Open Source GIS

presented by Adam Estrada
Agenda

- About me
- Background: Why Open Source?
- Review of Open Source Solutions
- Open Source Data
- Applications of Open Source Data
Background: Why Open Source!?!
The mix…

- GDAL
- OpenLayers
- OpenStreetMap
- OpenGeo
- python
- TileMill
- PostGIS
- MapServer
- GeoTools
- GeoServer
- MapProxy
- OSGeo
- GEOS

Comprehensive list: http://opensourcegis.org/
Why Open Source?

Highlight

Frank Warmerdam:
• ‘91 to ’98 - PCI Geomatics
• ‘98 to present - GDAL
• Co-founder of OSGeo
• 2011 to present Google

Commercial software using GDAL:
• ESRI
• ERDAS
• FME Workstation
• IDRISI
• Google Earth
• TerraGo
• Many, many more…
Why Open Source?

- **OSSIM**: Developed by ImageLinks, Inc. for the NRO
  - Licensed under GPL in 1999

- **GRASS**: Developed by USACOE
  - Licensed under GPL in 1999

- **Opticks**: Developed by Ball Aerospace
  - Licensed under GPL in 2002

- **Quantum GIS**: Developed by Gary Sherman
  - Licensed under GPL in 2002
  - Over 200 developers today!
Why Open Source?

Crowdsourcing your Centerlines!
The Open Source Community

“... given enough eyeballs, all bugs are shallow.”

Linus's Law
as described by Eric Raymond
The Open Source Community

“... given enough eyeballs, all bugs are shallow.”

Linus's Law

as described by Eric Raymond

Project Bug Tracking (eg. TRAC)

StackOverflow

OpenGeo

Internet Relay Chat (IRC)

StackExchange

GIS

Mailing Lists (Listserv)
The Open Source Community

**GDAL FGDB Domain Error**
Fixed in less than a week
http://trac.osgeo.org/gdal/ticket/4455

**Project Bug Tracking**
(eg. TRAC)

**New Driver**
Added to GDAL in 1 day
http://trac.osgeo.org/gdal/ticket/4478

*Linus's Law*
as described by Eric Raymond

*...given enough eyeballs, all bugs are shallow.*

Efficient Geospatial Solutions
Software Stack

Desktop

Database  Web Client

Web Mapping Services

Geometry Support

File Access
Software Stack

- Geometry Support
- Web Mapping Services
- Database
- Web Client
- Desktop

File Access

Efficient Geospatial Solutions
Software Stack

Desktop

Database

Web Client

Web Mapping Services

JTS Topology Suite

GEOS

Geometry Engine
Open Source

GeoTools

Geometry Support

GDAL

File Access

PROJ.4
Software Stack

Desktop

Database

Web Client

Web Mapping Services

JTS Topology Suite

GEOS

Geometry Engine
Open Source

GeoTools

Geometry Support

GDAL

File Access

Efficient Geospatial Solutions

Use, duplication or disclosure of this document or any of the information contained herein is subject to the restrictions on the title page of this document.
Software Stack

Desktop

Web Client

Database

Web Mapping Services

JTS Topology Suite

GEOS

Geometry Engine

Open Source

GeoTools

Geometry Support

GDAL

File Access

PROJ.4

Efficient Geospatial Solutions

Use, duplication or disclosure of this document or any of the information contained herein is subject to the restrictions on the title page of this document.
Software Stack

Desktop

Database

Web Client

Web Mapping Services

JTS Topology Suite

GEOS

Geometry Engine
Open Source

GeoServer

GeoTools

Geometry Support

GDAL

PROJ.4

File Access
Software Stack

Desktop

Database

Web Client

Web Mapping Services

JTS Topology Suite

GEOS

Geometry Engine
Open Source

GeoServer

GeoTools

Geometry Support

GDAL

PROJ.4

File Access
Interoperability

COTS

- esri
- erdas
- OVERWATCH

Standards

- WMS
- WFS
- WCS
- WPS
- KML
- Etc…

Open Source

- GeoServer
- MapServer
- OpenLayers
- GDAL
- OSSIM
- PostGIS
Emerging Technologies
“Big Data”
Hadoop™ Workflow

Large Dataset

Node 1
- a
- c
- e
- g
- i
- k
- b
- d
- f
- h
- j
- l

Node 2
- a
- c
- e
- g
- b
- d
- f
- h

Asynchronous process...

PostsGIS

Spatial PostgreSQL

Efficient Geospatial Solutions
Licensing Options

License Types
Apache License, 2.0 (Apache-2.0)
MIT license (MIT)
BSD 3-Clause “New” or “Revised” license (BSD-3-Clause)
BSD 3-Clause "Simplified" or "FreeBSD" license (BSD-2-Clause)
GNU General Public License (GPL)
GNU Library or "Lesser" General Public License (LGPL)

Creative Commons Attribution
Open Source Data!
# Data Sources

## Geotagged Photos
- Panoramio
- Flickr
- Picasa
- Locr
- YouTube
- Instagram
- Many more…

## Social Media
- Twitter
- Foursquare
- Orkut
- Many more…

## Open Web Harvesting
- Blogs
- Wikis
- Forums
- News site
  - CNN
  - Reuters
  - Al Jazeera
  - Etc…

## GIS Data
- OpenStreetMap
- GeoCommons/GeoIQ
- Wikimapia
- ArcGIS Online
Web Harvester

- Crowdsourced Data
- Open Source Harvest
- Microblog Data
- Geotagged Photo Data
- Crawl Engine
Data Store

- Continuous updates
- Persistent storage/archive
- Schema definition
- Geometry validation
- Taxonomy database
  - Categorizing
  - Building relevance ranking
  - Term boosting
  - Entity extraction
Translate

"Rupana रुपाणा"
"Rupana village"

GOSSIP Store

Database

Translation
Indexing the Data Store
Why Index?

