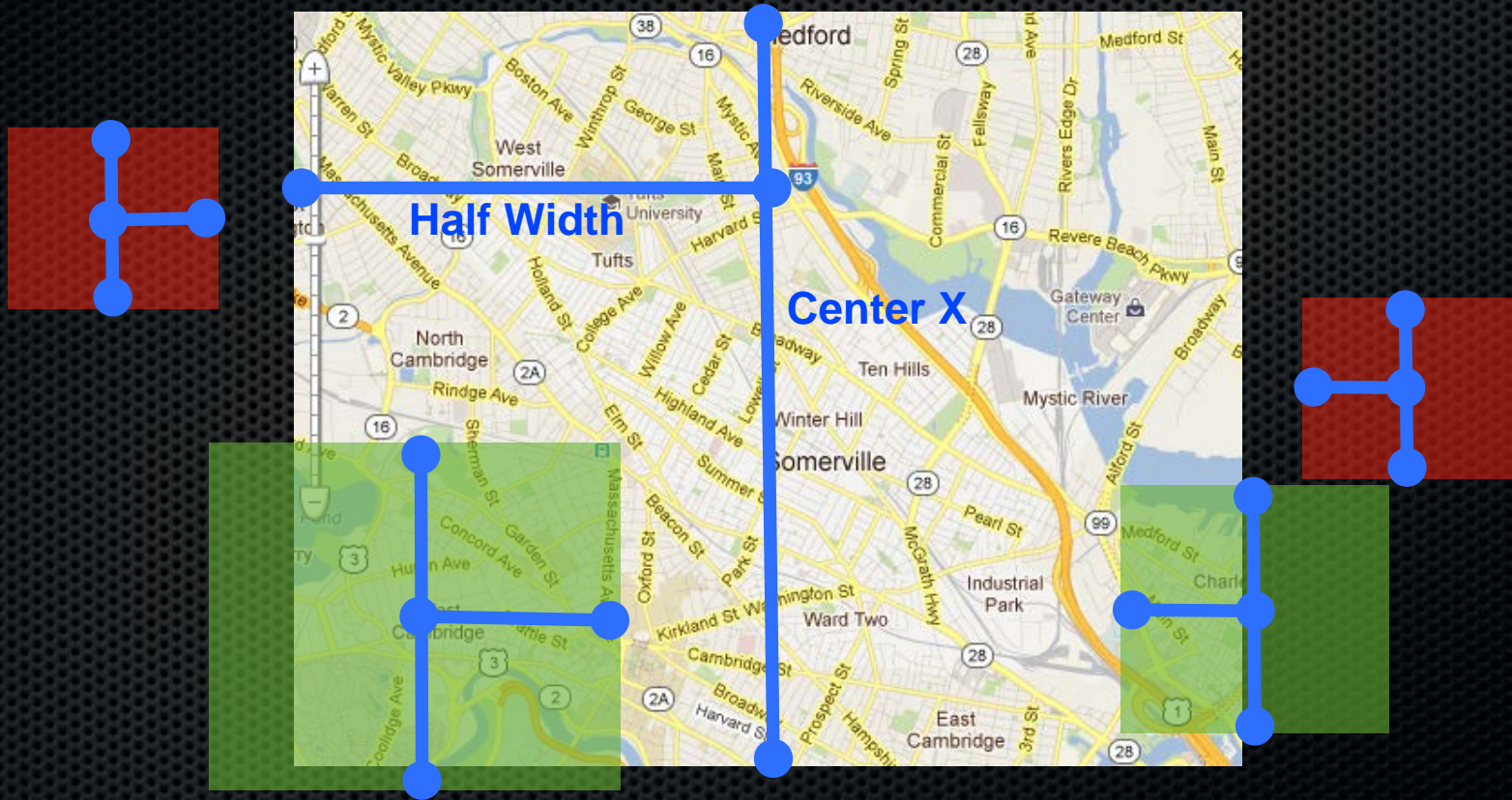
<http://code.google.com/p/opengeoportal/wiki/FgdcIngest>

