SANDIA REPORT

SAND2012-9916
Unlimited Release
Printed October 2012

Enhancing Safeguards Analysts' Geospatial Usage

Denise Bleakly, Karl Horak, Michael McDaniel

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

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.

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone:	(865) 576-8401
Facsimile:	(865) 576-5728
E-Mail:	reports@adonis.osti.gov
Online ordering:	http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd.
Springfield, VA 22161

Telephone:	(800) 553-6847
Facsimile:	(703) 605-6900
E-Mail:	orders@ntis.fedworld.gov
Online order:	http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2012-9916
Unlimited Release
Printed October 2012

Enhancing Safeguards Analysts' Geospatial Usage

Denise Bleakly[1], Karl Horak[2], Michael McDaniel[2]
[1] Technical Systems
[2] International Safeguards & Technical Systems
Sandia National Laboratories
P.O. Box 5800
Albuquerque, New Mexico 87185

Abstract

This report is the final summation of Sandia's Laboratory Directed Research and Development (LDRD) project #151316, ―Open Source Information Verification" (OSIV) which ran from FY11 through FY12. The aim of OSIV was to research, develop, and evaluate relevant geospatial analysis capabilities that address open-source information needs for international safeguards.

OSIV generated a number of technical, programmatic, and cultural advances, detailed in this report. There were new methodological insights and research that resulted in ten publications and presentations; this report concludes with an abstract-annotated listing of all materials. OSIV generated a substantial prototype, GeoSafeguards, that not only achieved its intended goal of testing our hypothesis, but which also served as a vehicle for customer education and program development. OSIV, as intended, has catalyzed future work in this domain; by the end of two years, it has already brought considerable attention to this work both domestically and with our international partners. Finally, the OSIV project knit together previously disparate research staff and user expertise in a fashion that not only addressed our immediate research goals, which has created cross- understanding, in service of Sandia's national security responsibilities in safeguards and nonproliferation.

ACKNOWLEDGMENTS

The authors would like to thank their managers, Dianna Blair and Richard Neiser, for support and guidance during this project. We also would like to acknowledge the contribution by our anonymous research participants, without whose efforts this would not have been possible.

TABLE OF CONTENTS

FIGURES

TABLES

NOMENCLATURE

AP	Additional Protocol
API	Applications Programming Interface
DOE	Department of Energy
DTRA	Defense Threat Reduction Agency
ESRI	Environmental Systems Research Institute
ESARDA	European Safeguards Research and Development Association
GML	Geographic Mark-up Language
GIS	Geographic Information System
GPS	Global Positioning System
IAEA	International Atomic Energy Agency
INMM	Institute of Nuclear Material Management
JSON	JavaScript Object Notation
KML	Keyhole Mark-up Language
LDRD	Laboratory Directed Research and Development
NPP	Nuclear Power Plant
OSIV	Open-Source Information Verification
SGIM	Safeguards Information Management
VGI	Volunteered Geographic Information
WMS	Web Mapping Service
XML	Extensible Markup Language

1 EXECUTIVE SUMMARY

This report is the final summation of Sandia‘s Laboratory Directed Research and Development (LDRD) project #151316, “Open Source Information Verification” (OSIV) which ran from FY11 through FY12. The aim of OSIV was to research, develop, and evaluate relevant geospatial analysis capabilities that address open-source information needs for international safeguards. OSIV generated a number of technical, programmatic, and cultural advances, detailed in this report. A key realization is that “open source” refers both to software and information. Therefore, the research team needed to survey software tools, information sources, search strategies, and methodologies for integrating these layers of information into an analyst‘s work process. This effort led to further awareness of the “digital footprints” left behind during research on the open Internet and potential operational security concerns.

OSIV, as intended, has catalyzed future work in this domain; by the end of two years, it has already gained considerable attention both domestically and with our international partners. There were new methodological insights and research that resulted in ten publications and presentations; this report concludes with an abstract-annotated listing of each. By means of these presentations, our research has been communicated widely to domestic and international audiences of like-minded subject matter experts and stakeholders. As a beneficial side-effect, these professional interactions led the researchers to form the Open-Source / Geospatial Information Working Group within the Safeguards Technical Division of the Institute of Nuclear Material Management (INMM). Now numbering over two dozen participants at dozens of institutions worldwide, the working group has created a LinkedIn online forum to strengthen and maintain professional relations in this field.

OSIV generated a substantial prototype, GeoSafeguards, that not only achieved its intended goal of testing our hypothesis, but which also serves as a vehicle for customer education and program development. Although sample sizes were small, significant trends were discovered and a number of qualitative findings came to light. We confirmed our hypothesis, that by enabling safeguards analysts to efficiently and effectively extract and utilize geospatially-referenced information from the Internet, these analysts will more often use these data in their analyses. GeoSafeguards is now under consideration for use by Sandia‘s Proliferation Sciences Department, DOE/HQ, the Defense Threat Reduction Agency (DTRA), and the International Atomic Energy Agency (IAEA).

Finally, the OSIV project knit together previously disparate research staff and user expertise in a fashion that not only addressed our immediate research goals, but has also created cross-disciplinary understanding, in service of Sandia‘s national security responsibilities in safeguards and nonproliferation.

2 INTRODUCTION

2.1 International Safeguards

International safeguards are implemented under the auspices of the IAEA pursuant to the Treaty on the Non-Proliferation of Nuclear Weapons (commonly known as the NPT or Non-Proliferation Treaty). While traditional safeguards seek to verify the correctness of a State‘s declared nuclear activities, enhanced verification measures under the Additional Protocol are required to allow the IAEA to provide reasonable assurances about the completeness of a State‘s declaration.

Because the Additional Protocol requires that the IAEA confirm the absence of undeclared nuclear activities for states with Additional Protocol (AP) agreements, it is necessary to collect and analyze new and broader sources of information, including the acquisition and analysis of open source information. Along with the transition from a facility-level assessment framework to a State-level assessment framework, new, efficient approaches for information utilization and management are required. Exploiting the geospatial dimension of information in open sources can help to make safeguards analysis more effective, while leveraging open source tools can make safeguards more efficient.

The International Atomic Energy Agency (IAEA) Department of Safeguards is facing a growing imbalance between an increasing workload and availability of resources. This imbalance is due to a combination of factors including the growing number and complexity of facilities under safeguards, an increased allocation of resources to high-profile investigations of suspected noncompliance, and the looming retirement of a significant fraction of safeguards personnel. The IAEA has acknowledged that adapting to these challenges will require a shift in the culture of safeguards implementation, from one that is facility-based and criteria-driven to one that is risk-informed and objectives-driven.

The IAEA is therefore evolving towards the implementation of the State-Level Concept, and Sandia National Laboratories has made the support of this transition a high-priority objective. From a technology standpoint, this will involve new emphasis on the development of information technologies that will allow for the collection, integration, and analysis of disparate data streams relevant to safeguards implementation.

Former Director of Concepts and Planning at the IAEA, Rich Hooper, asserts that all safeguards information exists along three common dimensions; technical, temporal, and geographic [1]. Based on this, we propose that by integrating all safeguards information along the geographic dimension, significant gains in efficiency can be realized.

In order to support such a transition, safeguards analysts, who generally lack training in geographic information science, must have tools to collect, integrate, visualize and share geographically referenced information. To achieve this, easily trainable, interoperable, and workflow-integrated tools must be developed that allow analysts to leverage the geospatial dimension of safeguards information.

Meanwhile, within the field of geospatial analysis, usability experts have explored the general use of maps for decision-making largely in terms of how to locate aerial and satellite imagery, how to find map layers of interest, and so forth. However, no research has been done with an international nuclear safeguards perspective. Our research will address this gap by examining the utility of a geospatial toolkit specifically as it applies to safeguards and nonproliferation.

While the quality of spatial data processing software is important within our analytical line of work, such preoccupations are not always reflected in research fields related to end-user applications [2]. Consequently, there is a missing link between the geospatial software effectiveness and efficiency in the safeguards domain and the information used in practice to overcome the consequences of complex and specialized geographical information system (GIS) software platforms.

Today‘s Internet is characterized by numerous interactive features that provide a plethora of avenues for user-contributed content. Ubiquitous cell and smart phones usually combine camera, web browser, global positioning system (GPS), and other tools that permit the person on the street to upload text, photographs, and video to any number of blogs, social media outlets, news networks, and other online repositories in near-real time. Growing numbers of mobile devices further increases the number of highly mobile, amateur and professional web contributors.

Much of this user-generated content is geospatially referenced. For example, modern digital cameras embed spatial coordinates by default, tweets (140-character Twitter messages) and Facebook updates allow users to –geotag” their location. Geospatial information can simply be entered manually, obtained from cell tower triangulation, or precisely derived from GPS.

Importantly for the purpose of this study, some of this web content may accidentally, incidentally or purposefully include information on nuclear facilities, materials, and perhaps even proliferation activities. However, as is obvious from the above examples, such information is often overlooked or not used because the routines to systematically extract these data must be hand-crafted. Even when a well-defined Application Programming Interface (API) is available, a solution must be constructed for each data source. In order to make this information available to safeguards analysts, web-based technologies need to be leveraged in novel ways so that the end user need not rely on web programming expertise.

2.2 Problem Statement

Several facets to the research have been identified for this project: (1) The identification of current open source tools with the potential to assist analysts in extracting and managing geospatial data; (2) an examination of the growing number of geospatial data types and the geotagging of typically non-geographic data like photographs; and (3) a test case to demonstrate the usefulness of these types of open-source geospatial information.

Previous research suggests that humans organize and process a great deal of information based on a –map-in-the-head” cognitive model [3]. It is therefore suggested that providing geospatial tools to analysts may decrease their cognitive load and increase their effectiveness in handling heterogeneous information that have common geospatial dimensions [4]. This research examined how geospatially referenced, safeguards-relevant information can be harvested using free, open-source tools to enable analysts to make timelier and more information intensive conclusions.

This research was designed to survey current geospatial resources on the open Internet and examines the feasibility of providing geospatial tools to analysts who do not have a high level of GIS or web programming fluency. Therefore, the hypothesis of this work is that, by enabling safeguards analysts to efficiently and effectively extract and utilize geospatially referenced information from the Internet, these analysts will more often use these data to produce more complete and context-rich analyses.

2.3 Open-Source Information for Safeguards

Safeguards open-source information management has involved labor-intensive, manual information processing and sophisticated, expensive, one-of-a-kind software systems [5] to deal with:

- Newspaper clipping services
- Hardcopy academic journals
- Industry and trade publications
- Traditional library research
- Paper maps, diagrams, and photographs, or
- Scanned versions of the same.

Now with the Internet (plus mobile phones, digital cameras, personal GPS 3 and 4-G networks, and Wi-Fi), it is a new world for information technology. Numerous examples demonstrate how rapidly open-source information has expanded to have a global impact. In Rwanda in 1994 warnings about the genocide were transmitted by fax. It never spread far enough to generate international assistance from the United Nations. In Kenya 2007, post-election violence similarly seemed about to spiral out of control. Eyewitness reports were transmitted by e-mail and text messages. An enterprising group called Ushahidi [6] aggregated these messages and placed them on a Google map layer, providing a meaningful visualization for the entire Internet community [7]. Statesmen noticed and the likes of Kofi Anan responded resulting in international response and abated violence.

The difference between 1994 and 2007 was not text messaging and Google Maps, which had been around for a number of years, but that the software and the Internet environment had matured to the point that people could make connections between disparate datasets and provide an online, map-based visualization of the complexities of the situation.

Implementation of the Additional Protocol [8] requires the utilization of new and expanded information, sources and analysis in order to confirm the absence of undeclared activity in member states. And while the use of advanced satellite imaging technology has been an area of development in international safeguards since the mid-1990s [9] there is a broader range of geospatial information sources that can be invaluable to the detection of undeclared nuclear activities. These rich sources of data exist in open sources but are largely unstructured and heterogeneous in nature, thus requiring specialized tools in order for them to be identified, extracted and integrated with existing safeguards information.

