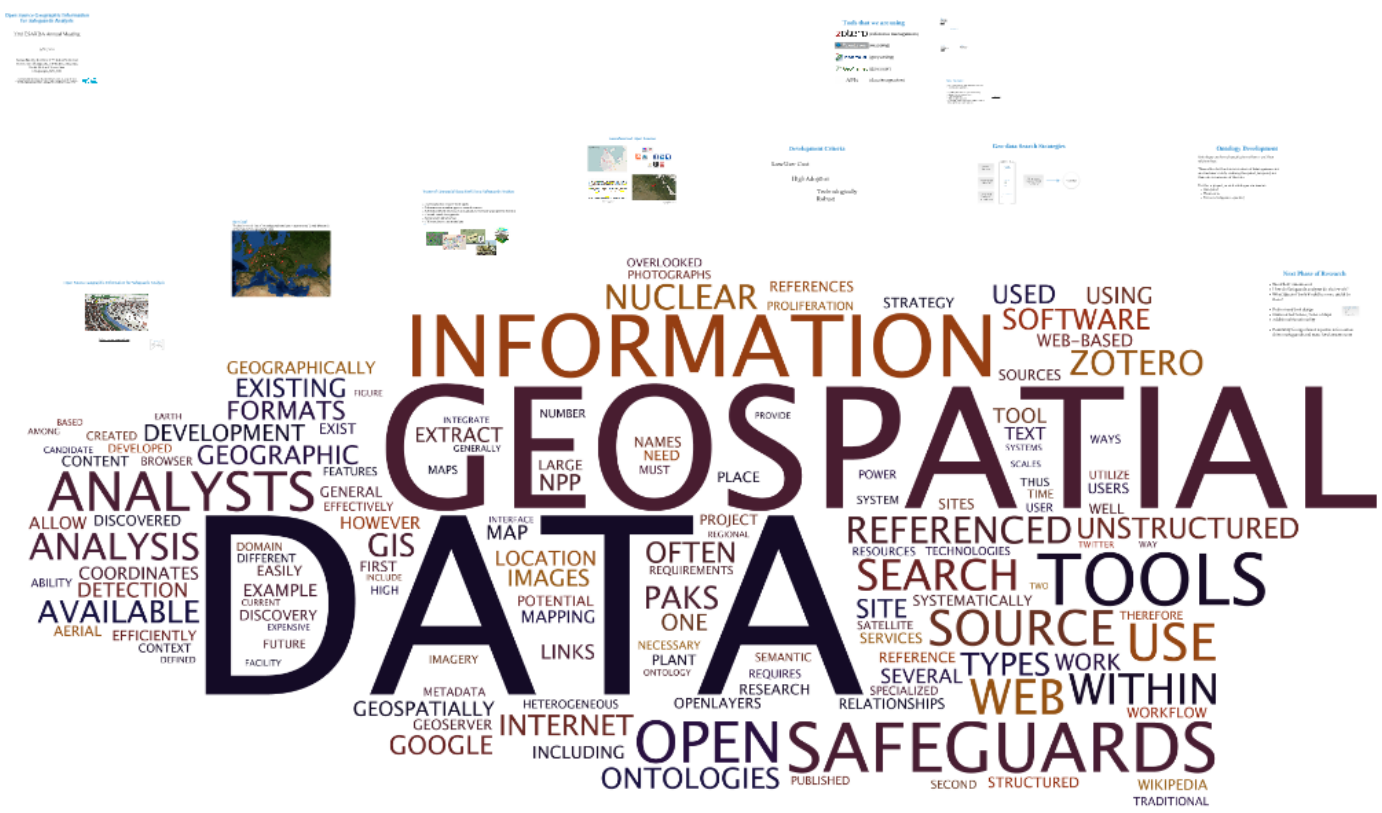
SAND2011-3260C

Questions?



Contacts

Dr. Karl Horak
kehorak@sandia.gov

Michael McDaniel
mmcdani@sandia.gov

Sandia National Laboratories

Open Source Geographic Information for Safeguards Analysis

33rd ESARDA Annual Meeting

19 May 2011

Denise Bleakly, Karl Horak & Michael McDaniel
International Safeguards and Technical Systems
Sandia National Laboratires
Albuquerque, NM, USA

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Sandia National Laboratories