Spatial Centered Search

Filter Out Layers That Do Not Intersect Map

Score Layers Based On Similarity To Map

Diff CenterXs > Sum Half Widths



But Ranking Really Matters

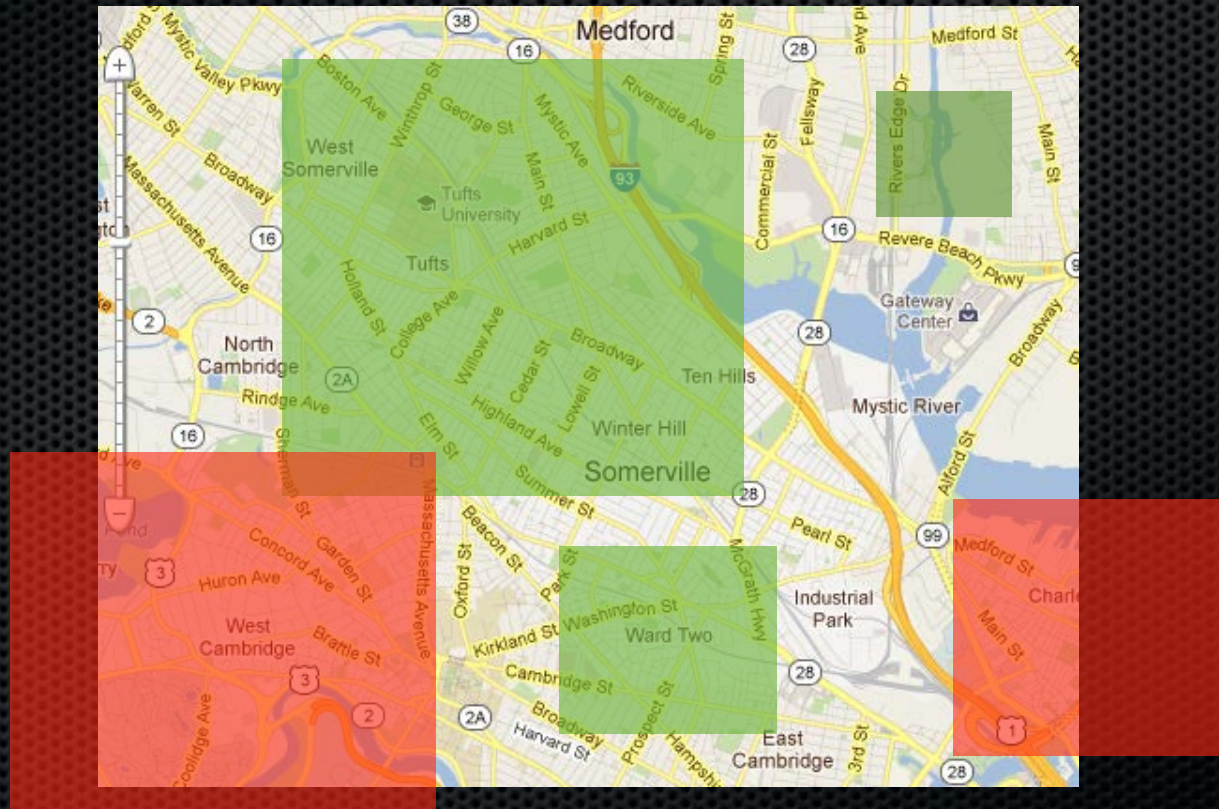
Is The Layer Completely Within The Map?

How Similar Is Layer Area To Map Area?

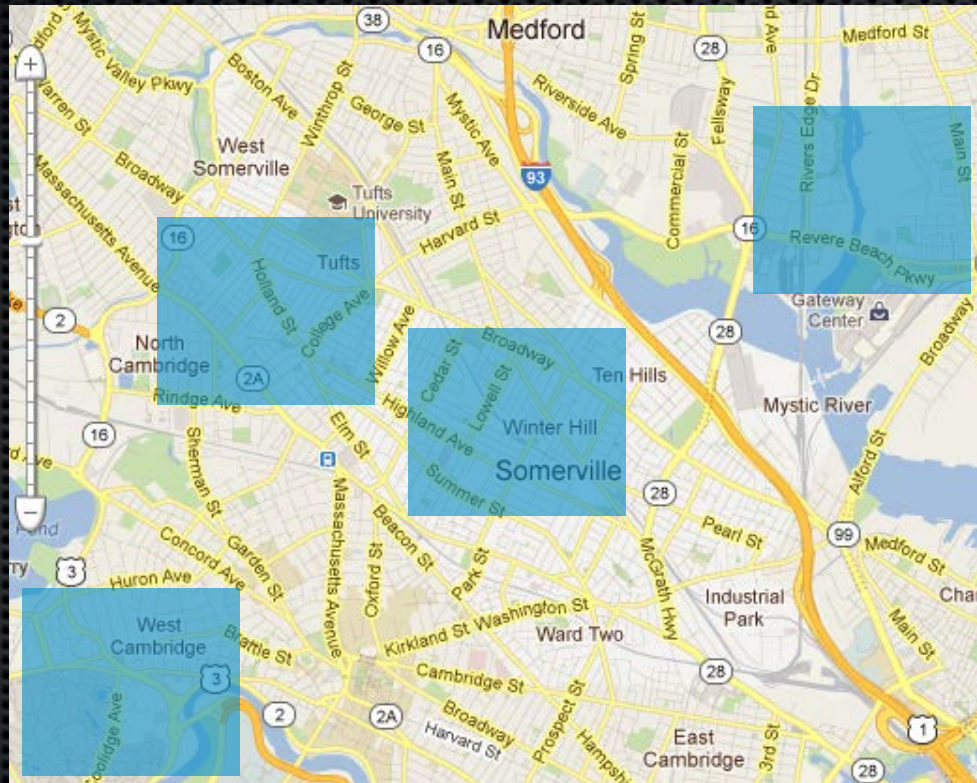
How Similar Is Layer Center To Map Center?