Based on discussions with analysts, desirable geospatial information for safeguards analysis includes aerial and satellite imagery, reference maps and images, detailed site information, terrain models, and digital geographic feature datasets. Recently, however, there has been a drastic increase in what one might call non-traditional geographic information available from open sources. Through the emergence of location-based services and online social networking technologies, user-generated content containing geospatial information, termed Volunteered Geographic Information (VGI), has become a ubiquitous part of the digital information landscape. These data vary from geotagged images and geographically referenced Tweets to place name references and aerial images in media reports. This ―citizens as sensors” phenomenon has been exploited aggressively by emergency response, humanitarian aid and human rights organizations [10]. And because data exhibit extensive global coverage with potentially high locational accuracy, and are generated in rapid response to events, these data may be important to the detection of undeclared activities.

While Wallace et al [11] note that information gleaned from open source alone is unlikely to lead to a definitive conclusion about the presence of undeclared activity, examining open source information though a geospatial filter can improve an analyst‘s ability to discover safeguards-relevant information and thus increase the effectiveness of the analysts work.

Recently, articles discussing nuclear proliferation detection using geospatial data have focused on the use of aerial and satellite imagery for change detection analysis [12][13]. This has led the way to the systematic use of satellite and aerial imagery within the safeguards community as one way of identifying undeclared nuclear activity [11]. Other geospatially referenced open source information, such as ground level images from tourists and visitors, ―crowdsourced” map data, and geospatial references in blogs or discussion wikis, are a resource that has not been systematically analyzed to determine their usefulness in safeguards analysis.

2.4 Geospatial Data

Traditionally, geospatial data have come in two primary categories, each with relatively few but well defined data types: vector (or geometric) and raster (or image) data types. While this distinction still holds in general, the emergence of collaborative web technologies has supported the rise of dozens of new ways of encoding geospatial data, as well as a shift in the ways in which geospatial data is produced and conceptualized. Because high quality geospatial data is no longer created and published only by large government agencies, or private sector firms, as has generally been the case in the past, traditional methods of geospatial data discovery fail to uncover many of these new data types and formats.

To guide the development of these tools and to determine how such tools might assist in the geospatial data discovery process, three different search strategies have been devised to discover geospatial data in all formats. First, a general Internet search strategy using search engines such as Google, Google Scholar, and Wikipedia leads to the discovery of unstructured geospatial data in text and images. These data require additional computational procedures to transform them into geospatial data types useable in a mapping context. Second, a geographically enabled search strategy using specific geospatial filters such as coordinate pairs, bounding box coordinates, or administrative boundary names, leads to the discovery of geotagged data and geospatial web services. These data are generally unstructured but have associated geospatial metadata.

Examples are geotagged images or blog posts. Third, structured geospatial data, such as ESRI shapefiles and GeoTIFF images, are discovered through geospatial data portals and clearing houses. In general these outlets are run by government or not-for-profit agencies, however, geospatial firms like Environmental Systems Research Institute (ESRI) the makers of the popular ArcGIS software, are now providing internet-based on-line data services. GeoSafegards is not a geospatial data search tool, but rather a tool that can be used to visualize geo-tagged and geospatial data. Note that these search strategies are not mutually exclusive and one search strategy can lead to the discovery of different types of data.

2.5 Open-Source Software for Safeguards Analysis

2.5.1 Open-Source Software

The utilization of open-source software to address safeguards needs represents the potential for great efficiency gains over the use of proprietary software. Open-source software projects are the result of many thousands of labor hours by a broad, global community of developers, and are distributed at no cost to the end user under Open Source Definition licenses [14]. These licenses are free from copyright restrictions and can therefore be further developed and redistributed to meet the needs of specific users such as safeguards analysts. Therefore, leveraging technologies developed in the open source can be both time and resource conserving. This potential efficiency has been recognized by the United States Government who, through the Department of Homeland Security, has funded a 5 year, 10 million dollar project to examine the use of open-source software to meet national cyber security needs [15].

The effective use of open-source information requires an integrated system for collecting, evaluating, structuring, analyzing, securing, and disseminating information [16]; however, working with geographic information, and particularly unstructured geographic information, requires a set of tools distinct from other analysis tools. Generally, proprietary software for working with geospatial information is costly and requires extensive training. However, there is a large community of software developers working on open source technologies for geospatial solutions.

2.5.2 Tools assessment

To guide the assessment of the tools available to achieve these goals, several requirements were established for the development of a safeguards toolset.

- First, the tool should be available at low or no cost to the end user. This requirement led to the examination of existing open source software that could be used, modified, and re-distributed free of charge and copyright restriction. Moreover, by leveraging open source software, existing capabilities need not be reinvented, thus saving significant development time.
- Second, the developed tools should integrate easily into analysts‘ existing workflows. This will help to ensure that the tools can be used widely among analysts by lowering the threshold of adoption. This also requires that tools have an intuitive design and

functionality, refined user interface, and that users are closely involved in the development. This further requires that these tools be compatible with existing software and data.

- Third, these tools must be extensible and thus adaptable to different analysis requirements, individual preferences, and future changes in workflow.

The tools that we assessed included, but were not limited to:

- Geospatial data servers: GeoServer
- Geoprocessing libraries: GeoTools
- Content management systems and Wikis
- Reference management systems: Zotero, EndNote
- Web browsers: Firefox, Internet Explorer
- Entity extraction services: Citrus, OpenCalais
- Meta-search engines
- Web crawlers
- Web gazetteers

Tools that were not included into the final product either went beyond the scope of the project or were excluded based on time limitations.

3 GEOSAFEGUARDS

In order to give analysts with little or no training in GIS the capability to explore the geospatial dimensions of unstructured information and to incorporate this information into their analysis, we have developed a lightweight tool that integrates with existing analyst tools and workflows. The tool, called GeoSafeguards, integrates several open-source software programs, web services, and data sources to create a unique capability for safeguards analysts to create and interact with geographically referenced safeguards information.

The GeoSafeguards tool has these characteristics:

- Lightweight tool and techniques that work within existing workflows
- Rapid geographic visualization of unstructured data
- Introduce analysts to a non-traditional view of open source and geospatial information
- Use as a starting point for further geospatial analysis
- Augmentation of open source research techniques to expose the geospatial dimensions not generally looked for in open source research
- Based on free and open-source technology which makes the tools
- Cost effective
- Flexible for a variety of needs
- Compatible with a variety of computing environments, and existing tools

3.1 Software of the GeoSafeguards Tool

The software used to create the GeoSafeguards tool includes the Mozilla Firefox web browser, Zotero reference management plugin for Firefox, and the OpenLayers JavaScript mapping library. Integrated web services include Yahoo! Placemaker and various GeoNames services. The GeoSafeguards tool also closely integrates with GeoHack and uses Google Maps and OpenStreetMap base layers on which additional information is overlaid. Appendix D, provides screen shots of the GeoSafeguards tool and some of the geographic data it draws upon.

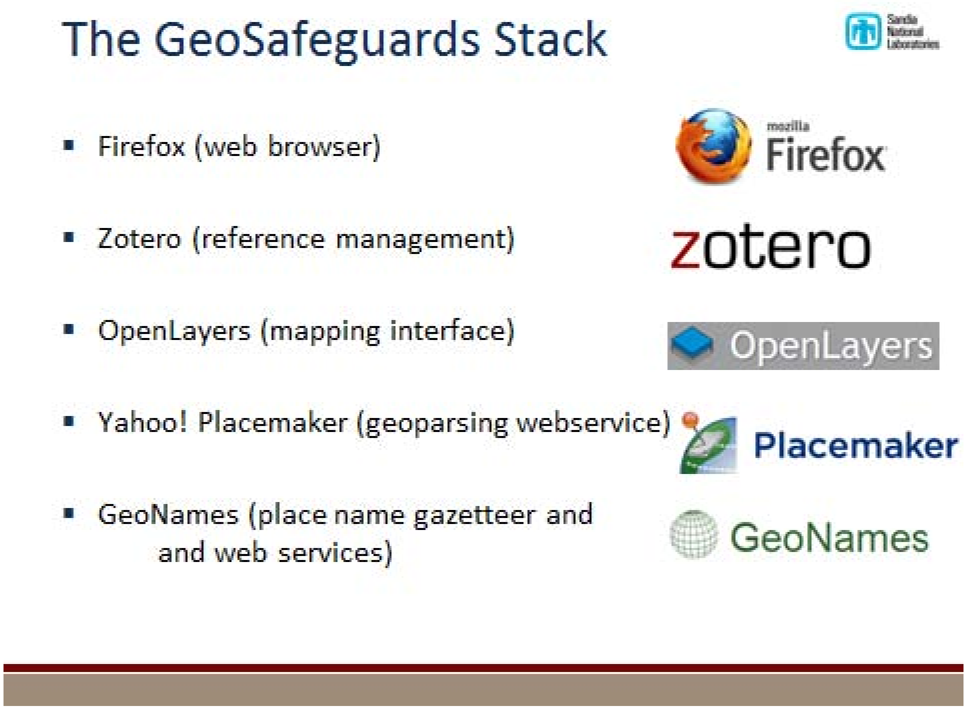

Figure 1. The GeoSafeguards software stack.

3.1.1 Firefox

Because web browsers are ubiquitous and essential to modern research and analysis, GeoSafeguards has been developed as an extension to the Firefox web browser. Most analysts, particularly those dealing with open-source information, use web browsers to access the majority of their information. Firefox is an open-source web browser developed by the non-profit Mozilla Foundation. Because Firefox is open-source it offers a greater degree of flexibility for developers to build additional capabilities. It tends to have a higher level of standards compliance and is generally more secure than Microsoft Internet Explorer.

3.1.2 Zotero

Zotero [17], an open-source reference management software developed by George Mason University, was chosen as an integral part of the GeoSafeguards tool. Because the workflow of safeguards analysts has a clear correspondence to the research process in which references are collected, tracked, and organized, and reports are generated, Zotero can be a valuable tool to safeguards analysts for collecting and managing open-source information. Zotero is deployed as an add-on for the Firefox web browser and therefore integrates into one of the primary tools of safeguards analysts—the web browser. Additionally, Zotero has a robust Application Programming Interface (API) with which additional capabilities can be developed, allowing the GeoSafeguards tool to leverage capabilities of Zotero such as document indexing.

3.1.3 OpenLayers

OpenLayers is a pure JavaScript library for displaying and manipulating geospatial data within a web browser [18]. With the ability to display standards compliant geospatial resources from anywhere on the Internet, OpenLayers is used to generate the front-end mapping interface of GeoSafeguards. OpenLayers is compliant with Open Geospatial Consortium standards [19] and thus interoperable with a wide variety of other geospatial software.

3.1.4 Yahoo! Placemaker

An important analysis capability due to the volume and complexity of the information gathered by safeguards analysts is that of entity extraction. This is an automated process by which unstructured text is analyzed and tagged according to the type of entities it contains (for example, people, places, companies, etc). Yahoo! Placemaker is a web service that extracts place names from text (a process known as geoparsing) and returns metadata for the document including geographic scope and coordinates [20]. This allows unstructured data to be tagged, organized, and visualized based upon the semantic, geospatial information contained within the text.

3.1.5 GeoNames

GeoNames is an online place name database, or gazetteer, amassed from dozens of sources including user contributions and contains greater than 8 million unique named geographic features [21]. GeoNames offers a convenient and nearly comprehensive place name search and also provides links to geographically referenced Wikipedia articles, which can provide valuable contextual information to analysts. While it provides free access via API, it can also be downloaded for local use for a fee.

3.1.6 GeoHack

Maintained as part of the Wikimedia Toolserver project, GeoHack offers access to dozens of global and regional mapping services and geographically referenced information based on geographic coordinate pairs [22]. Based on feedback from analysts, this resource provided unparalleled single-point access to unique geotagged information that is potentially directly or indirectly valuable to safeguards, so the decision was made to tightly integrate it into the GeoSafeguards tool.

The GeoSafeguards tool was developed to provide analysts with little or no training in GIS the ability to visualize unstructured information and generate structured geographic information, essentially bridging a capability gap between general safeguards analysts and geospatial analysts.