Open Source Geographic Information for Safeguards Analysis

http://www.osm-3d.org

Research Overview

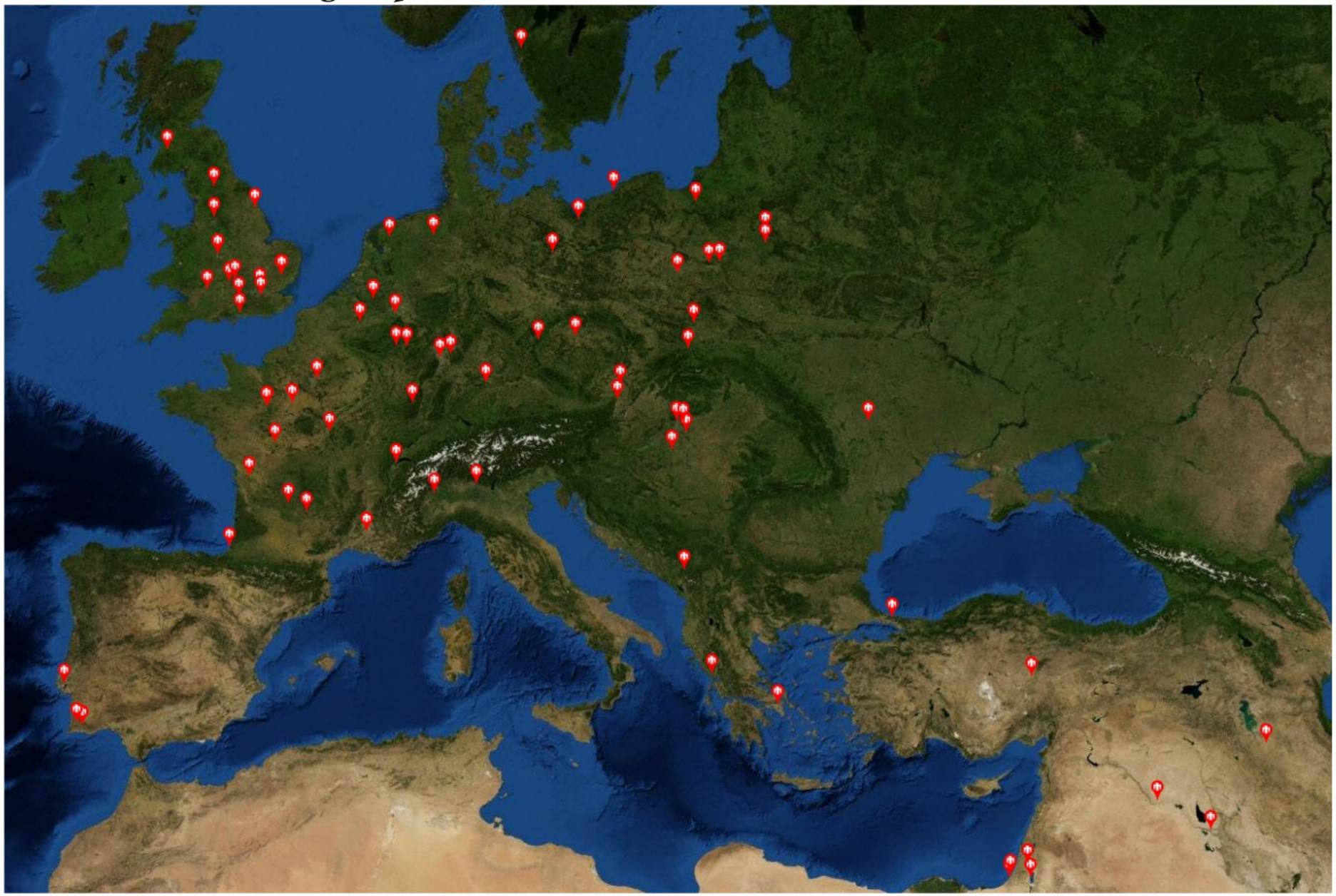
We are on a 2 year research project to explore ways to harvest geo-referenced information from Open Sources for Safeguards Analysis.

We seek to enable safeguards analysts to use open source geospatial information efficiently and effectively by means of novel web-based information technologies.

We are looking for participation from analysts to help guide our work.

Our Goal

To develop tools that allow safeguards analysts to systematically and efficiently extract and utilize geospatial data.

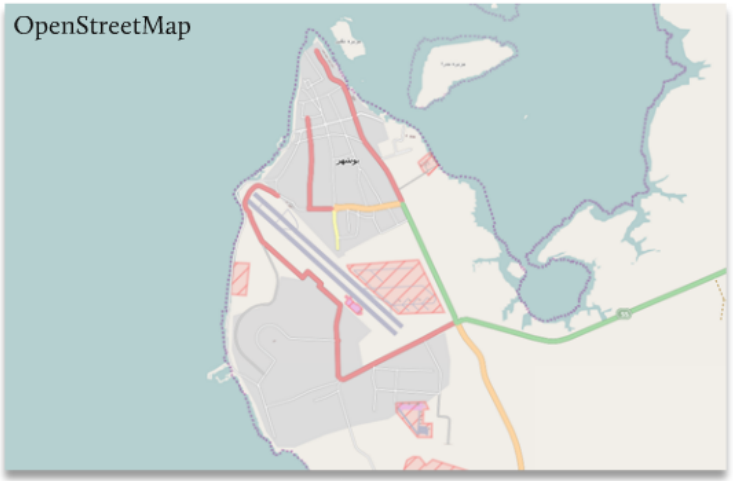


Types of Geospatial Data Useful to a Safeguards Analyst

- Overhead aerial or satellite imagery
- Reference maps and images to provide context
- Reference information such as roads and other nearby geographic features
- Ground-based photographs
- Detailed site information
- GIS/map data to use in analysis