Does The Layer Cover Most Of The Map?

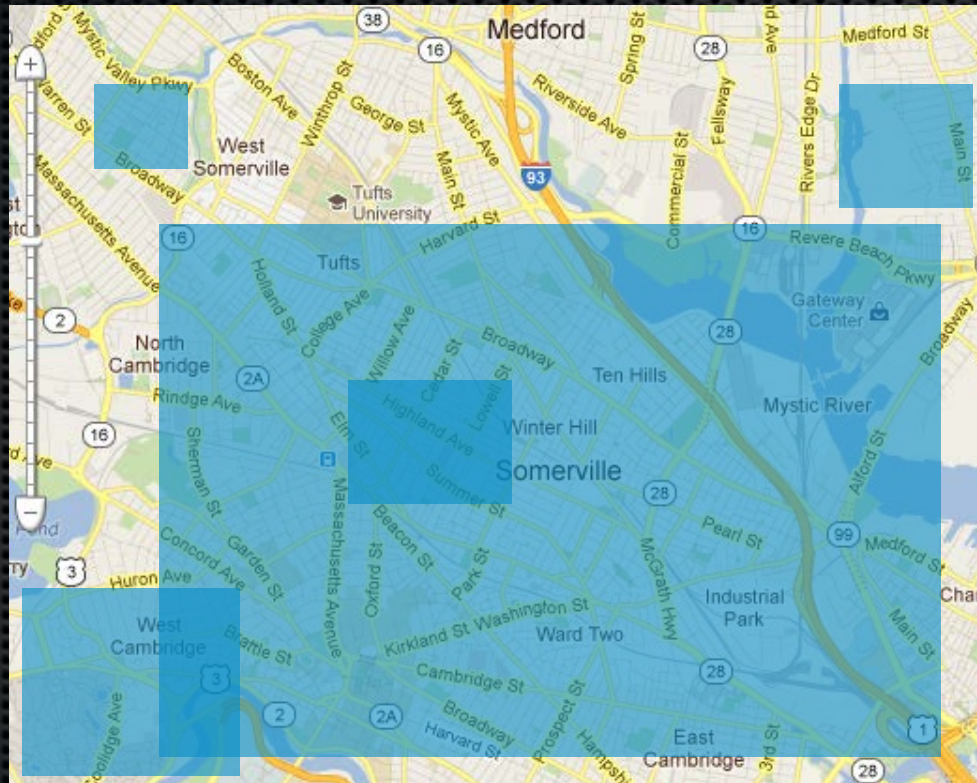
Layer Within Map



Similar Center



Similar Area



OGP Schema

.MinX, MaxX, CenterX

.MinY, MaxY, CenterY

.HalfWidth

.HalfHeight

.Area

.tdouble Field Types

Also Search By

Date

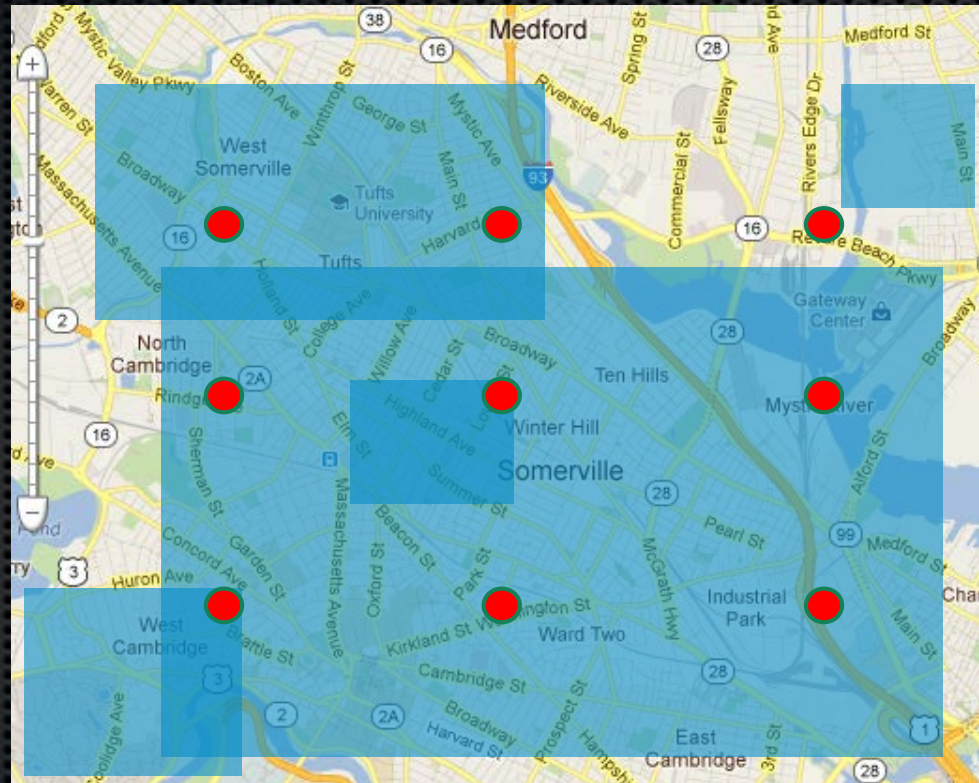
Keywords

DataType

Institution

Solr Filter Clause

Cover Area



OpenGeoPortal Query

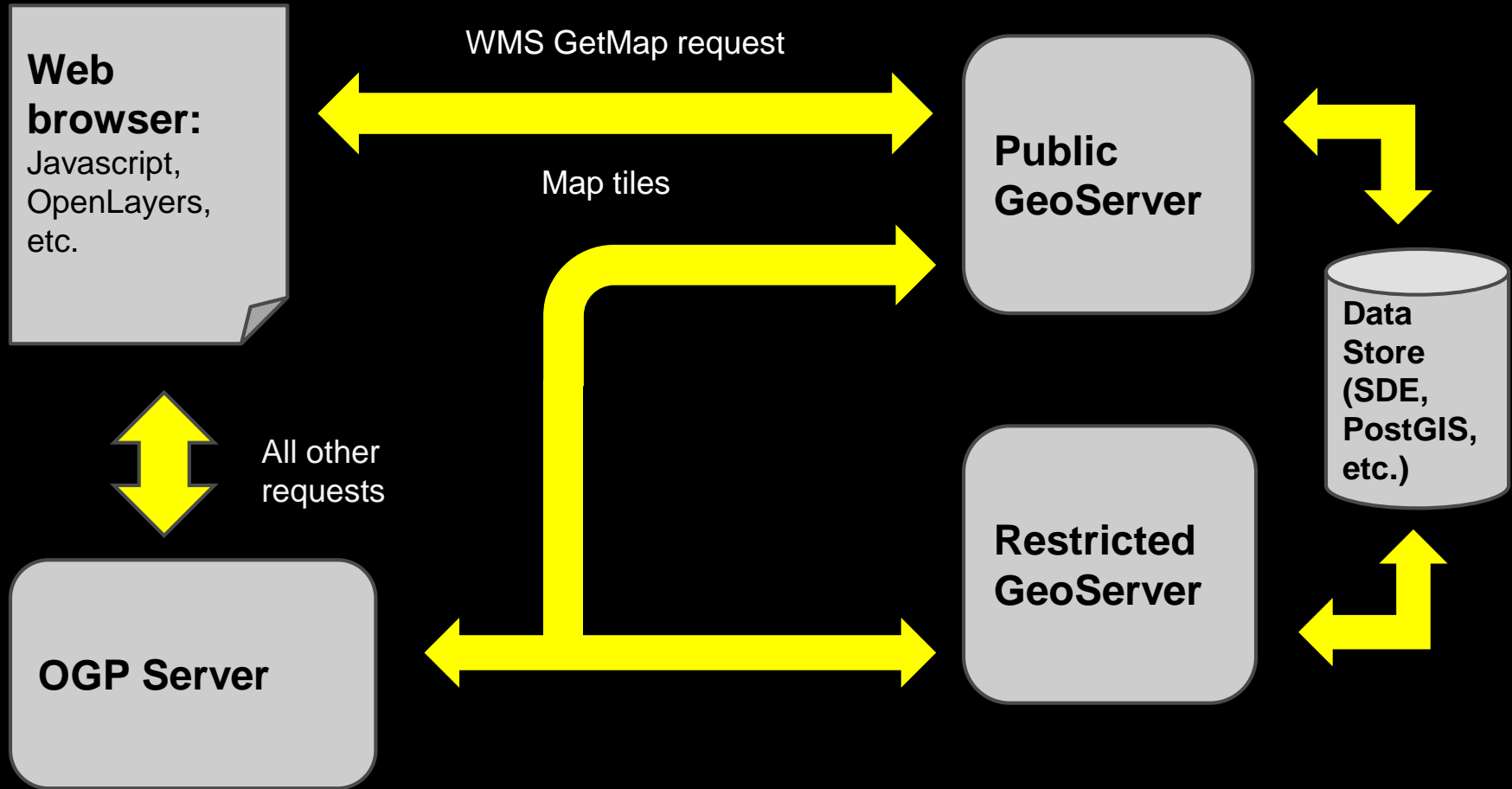
[.http://geodata.tufts.edu/solr/select?q=_val:%22product\(10.0,map\(sum\(map\(MinX,-71.143160023987,-71.096038976013,1,0\),map\(MaxX,-71.143160023987,-71.096038976013,1,0\),map\(MinY,42.385170824958,42.428266055761,1,0\),map\(MaxY,42.385170824958,42.428266055761,1,0\)\),4,4,1,0\)\)%22_val:%22product\(15.0,recip\(sum\(abs\(sub\(Area,0.002030692438118123\)\),.01\),1,1000,1000\)\)%22_val:%22product\(3.0,recip\(abs\(sub\(product\(sum\(MaxX,MinX\),.5\),-71.119599499999999\),1,1000,1000\)\)%22_val:%22product\(3.0,recip\(abs\(sub\(product\(sum\(MaxY,MinY\),.5\),42.4067184403595\),1,1000,1000\)\)%22+AND+%28LayerDisplayName:water^3+OR+ThemeKeywords:wate