3.1.7 Base layers

GeoSafeguards integrates a number of base layers that serve as the background upon which all other data are overlaid. These include four Google Maps layers, one OpenStreetMap layer, and one MapQuest Open Aerial layer. These are essential for providing the geographic context of the data that is created by GeoSafeguards. We have included multiple layer types and providers to allow the user to gain different perspectives on the areas of interest depending upon the base layer selected. These base layers are delivered via the Internet as image tiles for the geographic extent covered by the GeoSafeguards window. While accesses to Google Maps layers require an API key, the OpenStreetMap and MapQuest Open Aerial layers are available without a key. Accesses and display of all layers are managed by the OpenLayers library. All layers are available at no cost.

3.2 The GeoSafeguards workflow

Using Zotero, document references are generated and placed into the document library and organized into collections based on some thematic criteria (e.g. a particular state or facility, or technical theme.) These references may be created manually, requiring the user to input metadata relating to the reference, or automatically, based upon metadata contained within a document‘s header. Zotero indexes the text contained within the document and stores this information on the user‘s machine in a SQLite database. After selecting a document or collection to generate a map from, the user initiates the GeoSafeguards interface from within Zotero. Once GeoSafeguards has been launched, the text from the indexed document collection is sent via an XMLHttpRequest to the Placemaker web services which detects place names within the text and returns an XML document containing metadata for each of the detected place names, including but not limited to coordinate pairs, location of place name within the text, a confidence measure, and an importance measure. The confidence measure is an integer between 1 and 10 indicating the level of certainty that the web service returned the correct place for the place name (e.g. 1 for low certainty, 10 for high certainty.) The importance measure is a positive integer approximating the importance of the place name within the document collection which relates to the number of times the place name appears in the document collection.

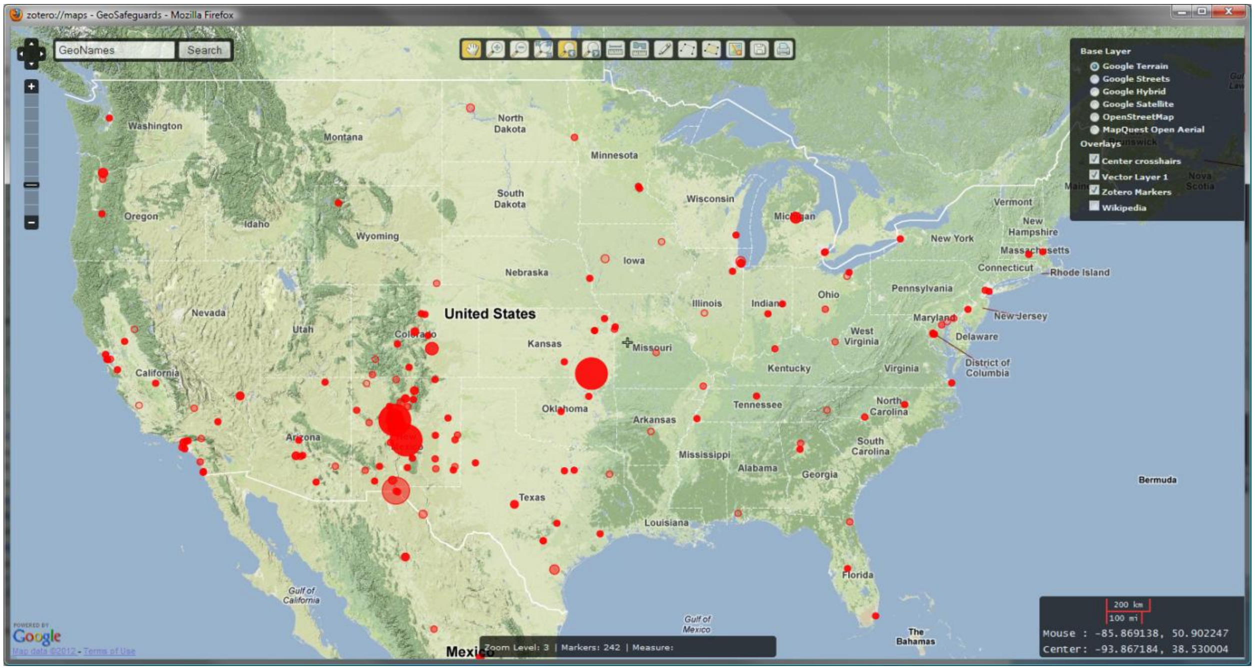

Figure 2 GeoSafeguards interface

Based upon the returned coordinate pairs each place name is mapped and displayed within the GeoSafeguards interface on top of the selected base layer. The confidence score for each place name is visualized on the opacity of the point where its importance is visualized by the size of the point. Figure 2 is an example of the GeoSafeguards interface with mapped locations displaying varying importance and confidence levels. The user can interact with the results, clicking on points to display the place name, place type (e.g. town, point of interest, etc.), confidence and importance measures, context in which the place name appears in the document, a link to the document(s) in which it appears, and a link to that specific location in GeoHack. See figure 21 in Appendix D for an example of the results presented when a point is selected.

The user can navigate within the map by a variety of means, using the keyboard, mouse or GUI buttons. GeoSafeguards also provides the user with a variety of other GIS capabilities, such as linear and area measuring and vector creation, and place name searching against the GeoNames web service. The user can turn geospatial layers, including the mapped place names, user generated vectors, and a layer for georeferenced Wikipedia articles, on and off. Geospatial data can be exported to a KML file and shared with others or a hardcopy map can be printed. Figure 7 in Appendix C shows the general system architecture and proposed future capabilities of GeoSafeguards.

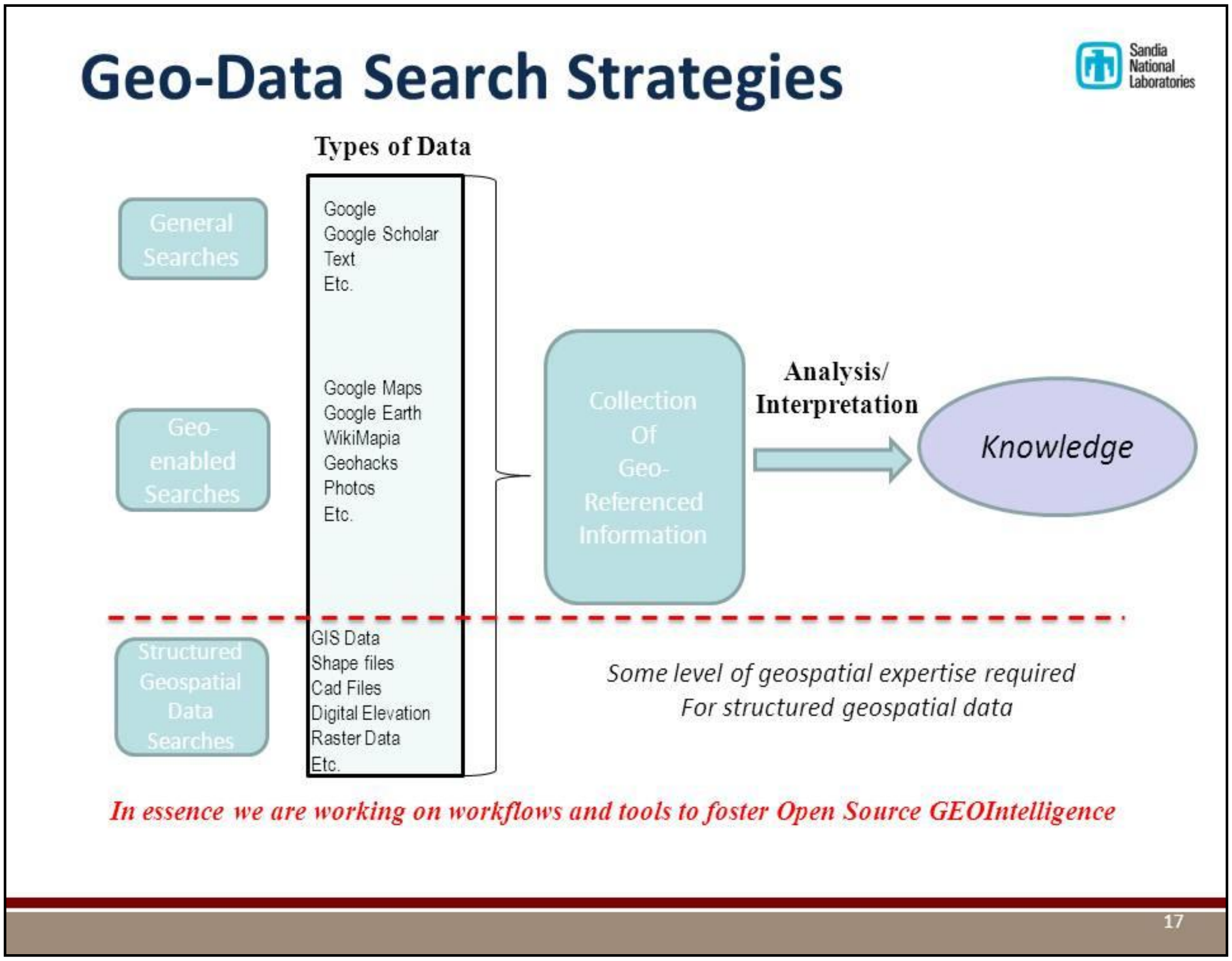

Figure 3 Hierarchy of search strategies for geospatial information

4 EXAMPLE OF HOW GEOSAFEGUARDS CAN BE USED

4.1 Test Case: Paks Nuclear Power Plant, Paks, Hungary

To demonstrate the wide variety of open source geographically referenced information available using this phased search strategy, a theoretical test case was developed based on the need to collect information to assist a safeguards and security analyst to understand Paks Nuclear Power Plant (NPP) near Paks, Hungary.

4.2 General Internet search

The first search included sites such as Google, Google Scholar and Wikipedia. Over 580,000 results were received on Google by searching for ―Paks Nuclear Power Plant”. The first entry returned was that of Paks NPP on Wikipedia. The second site listed was the home page of the power plant.

The Wikipedia site for ―Paks Nuclear Power Plant” had a wide variety of geographically referenced data [23]:

- A location map
- Latitude and longitude coordinates
- Multiple current and historic images of the site
- Links to the Paks NPP website and many other related sites
- Links to other papers and references

The Paks NPP home page [24] in English also provided some geospatially referenced information:

- An address of the facility
- A location map
- Images of the facility in their ―Virtual Tour and Gallery Links”

From these two websites alone, a substantial geospatial reference to the site can be built, note that these data are in text and image formats and as such cannot be easily utilized by traditional geographic information systems (GIS). Figure 4 shows the types of geographically referenced data found by doing these searches-- two different location maps, a satellite image and a 3D rendering of the site.