- Faceted search and filtering
- Term vectoring and tokenization
- Dynamic clustering based on
  - Topically related content
  - Frequent itemset mining
  - Etc…
- More like this…functionality
- Spatial relevancy
{ Analysis }
Working with Twitter

- Find top n users for an area
- Find most influential users
- Find similarities between Tweets
  - Jaccard Indices
  - Used to find similarities between sample sets
  - Useful for detecting copies, common topics in text corpus and making recommendations
- Output is a score between (0-1)
But what does it mean!?!?

1. Sally sells sea shells by the sea shore.
   2. Sally is at the sea shore.

The computer sees this…
B. [“sally”, “sea”, “shore”]

Result = 0.55555556 or 56% similar

NOTE: Could get better results with better machine translation!
<table>
<thead>
<tr>
<th></th>
<th>168691083033849856</th>
<th>168682158339260418</th>
<th>0.1111111111111111</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>168668394739085312</td>
<td>168668394739085312</td>
<td>0.2857142857142857</td>
</tr>
<tr>
<td>7</td>
<td>168744891147489280</td>
<td>168744887427145728</td>
<td>0.14285714285714285</td>
</tr>
<tr>
<td>8</td>
<td>168744891147489280</td>
<td>168691057897385984</td>
<td>0.125</td>
</tr>
<tr>
<td>9</td>
<td>168744891147489280</td>
<td>168738096945446912</td>
<td>0.1111111111111111</td>
</tr>
<tr>
<td>10</td>
<td>168744891147489280</td>
<td>168738072740114433</td>
<td>0.1111111111111111</td>
</tr>
<tr>
<td>11</td>
<td>168744891147489280</td>
<td>168744736163758080</td>
<td>0.1111111111111111</td>
</tr>
<tr>
<td>12</td>
<td>168723730376298496</td>
<td>174491414850310144</td>
<td>0.1818181818181818</td>
</tr>
<tr>
<td>13</td>
<td>168723730376298496</td>
<td>168707698483929089</td>
<td>0.2222222222222222</td>
</tr>
<tr>
<td>14</td>
<td>168723730376298496</td>
<td>168730920071872512</td>
<td>0.125</td>
</tr>
<tr>
<td>15</td>
<td>168707841941704704</td>
<td>168738188100251648</td>
<td>0.1333333333333333</td>
</tr>
<tr>
<td>16</td>
<td>168699482874200066</td>
<td>168699476423344128</td>
<td>0.1176470588248247</td>
</tr>
<tr>
<td>17</td>
<td>168699482874200066</td>
<td>168731038300913664</td>
<td>0.1578947368421053</td>
</tr>
<tr>
<td>18</td>
<td>168699482874200066</td>
<td>168723716203745282</td>
<td>0.1111111111111111</td>
</tr>
<tr>
<td>19</td>
<td>168699482874200066</td>
<td>16871573148274689</td>
<td>0.1052631578947368</td>
</tr>
<tr>
<td>20</td>
<td>168699482874200066</td>
<td>168723694192037888</td>
<td>0.14285714285714285</td>
</tr>
<tr>
<td>21</td>
<td>168699482874200066</td>
<td>16868356109545472</td>
<td>0.1052631578947368</td>
</tr>
<tr>
<td>22</td>
<td>168699482874200066</td>
<td>168682301251784705</td>
<td>0.1176470588248247</td>
</tr>
<tr>
<td>23</td>
<td>168699482874200066</td>
<td>168668340175380480</td>
<td>0.1578947368421053</td>
</tr>
<tr>
<td>24</td>
<td>168699482874200066</td>
<td>168744836453761024</td>
<td>0.1176470588248247</td>
</tr>
</tbody>
</table>
Mapping Languages
Mapping Languages
Mapping Languages
Sentiment Mapping
Sentiment Mapping
Suggested Discussion Topics

- OS provides the opportunity to improve processing speed
- Compare IT resources required to maintain software
- Unfamiliarity for experienced GIS users may lead to frustration
- OS provides building blocks that can be used to build out entire system
- OS can run without being “installed”
- Software update schedules
- Licensing restrictions: What does it mean to be “commercially” friendly?
- What is the actual cost?
Now how did we get to that decision?

Emphasize here that we have scratched the surface – encourage a follow-on full study comparing open source desktop GIS to COTS GIS. This will allow us to provide a better recommendation on what to go “open” on and what should remain commercial.

For example, can you build a model in open source, how do you do it? Is it as “easy” as COTS?

This is an ongoing project…
Thank you