r^2+OR+PlaceKeywords:water^2%29+AND+%28ThemeKeywords:geoscientificinformation^4%29&&fq={!frange+l%3D1+u%3D10}product\(2.0,map\(sum\(map\(sub\(abs\(sub\(-71.119599499999999,CenterX\)\),sum\(0.023560523986994042,HalfWidth\)\),0,400000,1,0\),map\(sub\(abs\(sub\(42.4067184403595,CenterY\)\),sum\(0.021547615401498632,HalfHeight\)\),0,400000,1,0\)\),0,0,1,0\)\)&wt=json&fl =Name,CollectionId,Institution,Access,DataType,Availability,LayerDisplayName,Publisher,GeoReferenced,O riginator,Location,MinX,MaxX,MinY,MaxY,ContentDate,LayerId,score,WorkspaceName,SrsProjectionCode& rows=27&start=0&sort=score+desc&fq=ContentDate:\[1950-01-01T01:01:01Z+TO+2012-01-01T01:01:01Z\]&fq=DataType%3APoint&fq=Institution%3ATufts+OR+Institution%3AHarvard&fq=Institution:T ufts+OR+Access:Public&json.wrf=jQuery16408675794449108286_1331937717696&_=1331941365233](http://geodata.tufts.edu/solr/select?q=_val:%22product(10.0,map(sum(map(MinX,-71.143160023987,-71.096038976013,1,0),map(MaxX,-71.143160023987,-71.096038976013,1,0),map(MinY,42.385170824958,42.428266055761,1,0),map(MaxY,42.385170824958,42.428266055761,1,0)),4,4,1,0))%22_val:%22product(15.0,recip(sum(abs(sub(Area,0.002030692438118123)),.01),1,1000,1000))%22_val:%22product(3.0,recip(abs(sub(product(sum(MaxX,MinX),.5),-71.119599499999999),1,1000,1000))%22_val:%22product(3.0,recip(abs(sub(product(sum(MaxY,MinY),.5),42.4067184403595),1,1000,1000))%22+AND+%28LayerDisplayName:water^3+OR+ThemeKeywords:wate