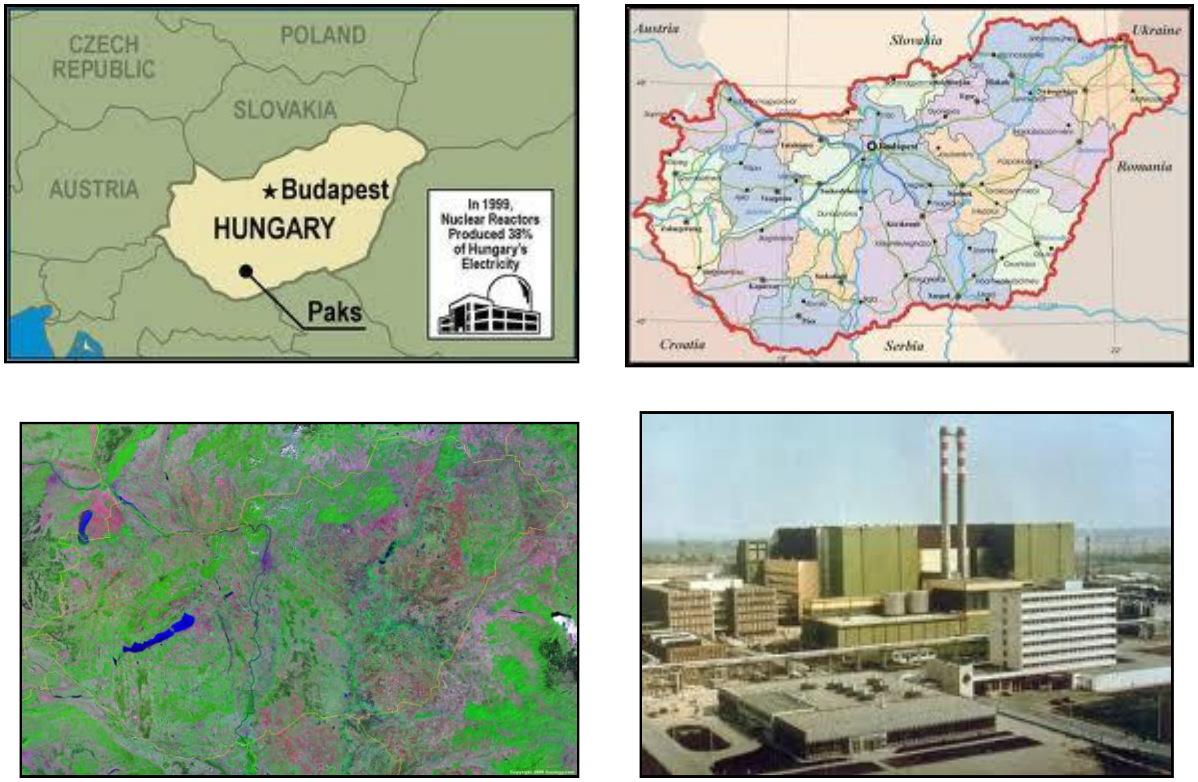

Figure 4 Examples of maps and images found during a general Internet search

4.3 Geo-enabled Search

From the information gained during the general internet search, specifically the coordinates of the Paks NPP (46.5725N, 18.854167E), Google Earth was used to get an aerial image of the site, dated 20 December 2006. Also available in Google Earth are 3-dimensional building renderings showing a photorealistic rendering of the main buildings at the Paks plant, including those housing the reactors, turbines and control rooms. Google Earth, which has become the layperson‘s geographic information system (GIS) of choice, has the ability to overlay data from dozens of already defined sources including Web Mapping Service (WMS) layers from any external source.

Next, Wikimapia [25], a crowdsourced mapping service that allows users to digitize and annotate geographic features, was examined. Users have digitized buildings and infrastructure at the NPP site, including reactor housings 1 through 4, cooling water input and output systems, switchyard, control room building, visitors center, fire station, meteorological tower, and bus station, among others. This information can be extracted through Wikimapia‘s API in XML, JSON, KML, and binary formats.

The third geo-enabled search was through the GeoHack [26] website. GeoHack is a tool developed by of the Wikimedia community‘s Toolserver project that aggregates mapping services that are capable of displaying georeferenced content from many different sources. By querying a latitude and longitude coordinate pair, GeoHack returns links to various mapping services that display data centered on these coordinates as well as links to other web-based resources related to these coordinates and thus serves as a valuable jumping off point to a large amount of geospatial data. From here, a large number other of websites containing geo-tagged information were discovered to include:

- 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

While each of these sites do not necessarily represent unique data points, as some links are coincident or contain identical data, this does illustrate the relative ease with which recent aerial and satellite imagery and geographic data visualizations are obtainable. Figure 5 shows the Google Earth image for the site and figure 6 shows one section of the GeoHacks report generated for the site.

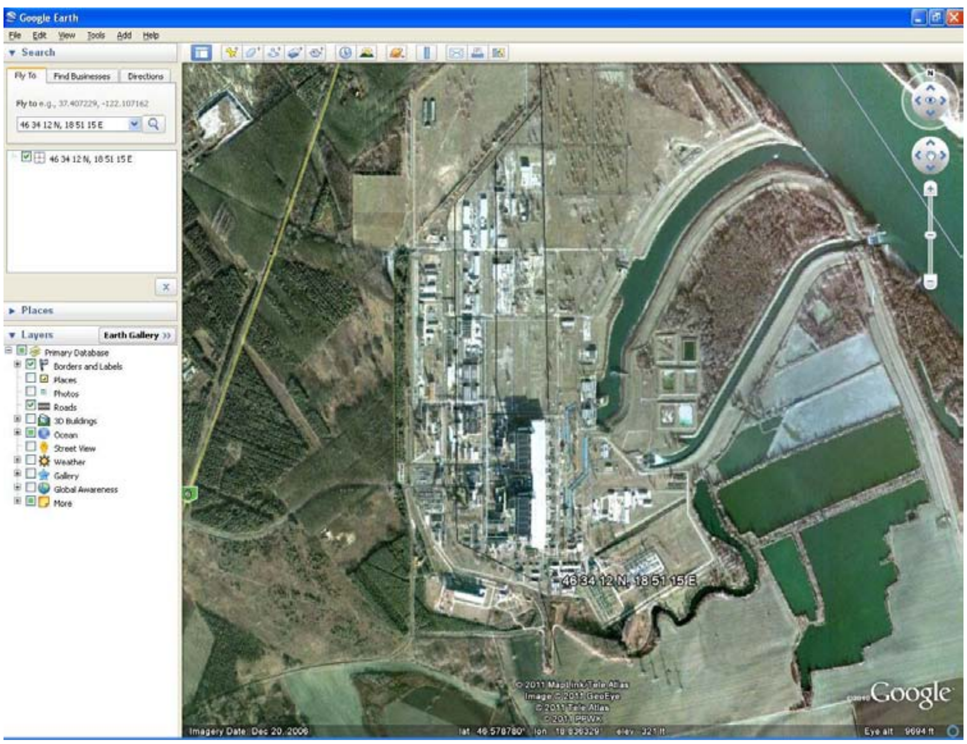

Figure 5 A Google Earth Image of the Paks Nuclear Power Plant

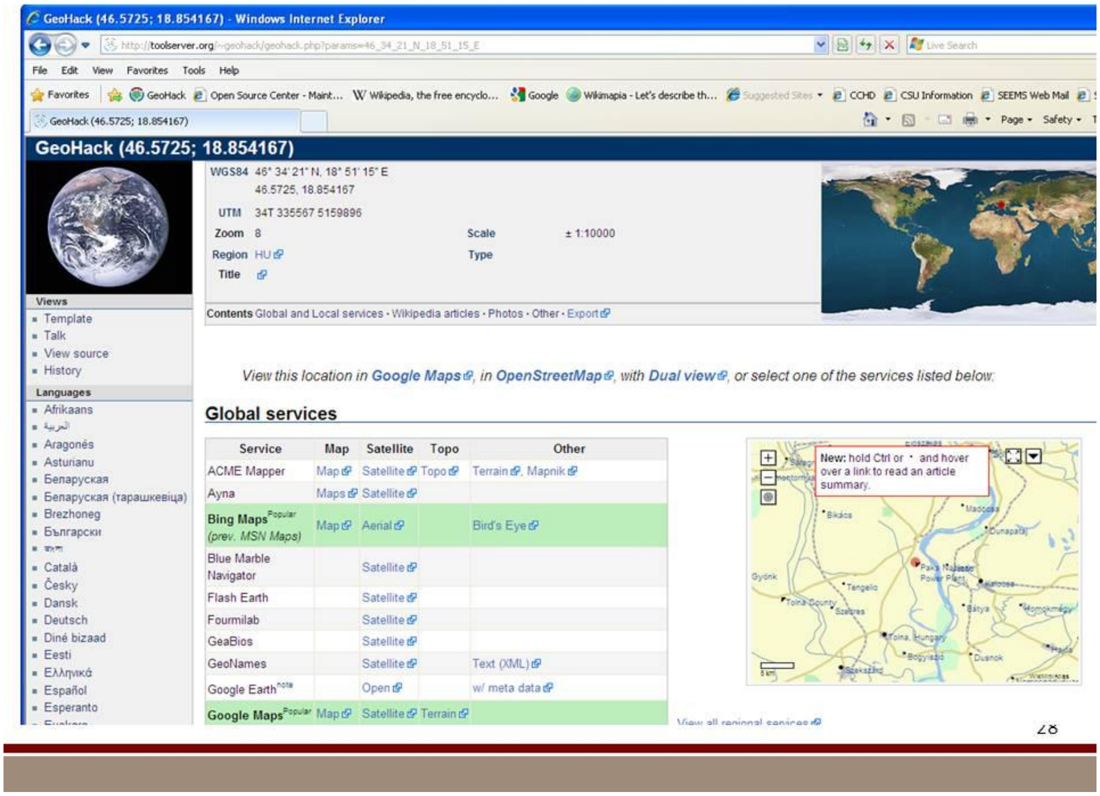

Figure 6 Screen shot of GeoHack information for Paks Nuclear Power Plant

4.4 Structured geospatial data search

Finally, an on-line search was conducted for standard structured geospatial data such as ESRI Shapefiles, digital elevation models (DEM), and GeoTIFF images. Effective use of data in most of these formats requires specialized GIS software (e.g. ArcGIS or MapInfo) and a trained geospatial specialist. However, several XML-based geodata formats (e.g. KML and GML) have emerged in recent years that allow these data to be used within a web-based computing framework and thus available to a larger number of analysts.

The quality and resolution of the GIS data discovered for the Paks NPP site ranged from very low to extremely high. While a large amount of data were discovered at state and regional scales, very little data were found at local and site-specific scales. For example, geospatial data for Hungary and Hungarian counties were abundant, while data for municipal scales and the Paks NPP site in particular were more difficult to come by. However, what one might consider ―micro-level" geodata, such as geotagged photographs, were widely available. This trend might indicate the need to, and benefit of, examining other sources of geographically referenced data to supplement this mid-scale data void.

4.5 Example Results

Based on discussions with analysts, desirable geospatial information for safeguards and security analysis includes:

- 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

Each of these data types were discovered on the open Internet with relative ease. Moreover, no tools exist (to the knowledge or the authors) that allow for the systematic detection, extraction and utilization of these data within a system that can be easily incorporated into the analysis workflow. Also, notably missing from this list are unstructured data (such as text data) containing geospatial references. Because these data are not easily used in a geospatial framework they are often ignored or overlooked.

5 TRAINING/TESTING

5.1 Background

For this research, we devised a combination training and testing program to test the ease of use and workflow of the GeoSafeguards tool. Our training and testing program was based on research related to end user applications [2]. There is a missing link between the geospatial software effectiveness and efficiency in the safeguards domain and the information used in practice to overcome the consequences of complex and specialized GIS software platforms.

A broad, exploratory research technique was required to study the phenomenon of effectiveness and efficiency in the use of geospatial software for analysis. Our research technique is predominantly inductive because it starts by asking subjects about their experiences and then uses the findings to induce theories. We explain our methods in the following sections. We used pre-and-post training surveys to ascertain how analysts learned GeoSafeguards, and how they may use it in their daily workflows. The methods section explains that the method includes surveys with software users from a range of backgrounds. The results section describes the overall results of the study and defines certain themes in order to describe the subjects‘ experiences.

Little research has been conducted into how general software user-experience standards (e.g. [27]) match with the users‘ concepts of usability, fitness for use, validity, or quality of geospatial information systems, not to mention their understanding of the terminology itself. Indeed, we believe there is very little empirical research relating to how people perceive and use spatial information for individual datasets in a real world environment. Accordingly, the challenge in this research is to define what determines usability to a subject who does not necessarily have formal training or education in spatial information theory. As such, it is a search for strategies and terminology to portray these important concepts to those who will be required to absorb any risk associated with use of the information.

In this study we collected data to measure pre/post changes in frequency and effectiveness in the use of open-source geospatial information. Our hypothesis is that analysts will more effectively utilize geographic information if provided with easily trainable, interoperable tools designed to systematically extract and store geospatial data from the Internet. This hypothesis was tested by collecting self-reported changes in ability or new insights gained after a 3-hour training session with the software.

5.2 Methodology

A training activity was conducted with individuals recruited from technical staff at Sandia National Laboratories. The aim was to make contact with subjects who have analytical needs similar to safeguards analysts but may not be making use of spatial information. Boin and Hunter [2] point out that while initial interviews of geospatial data users reveal many new themes (areas of concern, usability barriers, etc.), the rate of new themes occurring quickly disappeared after about ten interviews. We used 10 subjects in a single training session. This specific number is limited, based on the number of seats available in the computer training facility but appears adequate given the nature of our hypothesis.