The majority of this is unstructured data

Georeferenced Open Sources



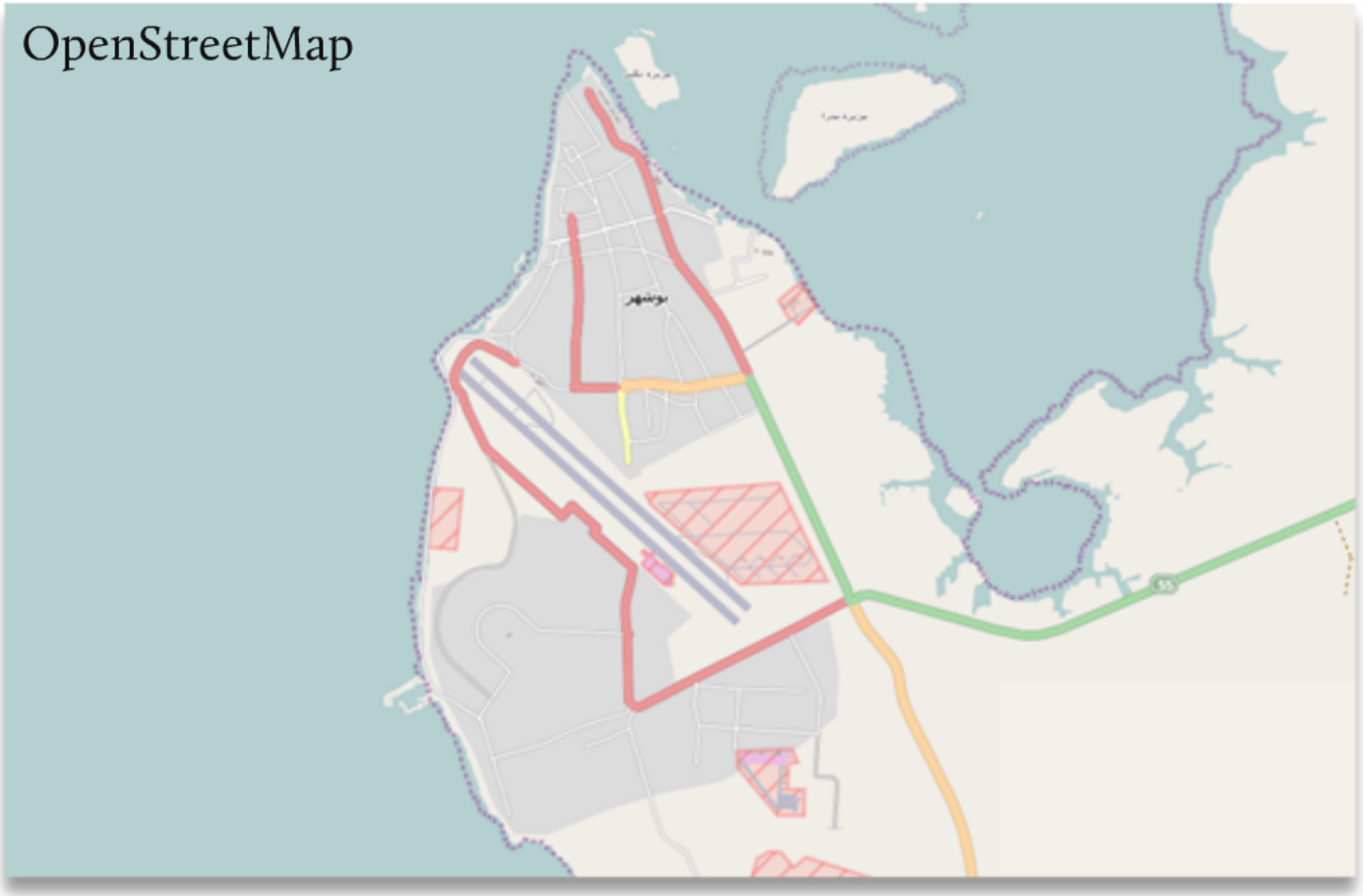
The LOCA at Fukushima began on March 11. The next day, 60,000 people for
human chain extending from the Neckarwestheim nuclear power plant (GKN-
one of the oldest German reactors and subject to life extension politics) to Stutt
forty miles away.

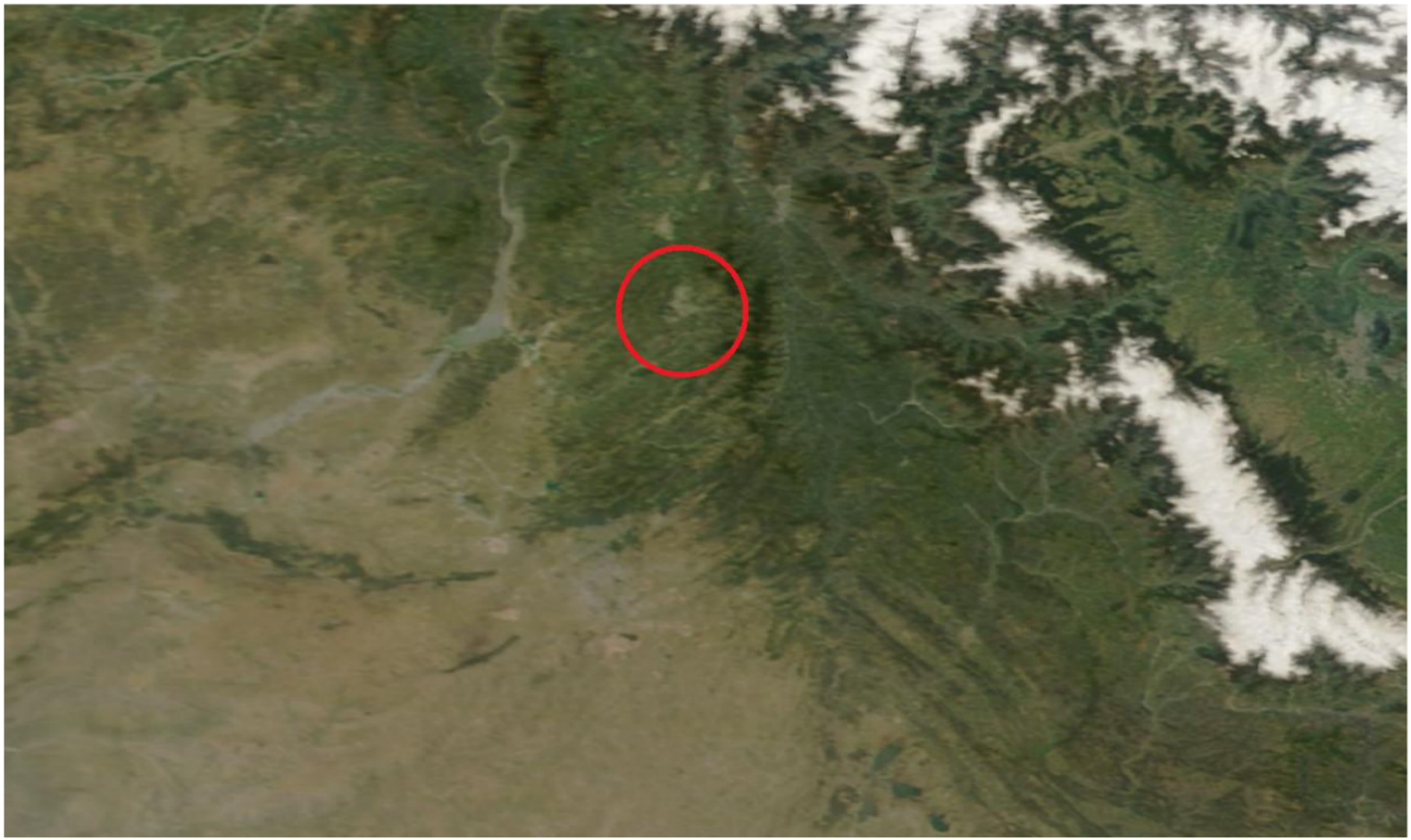
Stuttgart wasn't chosen as the Endziel of this demonstration by chance. (Endzi
terminus, BTW, and if you ever learned any German in a prior lifetime, you'll
refresher course by clicking on to some of the links on this post) Stuttgart is the
of the state of Baden-Wuerttemberg. A pivotal state parliamentary election wa
scheduled for March 27, sixteen days after the Fukushima LOCA began. The o
of that election would be critical to Merkel's efforts to assure that she has a maj
the upper house of the federal parliament (states' chamber, where voting is po