r^2+OR+PlaceKeywords:water^2%29+AND+%28ThemeKeywords:geoscientificinformation^4%29&&fq={!frange+l%3D1+u%3D10}product(2.0,map(sum(map(sub(abs(sub(-71.119599499999999,CenterX)),sum(0.023560523986994042,HalfWidth)),0,400000,1,0),map(sub(abs(sub(42.4067184403595,CenterY)),sum(0.021547615401498632,HalfHeight)),0,400000,1,0)),0,0,1,0))&wt=json&fl =Name,CollectionId,Institution,Access,DataType,Availability,LayerDisplayName,Publisher,GeoReferenced,O riginator,Location,MinX,MaxX,MinY,MaxY,ContentDate,LayerId,score,WorkspaceName,SrsProjectionCode& rows=27&start=0&sort=score+desc&fq=ContentDate:[1950-01-01T01:01:01Z+TO+2012-01-01T01:01:01Z]&fq=DataType%3APoint&fq=Institution%3ATufts+OR+Institution%3AHarvard&fq=Institution:T ufts+OR+Access:Public&json.wrf=jQuery16408675794449108286_1331937717696&_=1331941365233)

Open Geoportal: a front end to the
geospatial data index

Solr query is controlled by

- Search fields
- Spatial extent
- Display table

Preview And Download Processing



Preview of layers

Primarily through OGC standard web services, which are specified in the Solr index

Preview of layers

WMS GetMap:

Rasterized, tiled representation of a data set, indexed to a well-known grid & reprojected to a common projection.