Before the training activity, subjects were asked to fill out an online questionnaire to capture general demographic information, evaluate their feelings about the training activity, and determine their initial level of knowledge. After the training session, subjects were again asked to fill out an online questionnaire to determine any changes in ability as well as new insights into the geospatial data realm.

Quantitative results were tabulated from the ordinal survey questions and parametric statistical methods were applied. Paired t-tests and correlation coefficients were calculated using Microsoft Excel. Correlations between ordinal questions (on experience and background) and demographic groups (based on categorical questions) point towards factors effecting successful adoption of the software tool. Qualitative assessment of responses to open-ended text questions were made after the methods of Boin and Hunter [2], that is, categorization of themes by the research team followed by frequency analysis and summarization.

These testing methodologies were reviewed and approved by the Human Subjects Board for adherence to ethical human research practices.

5.3 Results

In spite of a small sample size, statistically significant results were apparent. Some were trivial: age of participants correlated highly with length of employment at Sandia. Training, on average, increased all participants‘ skill level by one category.

5.3.1 Quantitative Results

Other results reflected on institutional characteristics and culture. Analysts from Org. A were significantly more likely to use digital maps, satellite imagery, and geospatial information than those in the Org. B. Several reasons suggest themselves for this:

- Most of the analysts in Org. B are trained in ―old school" methods of research that did not include any paper maps or any digital maps.
- The projects are ―stove-piped" in that there is very little cross discussion between people so sharing of digital geographic data is very limited.
- Most staff in Org. B do not know how to access paper maps or digital geographic data resources, so they don‘t use them.
- Some sources of geographic data are hard to use, and the analysts are not trained to use them, so they don‘t use them.

Also, several of the analysts from Org. B were ―open-source" analysts who were formerly research librarians – they typically do not use maps (paper or digital). The others were analysts of a different sort, who used the open source data found by the research librarians.

We have found that unless a person or researcher knows and likes maps, generally, they will not use them. But for researchers who love maps (for recreational activity or because they have used them in their work) their analysis is generally geographic-data centric.

Other factors became apparent in our analysis of participant surveys. The GeoSafeguards tool was reported on average to be ―somewhat easy" to use and perceived software quality was rated ―high." Ease of use did not correlate with frequency of use of aerial photographs/satellite imagery (R^2= 0.09), GIS training (R^2= 0.08), age (R^2= 0.07), and use of laptops (R^2= 0.08) or mobile devices (R^2= 0.05).

80% of the participants plan to continue using GeoSafeguards in the future. Continued use did not correlate with frequency of use of digital maps (R^2= 0.04), aerial photographs/satellite imagery (R^2=0.09), and GIS skills (R^2= 0.02).

5.3.2 Qualitative Results

Table 1 lists in descending frequency of mention concerns and strengths identified during the training session. Not surprisingly, security concerns both for anonymous browsing and for protecting sensitive information topped the list. Some issues can be dealt with in a straight-forward way, such as by creating an installation script, revising the training material to include more time for case studies and open-source verification strategies, including a discussion of Zotero‘s timeline tool, or going into more depth about the disambiguation metrics.

Concerns about institutional support and foreign language queries are beyond the scope of our project, although usability of geospatial tools can inform decisions about agency software choices.

Issues	**Strengths**
Security of queries (anonymous browsing, firewalls)	Geospatial data visualization
Sensitive and classified information	GeoHack
Installation and setup	Integration
Veracity of query results	Bibliographic collections
More time for case studies	Use of Zotero‘s ‗Notes‘ field
Disambiguation	Applicability to vulnerability assessments
Temporal ribbon tool	
Lack of corporate support	
Language	
Increased scope of query results	

Table 1 Identified issues and strengths (in descending order of importance).

Clearly, the ability to visualize geospatial information buried within text documents and to have this integrated with the GeoHack tool is important to analysts. Merely providing ancillary capabilities like automating bibliographic processes and learning pro-tips like use of the ‗Notes‘ field for mapping also turned out to be important for our training participants.

6 CHALLENGES

6.1 Digital Footprint and OPSEC

Several security concerns have been noted when dealing with open source information and technologies that have been developed in the open source. A notable vulnerability of scouring the open Internet are the ―footprints" or digital browser signatures that are left on the servers from which data are accessed allowing the owners of the servers to deduce the identity and location of those who access the server. Users who are concerned about leaving a digital footprint while conducting open source searches are encouraged to work with cyber security specialists to address potential issues.

6.2 Data Quality of Open Sources

Evaluating the correctness and completeness of open source geospatial information is a complex task that has only recently been addressed. Volunteered Geographic Information (VGI) has become a very important source of geographical information used by both professionals and citizens [28]. One VGI dataset, OpenStreetMap [29] is a ―free wiki world map" that relies on users to add content and to make edits. The amount of information available on OpenStreetMap varies widely, because it relies on the interests of volunteers to add content. For example, most large metropolitain areas have extensive sets of data in OpenStreetMap, but the majority of the rural areas do not [30]. If the rural area is an area of interest for tourists, or has features of interest to others the area may have been mapped by an interested party. For our example in this paper, the Paks Nuclear Power Plant was chosen, and the OpenStreetMap data covers the area of the power plant in some detail.

In 2010, Zielstra and Zipf [31] did a comparison between OpenStreetMap and a commercial data set in Germany. In general, digital map datasets are compared to data of higher quality to determine data quality and accuracy. In this case, OpenStreeMap was compared to a proprietary commercial dataset created by the TeleAtlas for road data in Germany. The comparison was made for completeness of the data - where the commercial dataset was considered to be ―complete" while OpenStreetMap was compared to it. One measure of completeness was total length of street segments in both datasets, as well as measures of streets, roadways, pedestrian ways in individual locations. Statistics were calculated, and analysis performed. The study showed that OpenStreetMap had comparable data for roadways and pedestrian ways in urban areas in Germany, but in the rural areas, the data varied widely. The authors determined that OpenStreetMap can offer large amounts of digital map data, but that the data may not be appropriate for all applications. OpenStreetMap data does not contain attributes for routing purposes, or for navigation, but for a basic representation of roadways OpenStreetMap can serve as an input data layer if properly documented.

Finally, while open-source data can be an important supply of new types of information for safeguards analysis, it must be approached with some caution. Open-source data, especially crowdsourced information can be inaccurate, incomplete, biased or even fabricated [32].

7 CONCLUSIONS

In this study we investigated whether providing analysts with new information technologies that give them the ability to interact with information in a geospatial context could expand the types of information that experienced analysts used in their work without significantly increasing their workload. Previous research shows that humans organize and process a great deal of information based on a ―map-in-the-head" cognitive model [3]. We, therefore, suggest that providing geospatial tools to analysts may decrease their cognitive load and increase their effectiveness in handling heterogeneous information that have common geospatial dimensions [4]. To test this we built a suite of tools that allowed for efficient collection, organization and visualization of information. We then conducted a geospatial research training session and measured before and after use of open-source geospatial information. Our hypothesis was that analysts would more effectively utilize geographic information if provided with easily trainable, interoperable tools designed to systematically extract and store geospatial data from the Internet.

Using a phased search strategy that included general Internet search, geospatially-enabled search, and structured GIS data search, it is possible to assemble a basic geographically referenced set of data without a specialized GIS analyst or expensive GIS software.

Because of the ease of use and low life-cycle costs, the use of the GeoSafeguards tool to create a basic geospatially-referenced data set has the potential to increase the use of geospatially referenced data in future safeguards analysis. When configured to work within an existing safeguards analysis workflow, this tool can allow analysts to efficiently and effectively utilize both structured and unstructured geospatial data from the open Internet, a capability that generally is available only to those with specialized training and expensive, proprietary tools.

We conclude that by enabling safeguards analysts to efficiently and effectively extract and utilize geospatially-referenced information from the Internet, these analysts will more often use these data in their analyses to provide more complete and accurate results.

7.1 Derivative Works

This research and the GeoSafeguards software system have been widely published and presented. Appendix A lists the presentations and their abstracts. Associated papers were published in the corresponding proceedings.

Currently, the research team is working with others both internal and external to Sandia National Laboratories who have shown an interest in these capabilities and findings. These have included the Technical Analysis Department, Proliferation Sciences Department, the USG Open Source Center, DOE/HQ, DTRA, and IAEA Safeguards Information Management (SGIM) department.

As a beneficial side-effect of the presentations made over the last two years, these professional interactions led the researchers to form the Open Source / Geospatial Information Working Group within the Safeguards Technical Division of the Institute of Nuclear Material Management (INMM). Now numbering over two dozen participants at dozens of institutions worldwide, the working group has created a LinkedIn online forum to strengthen and maintain professional relations concerning the topic of open source information and open source tools.

7.2 Long-term Direction

To fully realize the efficient use of this heterogeneous and unstructured geographically referenced material, more work needs to be done to create standardized and automated processes for discovering, integrating and organizing the data. Development and application of ontologies and semantic technologies will be necessary to achieve this goal.

While open-source data can be an important supply of new types of information for safeguards analysis, it must be approached with some caution. A future goal of this research is to develop tools and methodologies that provide safeguards analysts the ability to differentiate valid and reliable geospatial data from those data that cannot be trusted.

The authors are currently going through the process of copyright review for GeoSafeguards, which will be required if the software is to be exported (for example, to the IAEA) or eventually released into the public domain.

The research team, along with others, hopes to build upon the current toolset to develop a prototype semantic-web system. Such a system would be capable of consuming all safeguards relevant information, organizing it based upon semantic content of the information along three dimensions of safeguards information (technical, geospatial and temporal) [1], and providing the information to appropriate users in usable forms based upon need-to-know, role-based access. This prototype system will be able to interface and serve other analytical tools with structured and semantically marked information.

Semantic information technologies move information systems from presentation-based representations (in which the computer systems simply present information to humans) to meaning-based representations (with embedded, computer readable metadata). Semantic information technology will enable the development of more sophisticated, knowledge-based analytical solutions.

To address potential security concerns and to broaden the type of information that can be utilized within the GeoSafeguards tool we are examining options for creating in-house capabilities to supplant the functionality provided by the web services. This could potentially allow GeoSafeguards to operate without an Internet connection and would not send information to remote servers, opening up the possibility to utilize sensitive information within the tool.

8 REFERENCES

[1] R. Hooper, ―Seminar on Strengthened Safeguards," Brookhaven National Laboratories, Nov-2011.

[2] A. T. Boin and G. J. Hunter, ―What communicates quality to the spatial data consumer," *Quality Aspects in Spatial Data Mining*, pp. 140–147, 2007.

[3] D. M. Mark, C. Freksa, S. C. Hirtle, R. Lloyd, and B. Tversky, ―Cognitive models of geographical space," *International Journal of Geographical Information Science*, vol. 13, no. 8, pp. 747–774, Dec. 1999.

[4] R. L. Bunch and R. E. Lloyd, ―The Cognitive Load of Geographic Information," *The Professional Geographer*, vol. 58, no. 2, pp. 209–220, 2006.

[5] J. Baute, ―Strengthening detection capability for safeguards," INMM-ESARDA, Santa Fe, NM, USA, 2005.

[6] ―Ushahidi." [Online]. Available: http://ushahidi.com/. [Accessed: 17-Sep-2012].

[7] ―Ushahidi: crowdmapping collective that exposed Kenyan election killings," *the Guardian*, 07-Apr-2011. [Online]. Available: http://www.guardian.co.uk/news/blog/2011/apr/07/ushahidi-crowdmap-kenya-violence-hague. [Accessed: 17-Sep-2012].

[8] ―Model Protocol Additional to the Agreement(s) Between State(s) and the International Atomic Energy Agency for the Application of Safeguards." International Atomic Energy Agency.

[9] B. Jasani, B. Richter, G. Stein, M. D. Ward, and M. Killinger, *Enhancing IAEA safeguards using commercial satellite imagery: A pilot study*. King's College London Department of War Studies, 1996.

[10] S. B. Liu, A. A. Iacucci, and P. Meier, ―Ushahidi Haiti and Chile: Next Generation Crisis Mapping," *ACSM Bulletin*, no. 246, Aug-2010.

[11] R. Wallace, G. Anzelon, and J. Essner, ―Safeguards information from open sources," *JNMM*, vol. 37, no. 4, pp. 30–40, 2009.