Open-source text with place names automatically highlighted. Note the ambiguity error with "Merkel."

Abbottabad, Pakistan
NASA MODIS Satellite (Terra)
2 May 2011

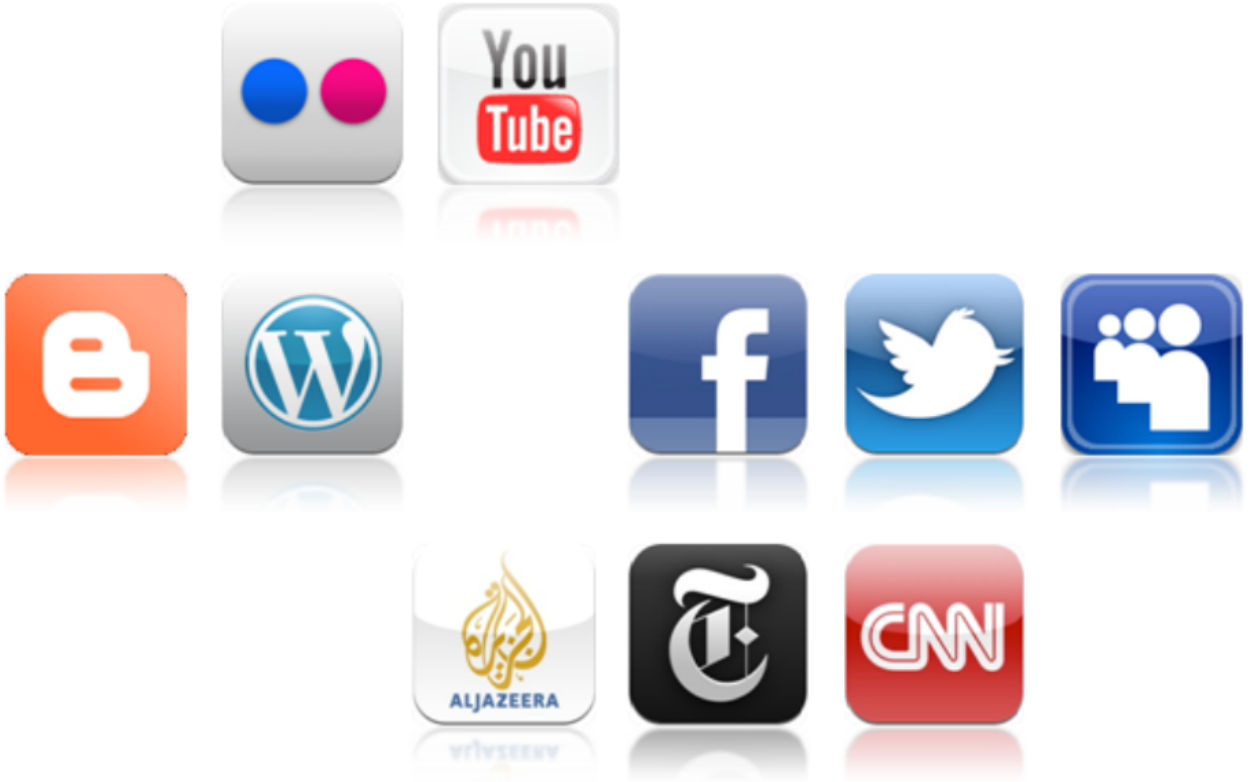
Georeferenced Op

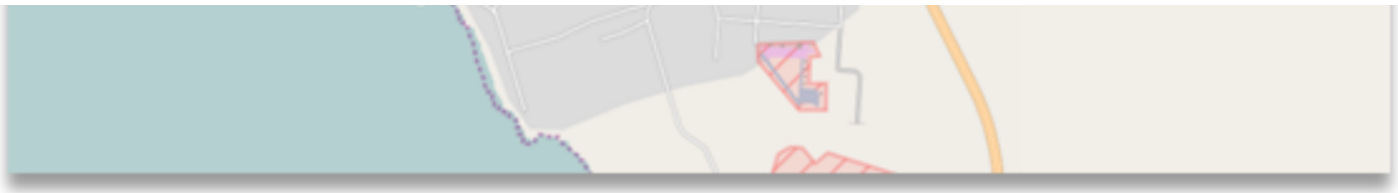




Abbottabad, Pakistan
NASA MODIS Satellite (Terra)
2 May 2011

renced Open Sources





The LOCA at Fukushima began on March 11. The next day, 60,000 people for
human chain extending from the Neckarwestheim nuclear power plant (GKN-
one of the oldest German reactors and subject to life extension politics) to Stutt
forty miles away.