Requests are constructed by OpenLayers from fields in the Solr index.

Preview of layers

WMS GetFeatureInfo: preview of attribute data, returned as an html table. In OGP, the table is parsed and styled for common presentation. Additionally, for layers with FGDC metadata, attribute definitions are retrieved from the metadata document to facilitate discovery.

Layer Download

A request is made to the download controller by the web client with :

Layer Ids, requested formats,
bounds

Layer Download

The download controller queries Solr for layer info.

If applicable, WMS

DescribeLayer or

“serviceStart” link may be called.

Layer Download

Based on the returned data and the contents of *ogpDownloadConfig.json*, a download method is chosen.

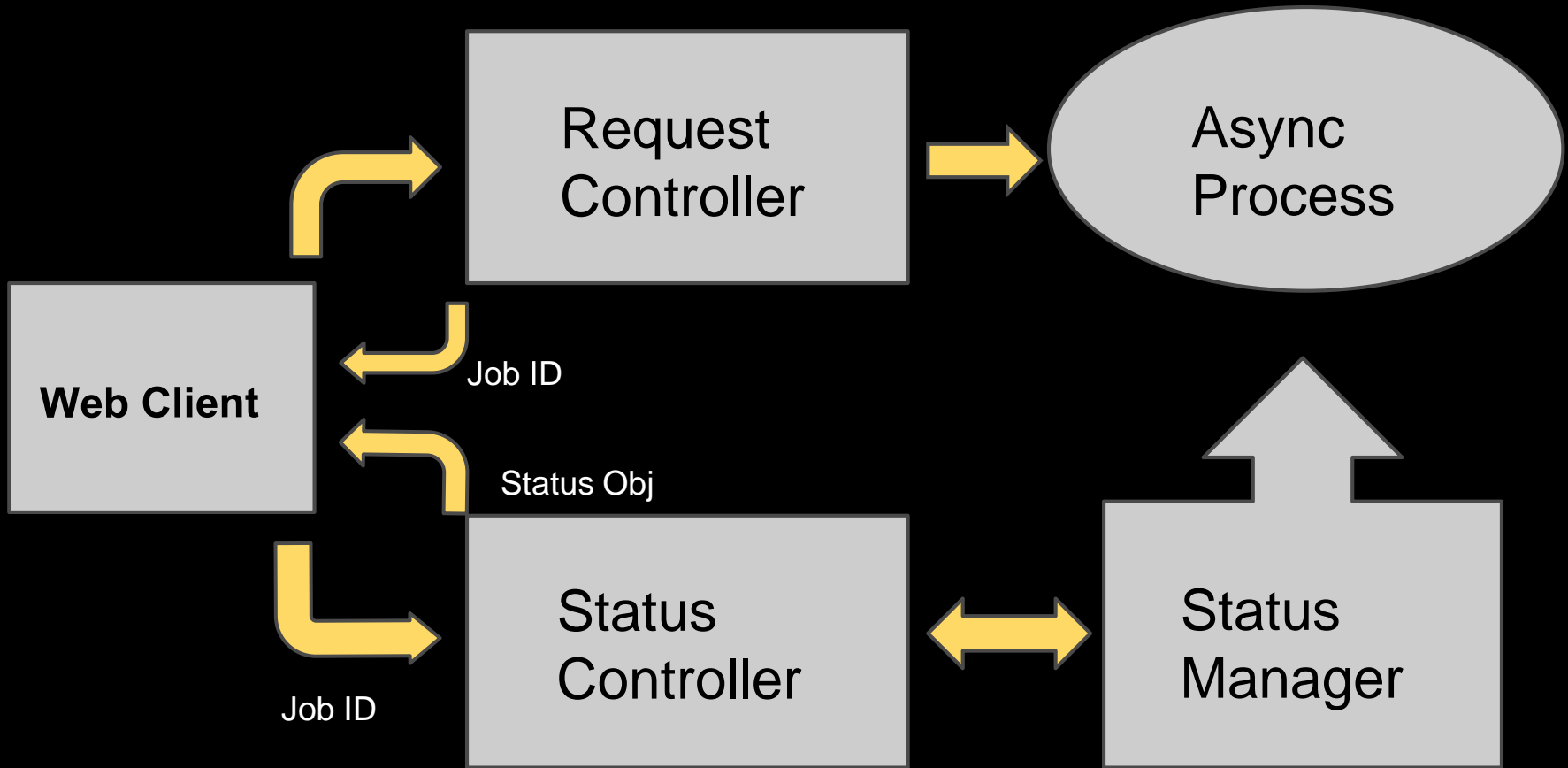
ogpDownloadConfig.json

```
"config":{
  "institutions":{
    "tufts":[
      {"classKey": "layerDownloader.wfs.proxied",
        "accessLevel":["restricted"],"dataType":["vector"],
        "outputFormats":["shp"]},
      {"classKey": "layerDownloader.wcs.proxied",
        "accessLevel":["restricted"],"dataType":["raster", "paper map"],
        "outputFormats":["geotiff"]}],
    "harvard":[
      {"classKey": "layerDownloader.email.HGL",
        "accessLevel":["public","restricted"],"dataType":["raster","paper map"],
        "outputFormats":["geotiff"]}],
    "default":
      [{"classKey": "layerDownloader.wfs",
        "accessLevel":["public"],"dataType":["vector"],
        "outputFormats":["shp"], "preference": 1},
      {"classKey": "layerDownloader.wcs",
        "accessLevel":["public"],"dataType":["raster", "paper map"],
        "outputFormats":["geotiff"], "preference": 1},
      {"classKey": "layerDownloader.wms", "accessLevel":["public"],"dataType":
```