[12] R. R. Vatsavai, B. Bhaduri, A. Cheriyadat, L. Arrowood, E. Bright, S. Gleason, C. Diegert, A. Katsaggelos, T. Pappas, R. Porter, and others, ―Geospatial image mining for nuclear proliferation detection: Challenges and new opportunities," in *International Geoscience and Remote Sensing Symposium*, 2010, pp. 48–51.

[13] I. Niemeyer, S. Nussbaum, and M. J. Canty, ―Automation of change detection procedures for nuclear safeguards-related monitoring purposes," in *International Geoscience and Remote Sensing Symposium*, 2005, vol. 3, p. 2133.

[14] ―The Open Source Definition | Open Source Initiative." [Online]. Available: http://www.opensource.org/docs/osd. [Accessed: 27-May-2011].

[15] ―HOST ::: Cyber Security R & D Center." [Online]. Available: http://www.cyber.st.dhs.gov/host.html. [Accessed: 27-May-2011].

[16] M. Barletta, A. El Gebaly, W. Hammond, T. Lorenz, S. Robb, N. Zarimpas, and R. Zarucki, “Open Source Information Acquisition and Analysis in the IAEA Department of Safeguards,” presented at the 33rd ESARDA Annual Meeting, Budapest, Hungary, 2011.

[17] “Zotero.” [Online]. Available: http://www.zotero.org/. [Accessed: 12-Apr-2011].

[18] “OpenLayers: Home.” [Online]. Available: http://openlayers.org/. [Accessed: 18-Apr-2011].

[19] “Welcome to the OGC Website | OGC(R).” [Online]. Available: http://www.opengeospatial.org/. [Accessed: 21-Apr-2011].

[20] “Yahoo! GeoPlanet™ - YDN.” [Online]. Available: http://developer.yahoo.com/geo/geoplanet/. [Accessed: 20-Apr-2011].

[21] “GeoNames.” [Online]. Available: http://www.geonames.org/. [Accessed: 20-Apr-2011].

[22] “GeoHack - Toolserver wiki.” [Online]. Available: https://wiki.toolserver.org/view/GeoHack. [Accessed: 05-Jun-2012].

[23] “Paks Nuclear Power Plant,” *Wikipedia*. [Online]. Available: http://en.wikipedia.org/wiki/Paks_Nuclear_Power_Plant. [Accessed: 04-Apr-2011].

[24] “Paks Nuclear Power Plant Ltd. - Home.” [Online]. Available: http://paksnuclearpowerplant.com/. [Accessed: 13-Apr-2011].

[25] “Wikimapia - Let‘s describe the whole world!” [Online]. Available: http://wikimapia.org/#lat=35.1025&lon=-106.6117&z=10&l=0&m=b. [Accessed: 13-Apr-2011].

[26] “GeoHack - Paks Nuclear Power Plant.” [Online]. Available: http://toolserver.org/~geohack/geohack.php?pagename=Paks_Nuclear_Power_Plant&params=46_34_21_N_18_51_15_E_region:HU_type:landmark. [Accessed: 04-Apr-2011].

[27] J. Nielsen and J. Levy, “Measuring usability: preference vs. performance,” *Communications of the ACM*, vol. 37, no. 4, pp. 66–75, 1994.

[28] M. Goodchild, “NeoGeography and the nature of geographic expertise,” *Journal of Location Based Services*, vol. 3, no. 2, pp. 82–96, 2009.

[29] “OpenStreetMap.” [Online]. Available: http://www.openstreetmap.org/. [Accessed: 11-Sep-2012].

[30] M. Haklay, “How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets,” *Environment and planning. B, Planning & design*, vol. 37, no. 4, p. 682, 2010.

[31] D. Zielstra and A. Zipf, “A comparative study of proprietary geodata and volunteered geographic information for Germany,” in *13th AGILE International Conference on Geographic Information Science*, 2010, vol. 2010.

[32] A. Berriman, R. Leslie, and J. Carlson, “Information Analysis for IAEA Safeguards,” in *Proceedings of the 45th INMM Annual Meeting*, Orlando, 2004.

Appendix A: Work Products Derived from this Research

Listing of OSIV Papers and Presentations					
Title	**Doc No**	**SAND NO**	**Type**	**Date**	**Description**
Open Source Geographic Information for Safeguards Analysis	5288650	2010-7745 P	Fact Sheet	10/26/2010	Fact Sheet 1
Open Source Geographic Information for Safeguards Analysis	5289406	2010-8292 A	Abstract	11/16/2010	
Harvesting Geospatial Safeguards Information with Open Source Tools	5291071	2011-0534 C	Conference Paper	1/24/2011	INMM July 2011
Open Source Geographic Information for Safeguards Analysis	5293808	2011-2719 P	Presentation	4/26/2011	Open Source Center Briefing
Open Source Geographic Information for Safeguards Analysis	5294151	2011-3018 C	Conference Paper	4/22/2011	ESARDA
Open Source Geographic Information for Safeguards Analysis	5294728	2011-3260 C	Conference Paper	5/9/2011	ESARDA PDF of Prezi
Harvest Geospatial Safeguards Information with Open Source Tools	5295552	2011-4025 C	Conference Paper	6/6/2011	INMM July 2011
Harvesting Geospatial Safeguards Information with Open Source Tools	5296579	2011-4886 C	Conference Paper	7/7/2111	INMM July 2011
Harvesting Geospatial Safeguards Information with Open Source Tools	5293808	N/A	Presentation	8/16/2011	DOE GIS Summit
Exploiting the Geospatial Dimension of Data in Support of IAEA Safeguards	5299623	2011-7107 J	Journal Paper	9/15/2011	ESARDA Journal (never published)
Advances in the Use of Open-source Information	5300204	2011-8305 C	Conference Paper	9/29/2011	INMM-ESARDA
Advances in the Use of Open-source Information	5300207	2011-8303 C	Presentation	9/29/2011	INMM-ESARDA Pdf of Prezi
Harvesting Open-Source Geospatial Information with Open-Source Tools	5300565	N/A	Presentation	10/10/2011	GEOINT 2011 Programatic Review
Open-Source and Geospatial Information and Software Tools for	5303523	2012-0771 C	Conference Paper	1/19/2012	INMM July 2012

Safeguards Analysis					
Open-Source/Geospatial Information Working Group	5303524	2012-0770 C	Conference Paper	1/19/2012	INMM July 2012
Tools for Open-Source Geospatial Information	5306527	N/A	Presentation	4/5/2012	Training Presentation
Open-Source/Geospatial Information Working Group	5308472	2012-4794 C	Conference Paper	5/23/2012	Working Group Charter
Open-Source Geospatial Information and Software Tools for Safeguards Analysts	5309152	2012-4909 C	Conference Paper	6/8/2012	INMM July 2012
Open-Source/Geospatial Information Working Group	5309792	2012-5949 P	Poster	6/26/2012	Poster for INMM July 2012
Open-Source Geospatial Information and Tools for Safeguards Analysts	5310181	2012-5656 C	Conference Paper	7/9/2012	INMM July 2012
GeoSafeguards Overview	5312021	Pending	Presentation	8/30/2012	DOE Headquarters
Factsheet 2		Pending	Fact Sheet		

Factsheet, version 1
"Open Source Geographic Information for Safeguards Analysis"

No abstract available (in both 8.5x11 and A4 formats).

ESARDA Presentation and Paper, May 2011
"Open Source Geographic Information for Safeguards Analysis"

In this era of user-generated Web content, geographically referenced information is being published to open sources at an astounding rate. One might conceptually understand these data as the product of a distributed, decentralized sensor network capable of detecting the geographic signals of nuclear proliferation. Within an information-driven safeguards regime, these data, (often created and shared by common citizens) can be invaluable to the detection of undeclared nuclear activity. Such information, however, is often overlooked and underutilized because, at present, no tools exist to systematically and efficiently extract and utilize these data. This work seeks to enable safeguards analysts to efficiently and effectively use open source geospatial information by leveraging web-based information technologies in novel ways.

While a great deal of geospatial data are published in well defined, easily detectable formats, most data are unstructured, heterogeneous and complex. By implementing geospatial and domain-specific ontologies, these data can be detected and converted into usable and semantically interoperable formats that can be effectively incorporated into an analyst's work. We are working closely with safeguards analysts and other stakeholders to establish high-level requirements and derive use cases to ensure that these tools are integrated into analysts' existing workflow for efficient use and high adoption.

INMM Presentation, July 2011
"Harvesting Geospatial Safeguards Information with Open Source Tools"

Because of the inherently geographic nature of the processes of nuclear proliferation, successful analysis requires appropriate tools for collecting and utilizing geographically referenced data to supplement safeguard activities. Although geographic information systems or GIS-based tools are currently being developed for use in a wide variety of nuclear safeguard activities including site inspection, verification and wide-area environmental sampling, analysts tasked with gathering information to support safeguards activities frequently lack the capabilities and tools necessary for extracting and making efficient use of geospatial data from open sources, notably the Internet.

The authors have surveyed a number of open source GIS tools with the goal of integrating and/or modifying them to create an easily learned, interoperable toolset. Such a toolset is designed to systematically extract, organize, and store geospatial data from the Internet. Because open source software is by definition fully available to developers, it is ideally suited to customization to better fit specific use-cases, such as nonproliferation analyses.

This paper will summarize our results to date to evaluate these tools and integrate them into the analyst's workflow. Existing and emergent information systems will be discussed, especially our use of a wide spectrum of Internet resources and an enhanced, customized version of Zotero, an open source reference management system..

DOE GIS Summit, August 2011
"Harvesting Geospatial Safeguards Information with Open Source Tools"

No abstract available.

GEOINT Presentation, September 2011
"Harvesting Open-Source Geospatial Information with Open-Source Tools"

No abstract available.

ESARDA-INMM Presentation, October 2011
"Advances in the Use of Open-source Information"

Great advances have been made in the application of open-source information for Safeguards. However, the open-source ―ecosystem" is rapidly evolving. Recently, the combination of powerful smartphones and the World Wide Web have led to novel developments for mapping the impact of natural and man-made disasters. Researchers are learning to harness online volunteers as citizen scientists. Online social media such as Facebook, LinkedIn, and Twitter are becoming accepted channels of credible information. Geospatially-aware applications (augmented reality apps) are capable of overlaying data upon a smartphone's view screen.

These and other developments will have a significant effect on the collection, evaluation, structuring, analysis, and dissemination of safeguards-relevant information. In addition, the modern, open Internet is posing new security threats—not just viruses and their ilk, but threats based on digital traces left by browsing the open Web.

The authors present a case study of the disastrous toxic flood near Devecser, Hungary in October 2010 and then give recommendations for the future use of open-source information and software tools.

INNM Presentation, July 2012
"Open-Source Geospatial Information and Software Tools for Safeguards Analysts"

Interim results from a two-year study at Sandia National Laboratories concerning open source geospatial data and open source software are now available. An easily trainable, interoperable toolset has been designed to systematically extract and store geospatial data from the Internet. Testing is underway to measure the utility of this integrated suite of open-source tools for geospatial data collection, analysis, and visualization. Test subjects have been initially surveyed, trained on the use of the software suite, and follow-up interviews are being conducted. Because special care was taken to integrate the system into the analyst_s existing workflow, initial reports point toward frequent acceptance. Variability in reported user experience is associated with several key attributes among test subjects. Based on this, we have created a profile of the user/analysts that will most benefit from this software system.

Demonstration to IAEA SGIM, August 2012

No abstract available.

Presentation to DOE/HQ, August 2012
"GeoSafeguards Overview"

No abstract available.

Factsheet, version 2

No abstract available.

Appendix B: Training Materials

The full complement of training material is comprised of:

Part I – Introduction
Part II – Open Source and GEOINT
Part III – The GeoSafeguards Tool
Part IV – Using GeoSafeguards
Part V – Case Study
Part VI – Self Study

The most current version of these materials is available within the Sandia internal network at https://snl-wiki.sandia.gov/display/OSIV/Training. Outside of Sandia National Laboratories, please contact the authors.