Stuttgart wasn't chosen as the Endziel of this demonstration by chance. (Endzi
terminus, BTW, and if you ever learned any German in a prior lifetime, you'll g
refresher course by clicking on to some of the links on this post) Stuttgart is the
of the state of Baden-Wuerttemberg. A pivotal state parliamentary election wa
scheduled for March 27, sixteen days after the Fukushima LOCA began. The ou
of that election would be critical to Merkel's efforts to assure that she has a maj
the upper house of the federal parliament (states' chamber, where voting is pop

Open-source text with place names automatically highlighted. Note the ambiguity error with "Merkel."

Development Criteria

Low User Cost

High Adoption

Technologically Robust

Developm

Low User Cost

High Ado

Open source

set Cost

Open-source

High Adoption

Existing workflow integration
Refined user interface
Mild learning curve
Intermediate utilization

Techno

Robust

Existing workflow integration
Refined user interface
Mild learning curve
Internationalization

loption

Existing workflow integration
Refined user interface
Mild learning curve
Incremental cost

Technologically Robust

Easily expanded to handle new and different data sets, functionality, and workflows

Easily expanded to handle new and different data sets, functionality, and workflow.

Tools that we are using

zotero (reference management)

(mapping)

(geoparsing)

(gazetteer)

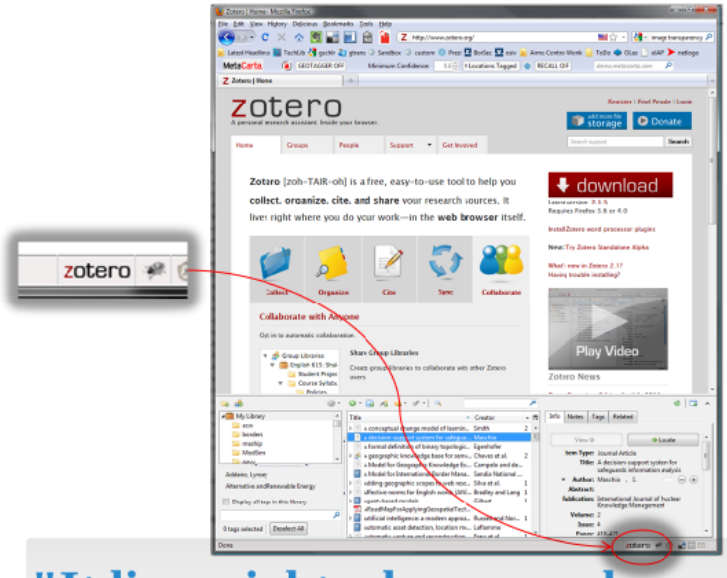
APIs (data integration)

Zotero

An open source citation management tool deployed as an add-on for the Firefox web browser, which is also being developed for other browsers and as a standalone tool

Used to collect, organize, cite and share electronic research sources.

http://www.zotero.org/



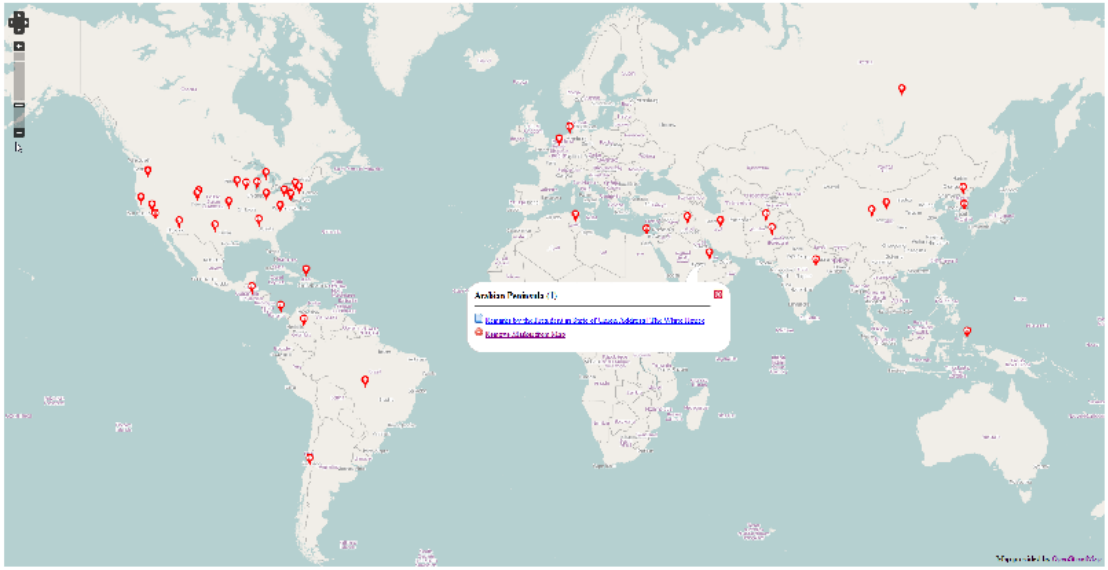
"It lives right where you do your work– in the web browser itself."