Layer Download

A request status object is registered to send info to the user and asynchronous processes are spawned to perform the actual downloads concurrently.

Asynchronous Processing



Spring Framework

- Java Framework with a large user community and support base
- Used by GeoServer and many others
- Proven and tested
- A framework for dependency injection, XML & JSON data binding, MVC, thread pooling, etc. that allows rapid and reliable development of code

Dependency Injection

Write code to an interface and specify implementation in an xml config file. (or annotation)

allows modular and flexible design (ex. download)

from *applicationContext.xml*

```
<!-- LayerDownloader -->
<beans:bean id="layerDownloader.wfs"
  class="org.opengeoportal.download.PerLayerDownloader" scope="prototype">
  <beans:property name="perLayerDownloadMethod" ref="downloadMethod.wfs" />
</beans:bean>
<beans:bean id="layerDownloader.wcs"
  class="org.opengeoportal.download.PerLayerDownloader" scope="prototype">
  <beans:property name="perLayerDownloadMethod" ref="downloadMethod.wcs" />
</beans:bean>
<beans:bean id="layerDownloader.wfs.proxied"
  class="org.opengeoportal.download.PerLayerDownloader" scope="prototype">
  <beans:property name="perLayerDownloadMethod" ref="downloadMethod.wfs.proxied" />
</beans:bean>
```

Model-View-Controller (MVC)

MVC offers extensibility and flexibility

A new service endpoint that builds on existing code can be added easily (by annotation)

The controller itself simply returns a data object. How it is rendered is determined by a separate view, which can be changed and customized without touching/breaking known-good business logic. (ex.: spreadsheet view from ingest)

JavaScript / Client

Libraries:

JQuery, JQueryUI, Backbone.js, Underscore,
DataTables, OpenLayers

Freedom of design, Open Source, active user
communities

Object Oriented

Separation of design from content

OpenGeoIngest

A Separate GitHub Project:

<https://github.com/OpenGeoportal/ogpIngest>

Prepares And Ingests Data

Solr, Geoserver

6 JSPs, 11000 Lines Of Java

Supports Batch And Web Forms

Local Layers

Web based metadata file upload

Ingests layer metadata into Solr

- Verifies Solr compatibility

Publishes layer in GeoServer

Records ingest event in SQL database

Ingest Via Directed Crawl

User Provides URL

URL And All Child Pages Searched

Links To Xml And Zip Files Processed

Planned:

Page Scraping To Augment Metadata

Wrap Data In Service



INGEST ACTIONS

[Preprocess Metadata](#)

[Upload Metadata](#)

[Ingest Records from Remote Solr Instance](#)

[Delete Solr Layers](#)

Metadata Ingest

Use this page to ingest FGDC metadata files into GeoServer (Local layers only) and Solr.

Institution (select one)

- Ingest to GeoServer and Solr Ingest to GeoServer only Ingest to Solr only

Note: This option only applies to local layers. Ingest will not attempt to configure GeoServer for remote layers.

Required fields

- Title Publisher Originator Abstract Data Type
 Theme Keywords Place Keywords Content Date Bounds Access

Note: If you require a field, Solr ingest will fail for a data layer if the metadata does not contain a valid value. If you do not require the field, a warning will still be produced for invalid values.