Appendix C: GeoSafeguards Structure Diagram

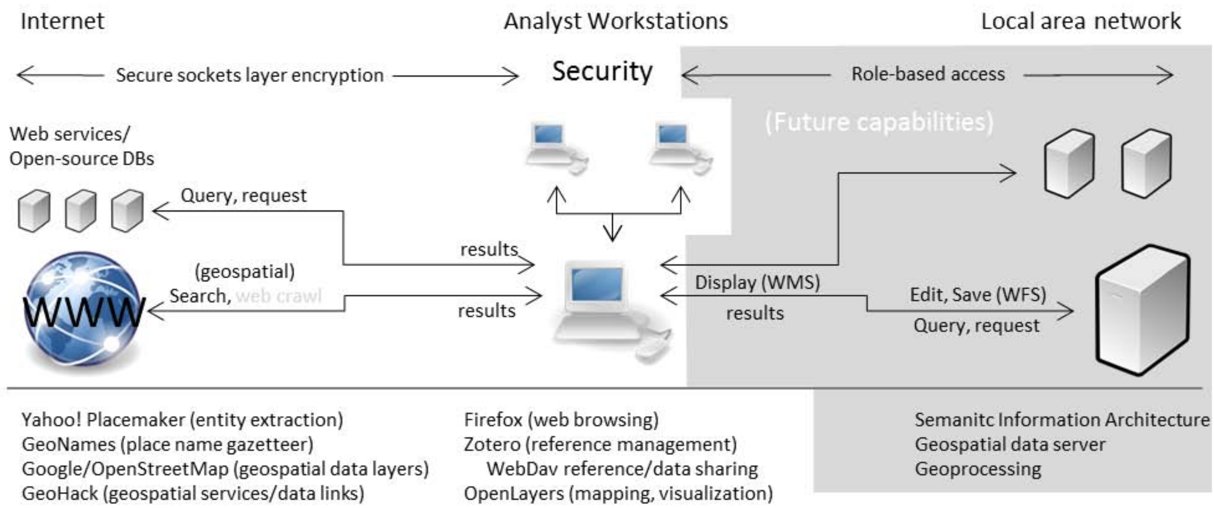

Figure 7 GeoSafeguards architecture and future capabilities

Appendix D: Screen Shots of the GeoSafeguards Tool

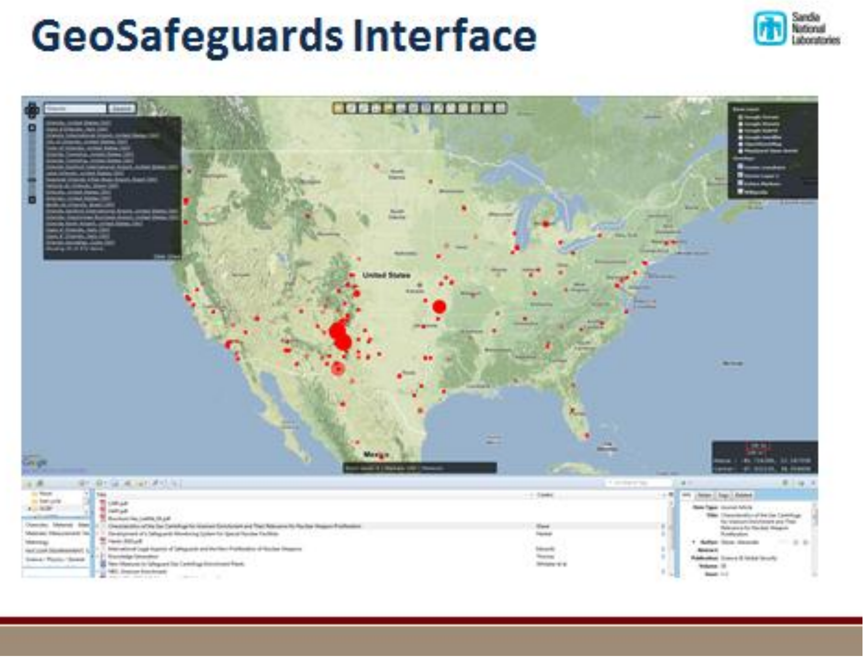

Figure 8 Mapping Interface with Zotero

Figure 9 GeoSafeguards Mapping Interface: Base layers

Figure 10 GeoSafeguards Mapping Interface: Toolbar

Figure 11 GeoSafeguards Mapping Interface: Place name search

Figure 12 GeoSafeguards Mapping Interface: Scale bar and coordinates

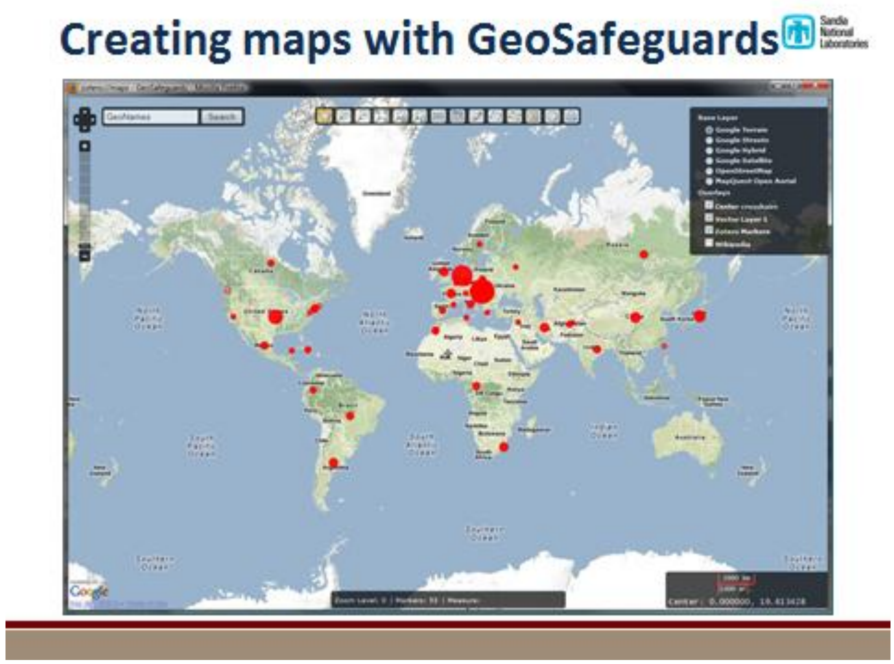

Figure 13 Creating Maps with GeoSafeguards

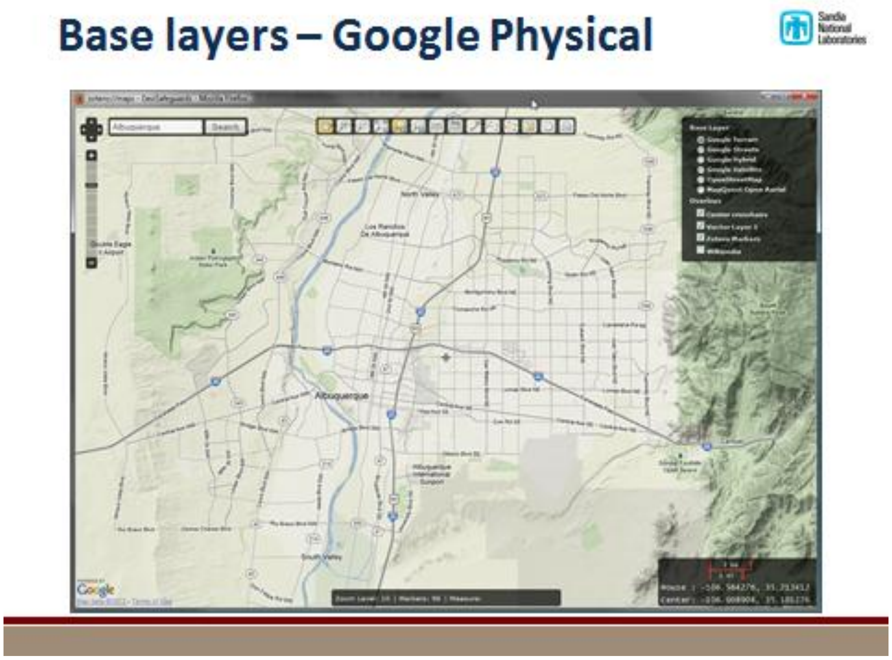

Figure 14 Base Layers – Google Physical

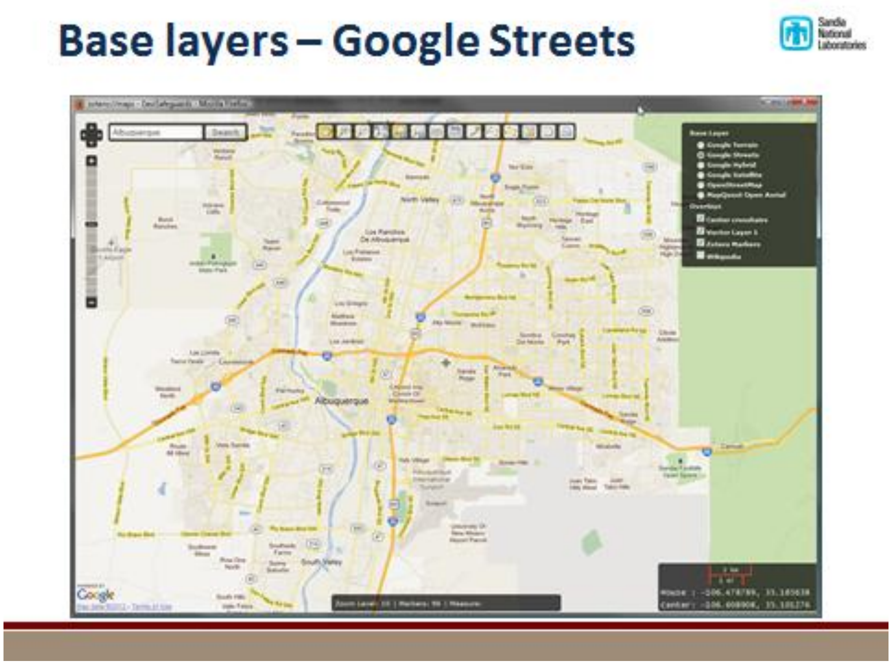

Figure 15 Base Layers – Google Streets

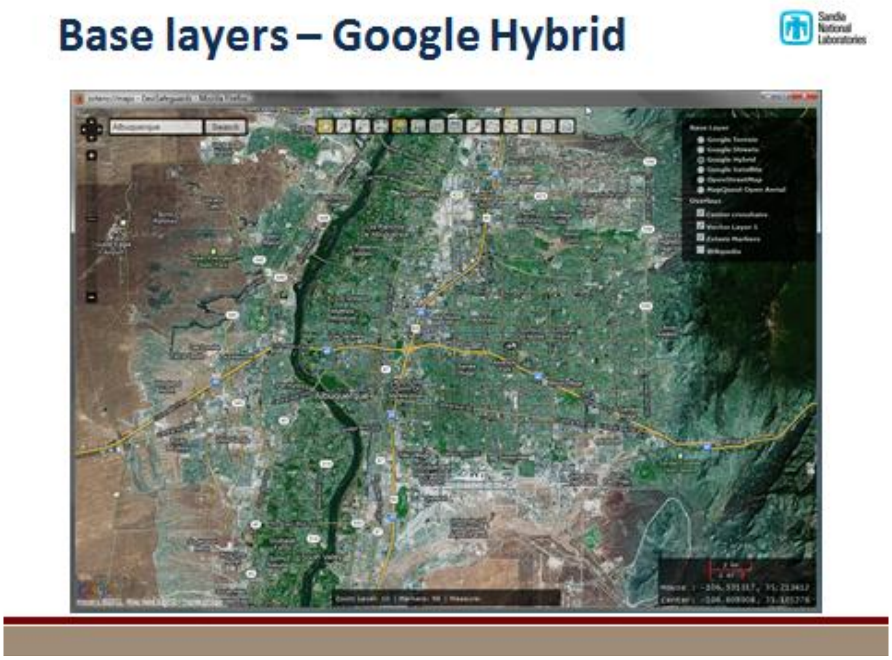

Figure 16 Base Layers – Google Hybrid

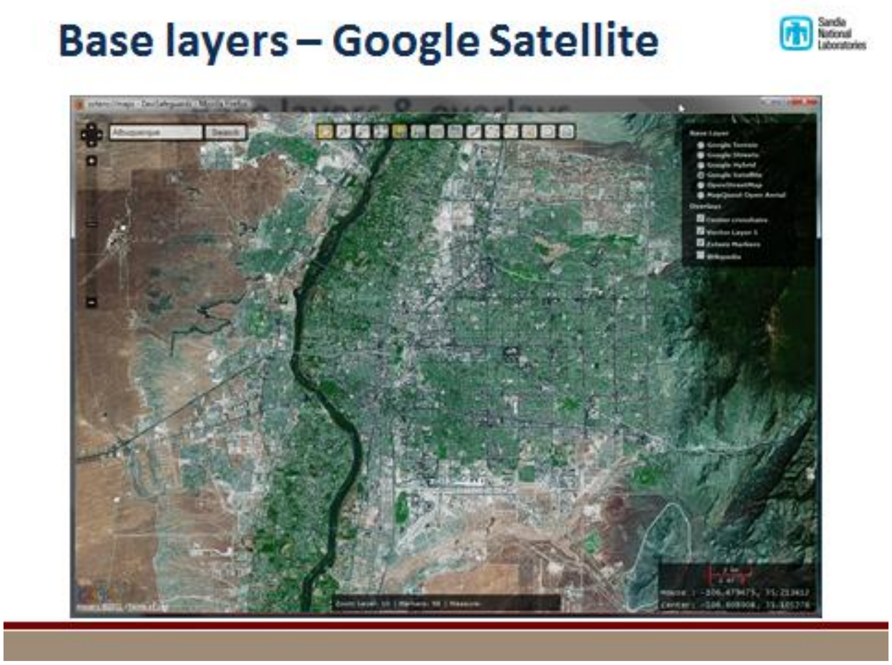