OpenLayers

- Open source (BSD-style license)
- Pure JavaScript
- Integrate information from multiple sources
- Support dozens of different geospatial data types
- Project of OSGeo (widely supported)
- OGC compliant (interoperable)
- Internationalization (i18n)
- http://openlayers.org

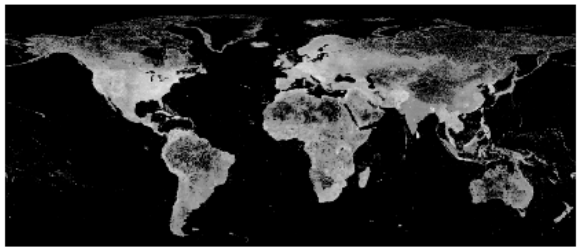
Zotero Maps

- An extension to Zotero.
- Rapid map creation from place names in text.
- Based on OpenLayers mapping library.
- No current support for disambiguation.



Yahoo! PlaceMaker

- Free webservice that extracts placenames from text
- Disambiguation capabilities



http://www.geonames.org

- 10 million place names (~7.5m unique features)
- Classified under 645 feature codes
- Dozens of languages
- Millions of alternate names
- Semantically enabled (each feature is liked to parent, child, neighbor and nearby features)

Geo-data Search Strategies

General Searches

Geo-enabled Searches

Structured Geospatial Data Searches

Types of Data

Collection of Geo-referenced Information

Knowledge

General Searches

General Internet Search

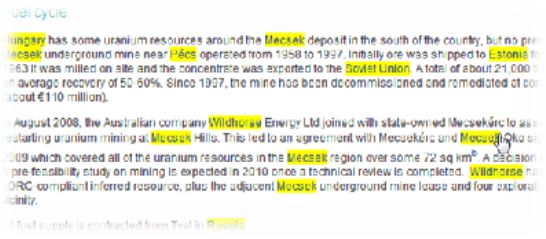
Text, images, etc. (unstrucutred data)

"Paks Nuclear Power Plant"

46° 34′ 21″ N, 18° 51′ 15″E

Google

WIKIPEDIA
The Free Encyclopedia



Geo-enabled Searches

Geo-enabled Search

wikimapia

- Crowdsourced geospatial information
- User digitzied and annotated features
- > 1m registered users
- >15m places
- Accessible through API

GeoHack

- 28 global map services sites (Google™ Maps, Wikimapia, OpenStreetMap, etc.)
- 12 Wikipedia links
- 10 photo hosting websites
- 19 "other sites"
- Over 100 Regional map services
- It linked to dozens of non-English sites

Google Earth

- Connection to existing WFS
- Large user community
- Reads/writes many geospatial dataformats (Pro version)

Geo-enabled

wikimapia

- Crowdsourced geospatial information
- User digitzied and annotated features
- > 1m registered users
- >15m places
- Accessible through API

GeoHack

- 28 global map services sites (Google™ Maps, Wikimapia, OpenStreetMap, etc.)
- 12 Wikipedia links
- 10 photo hosting websites
- 19 “other sites”
- Over 100 Regional map services
- It linked to dozens of non-English sites

S

ial dataformats (Pro version)

Google Earth

- Connection to existing WFS
- Large user community
- Reads/writes many geospatial dataformats (Pro version)

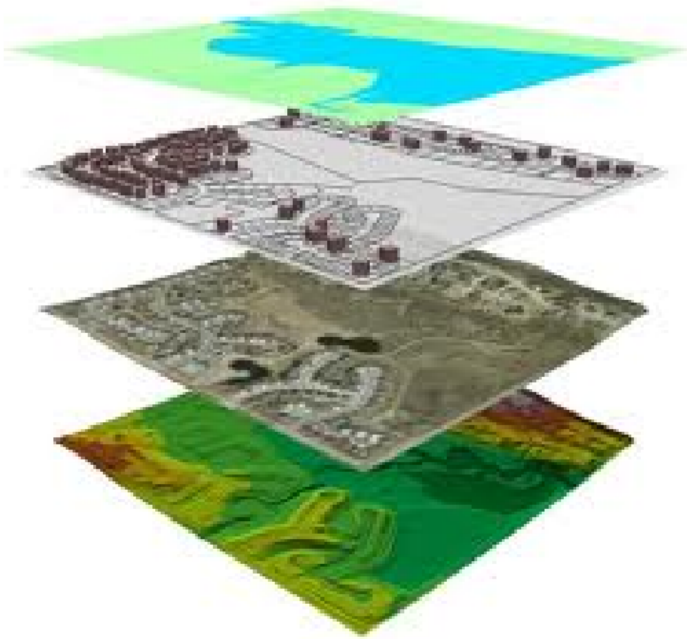

© 2011 Tele Atlas
Image © 2011 DigitalGlobe
© 2011 Cnes/Spot Image
Image © 2011 GeoEye
©2010 Google
Imagery Date: Dec 20, 2006
45°35'05.22" N 18°50'49.15" E elev 307 ft
Eye alt 753 ft

Structured Geospatial Data Searches

Structured Geospatial Data

Structured GIS data, like Shapefiles, Digital Elevation Models, CAD layers, etc are scale dependent.