Specify a URL to crawl:

(for example, <http://cod.humanitarianresponse.info/>)

Or, add XML metadata file(s) or zipped directory of XML metadata files

[+ Add files...](#)

[Start ingest](#)

Ingest Via Undirected Crawl

Crawl Entire Web For Spatial Data

Use CommonCrawl.org Data

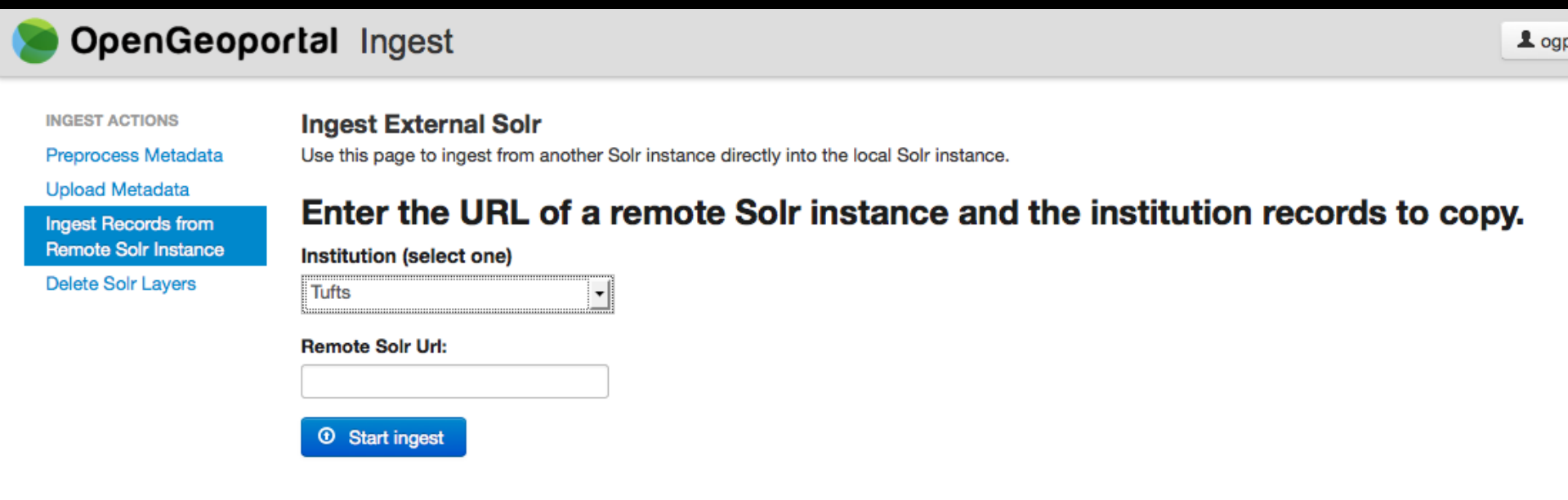
Run Hadoop Job On AWS Cluster

Easily Ingest Shape, KML/KMZ Files

Remote Layers

Solr Slave To Tufts

Copy Layers Solr To Solr



The screenshot shows the OpenGeoportal Ingest interface. At the top left is the OpenGeoportal logo and the text 'OpenGeoportal Ingest'. At the top right is a user profile icon labeled 'ogp'. On the left side, there is a vertical menu under the heading 'INGEST ACTIONS' with the following items: 'Preprocess Metadata', 'Upload Metadata', 'Ingest Records from Remote Solr Instance' (highlighted in blue), and 'Delete Solr Layers'. The main content area is titled 'Ingest External Solr' and includes the instruction: 'Use this page to ingest from another Solr instance directly into the local Solr instance.' Below this is a large bold instruction: 'Enter the URL of a remote Solr instance and the institution records to copy.' There is a dropdown menu for 'Institution (select one)' with 'Tufts' selected. Below that is a text input field for 'Remote Solr Url:'. At the bottom is a blue button with a play icon and the text 'Start ingest'.

OpenGeoportal Ingest

INGEST ACTIONS

- Preprocess Metadata
- Upload Metadata
- Ingest Records from Remote Solr Instance**
- Delete Solr Layers

Ingest External Solr

Use this page to ingest from another Solr instance directly into the local Solr instance.

Enter the URL of a remote Solr instance and the institution records to copy.

Institution (select one)

Tufts

Remote Solr Url:

Start ingest

Solr Cloud Issues

Federated Search

The Easy Way: Same Schema

Searching Shards

Local Solr Slave Of Cloud

Local Solr For Private/New Data

DeDuplication

A Problem

Questions?

Chris Barnett
Steve McDonald
Tufts University