Figure 17 Base Layers – Google Satellite

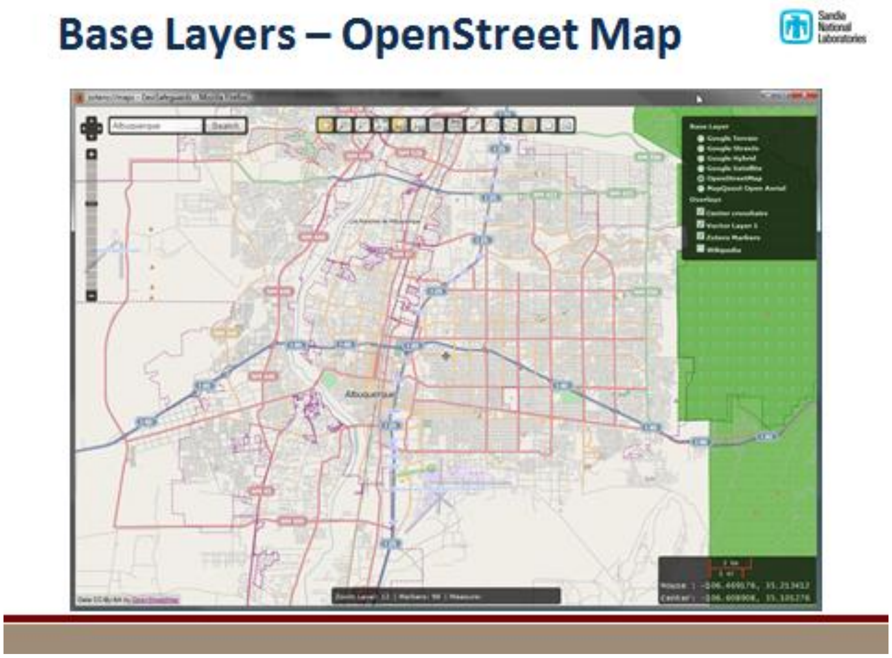

Figure 18 Base Layers – Open Street Map

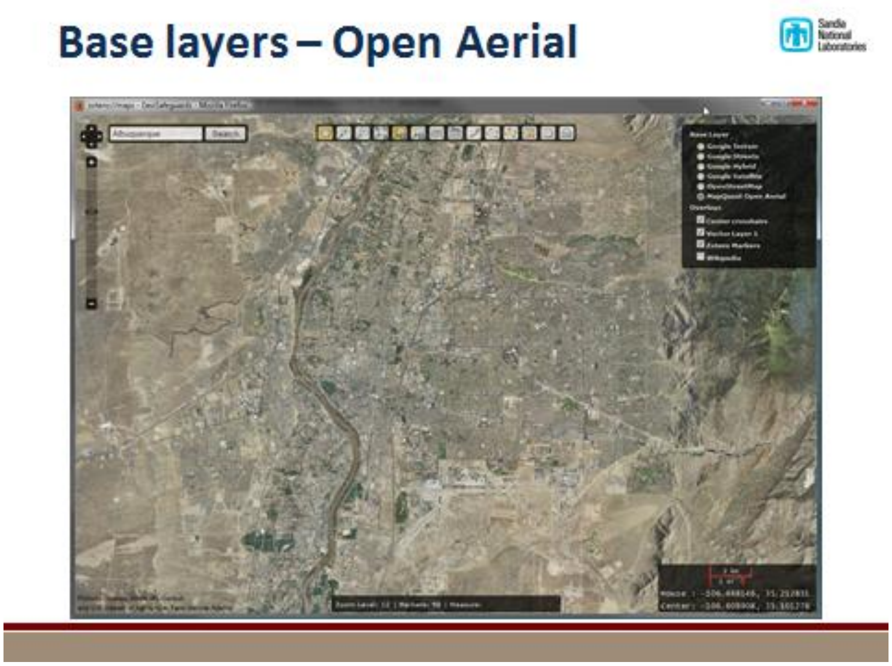

Figure 19 Base Layers – Open Aerial

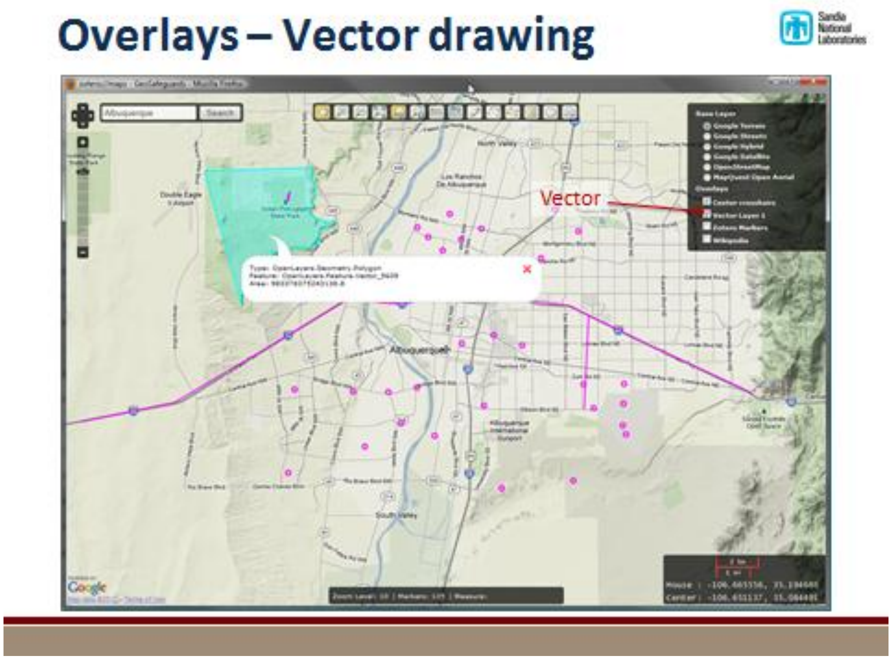

Figure 20 Overlays – Vector Drawing

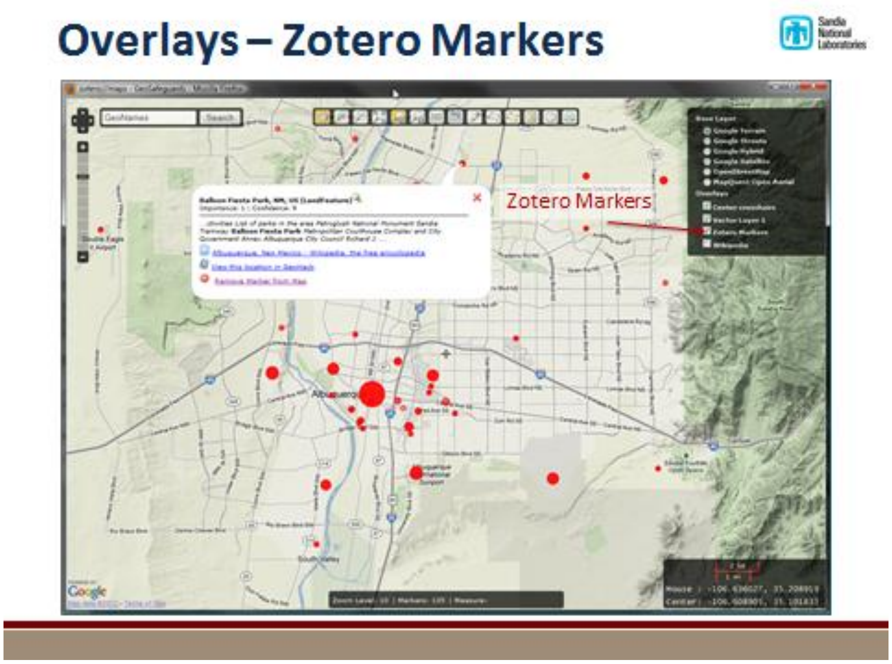

Figure 21 Overlays – Zotero Markers

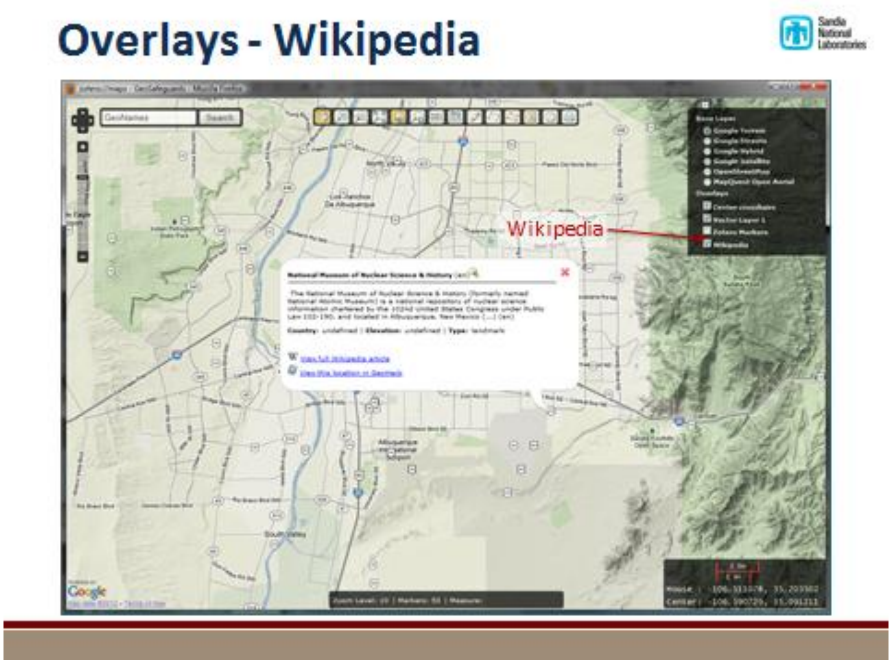

Figure 22 Overlays – Wikipedia

Distribution

1	MS 0161	Legal Technology Transfer Center, 11500 (electronic copy)
1	MS 0359	D. Chavez, LDRD Office, 1911 (electronic copy)
1	MS 0899	Technical Library, 9536 (electronic copy)
1	MS 1218	D. Bleakly, 5944 (electronic copy)
1	MS 1363	B. Marshall Jr., 6830 (electronic copy)
1	MS 1371	D. Blair, 6832 (electronic copy)
1	MS 1371	K. Horak, 6832 (electronic copy)
1	MS 1371	M. McDaniel, 6832 (electronic copy)
1	MS 1375	R. Wilson, 6800 (electronic copy)

Sandia National Laboratories