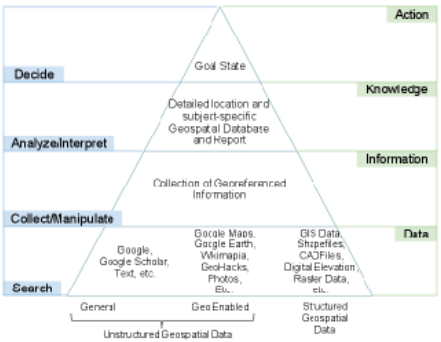
We found data for national, regional and local scales – enough to make representative maps



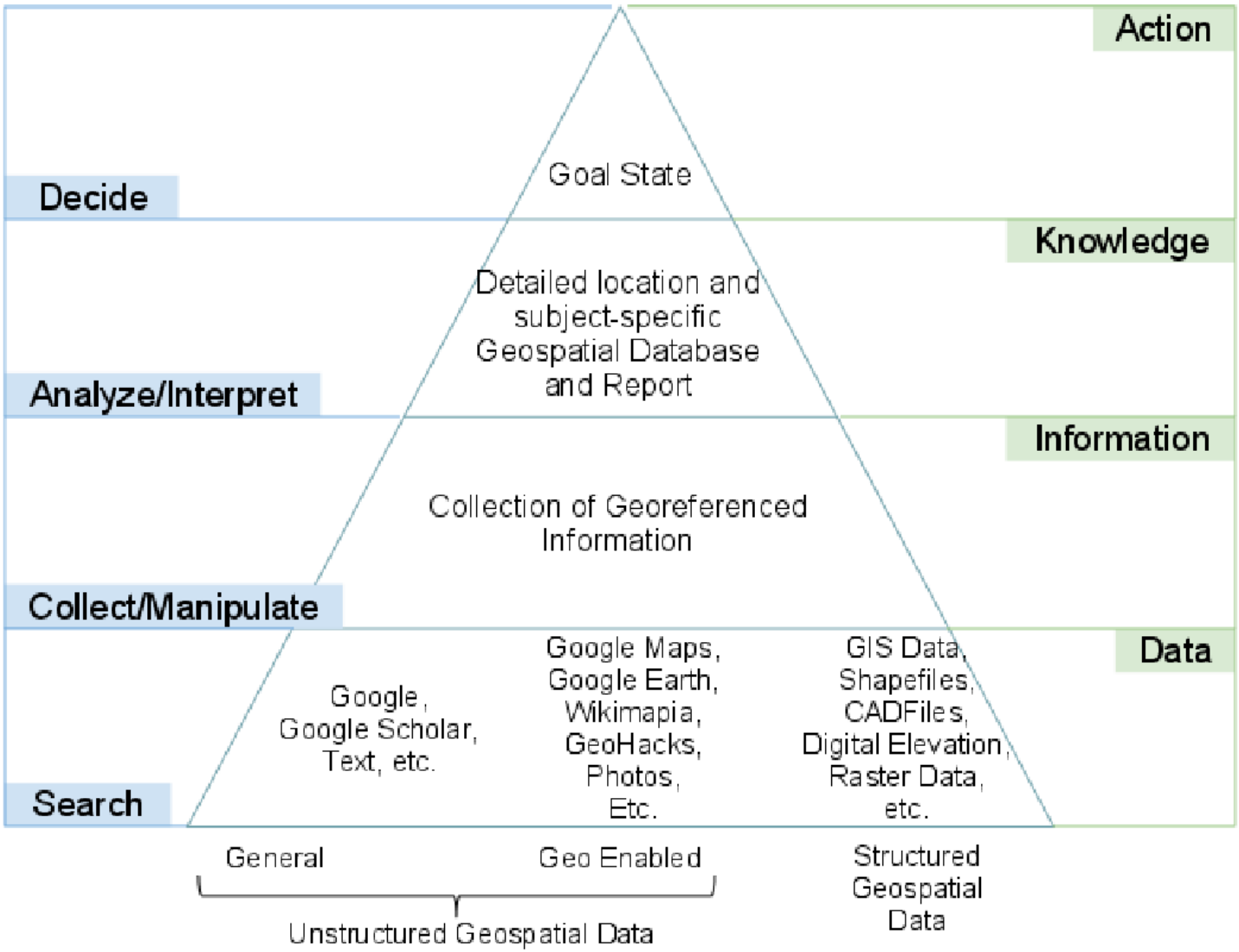
Collection of Geo-referenced Information

Knowledge

Action



Action


Action
Goal State
Decide
Knowledge
Detailed location and subject-specific Geospatial Database and Report
Analyze/Interpret
Information
Collection of Georeferenced Information
Collect/Manipulate
Data
Google, Google Scholar, Text, etc.
Google Maps, Google Earth, Wikimapia, GeoHacks, Photos, Etc.
GIS Data, Shapefiles, CADFiles, Digital Elevation, Raster Data, etc.
Search
General
Geo Enabled
Structured Geospatial Data
Unstructured Geospatial Data

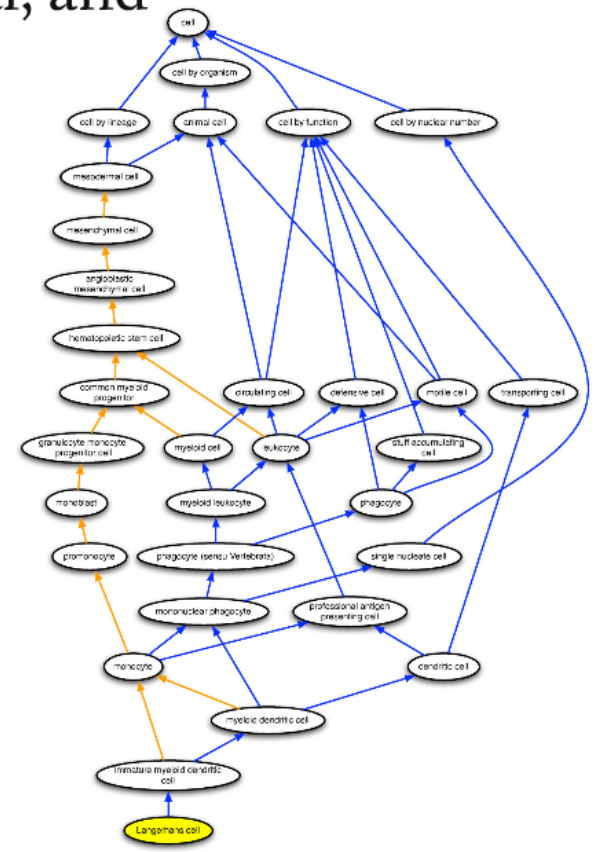
Ontology Development

Ontologies are formal specification of terms and their relationships.

These allow for the standardization of heterogeneous and unstructured data by defining the spatial, temporal, and thematic dimensions of the data.

For this is project, several ontologies are needed:

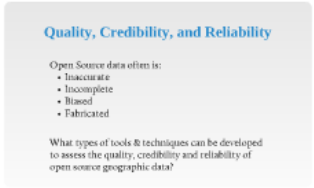
- Geospatial
- Placenames
- Domain (Safeguards specific)



Next Phase of Research

- Workflow Assessment
- How do Safeguards analysts do their work?
- What types of tools would be more useful to them?

- Preliminary tool design
- Customized Zotero, Zotero Maps
- Additional functionality



- Possibility for significant input to information-driven safeguards and state-level assessments

Quality, Credibility, and Reliability

Open Source data often is:

- Inaccurate
- Incomplete
- Biased
- Fabricated

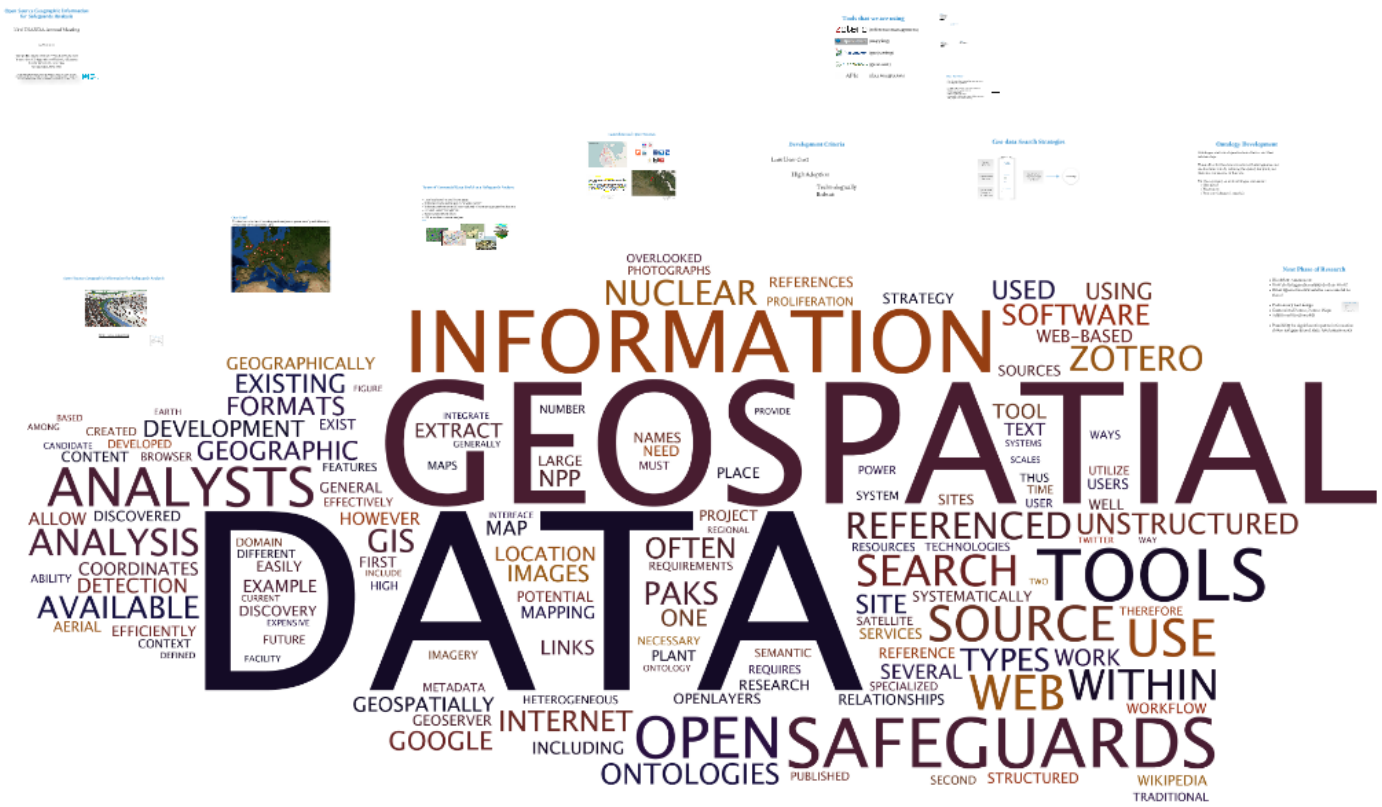
What types of tools & techniques can be developed to assess the quality, credibility and reliability of open source geographic data?

Participation

We are seeking analysts to provide us with input on their analysis work flow – to help us create a better, more integrated toolset.

An on-line survey is available at
http://www.surveymonkey.com/s/SYYPY6F

Questions?



Contacts

Dr. Karl Horak
kehorak@sandia.gov

Michael McDaniel
mmcdani@sandia.gov

Sandia National Laboratories