



An interactive ontology-driven information system for simulating background radiation and generating scenarios for testing special nuclear materials detection algorithms

Alexandre Sorokine ^{a,*}, Bob G. Schlicher ^a, Richard C. Ward ^a, Michael C. Wright ^a, Kara L. Kruse ^a, Budhendra Bhaduri ^a, Alexander Slepoy ^{b,1}

^a Oak Ridge National Laboratory, 1 Bethel Valley Rd., Oak Ridge, TN 37831-6085, United States

^b Lawrence Berkeley National Laboratory, Berkeley, CA 94720-8099, United States



ARTICLE INFO

Article history:

Received 29 October 2014

Received in revised form

31 March 2015

Accepted 20 April 2015

Keywords:

Ontology

Ontology driven information system

Scenario

Geographical information system

Database

ABSTRACT

This paper describes an original approach to generate scenarios for the purpose of testing the algorithms used to detect special nuclear materials (SNM) that incorporates the use of ontologies. Separating the signal of SNM from the background requires sophisticated algorithms. To assist in developing such algorithms, there is a need for scenarios that capture a very wide range of variables affecting the detection process, depending on the type of detector being used. To provide such a capability, we developed an ontology-driven information system (ODIS) for generating scenarios that can be used for testing of algorithms for SNM detection. The Ontology-Driven Scenario Generator (ODSG) is an ODIS based on information supplied by subject matter experts and other documentation. The details of the creation of the ontology, the development of the ontology-driven information system, and the design of the web user interface (UI) are presented along with specific examples of scenarios generated using the ODSG. We demonstrate that the paradigm behind the ODSG is capable of addressing the problem of semantic complexity at both the user and developer levels. Compared to traditional approaches, an ODIS provides benefits such as faithful representation of the users' domain conceptualization, simplified management of very large and semantically diverse datasets, and the ability to handle frequent changes to the application and the UI. The approach makes possible the generation of a much larger number of specific scenarios based on limited user-supplied information.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Detection of special nuclear material (SNM) represents one of the most important challenges of national security. Reliable SNM detection is critical for the enforcement of nuclear non-proliferation international agreements, preventing nuclear terrorism and illicit trafficking of SNM (ITDB, 2014; Runkle et al., 2009; Murauskaitė, 2014). But as noted by Runkle et al. (2009) "it is a daunting technical challenge to identify special nuclear materials via gamma ray detection." From the physics point of view detection of SNM poses a difficult task as its photon signature has to be extracted from the background radiation (Runkle et al., 2009). The background radiation field varies due to a very large number of factors such as geographic location, geology, human activity, weather, and others. The background radiation is generated by macroscopic objects, like the

materials used in buildings or anthropomorphic sources such as individuals receiving cancer or diagnostic treatment, is highly situation specific, and cannot be well-represented by a generic instance. Algorithms used for extracting SNM signatures, such as GADRAS (Medalia, 2010), must be tested and validated to assure their quality, i.e., low rate of mis detections and false alarms under diverse background radiation conditions. Due to high variability of environments and high costs or regulatory impracticality of field radiation measurements with significant quantities of SNM, testing of the algorithms is most effectively achieved in simulated environments using synthetic data.

Separating the SNM signal from the background requires sophisticated algorithms because SNM signal may be masked by other sources of radiation such as medical isotopes (Wald, 2014). Search mode algorithms in an urban setting are hampered by a structured background radiation that resists simple noise models. To assist in developing such algorithms, there is a need for scenarios which capture a very wide range of variables affecting the detection process, depending on the type of detector being used. Each urban segment produces a significantly unique signature background, complicated by the dynamic entry and exit of cars, trucks, and pedestrians. A

* Correspondence to: One Bethel Valley Rd., P.O. Box 2008, MS-6017, Oak Ridge, TN 37831-6017, United States. Tel.: +865 574 4966.

E-mail addresses: sorokina@ornl.gov (A. Sorokine), aslepoy@lbl.gov (A. Slepoy).

¹ Deceased March 19, 2013

particular instance of such a setting can only be described by a *scenario*, i.e., description of a sequence of events and a setting where these events occur. Our goal was to use ontologies to expand the number and variety of scenarios generated, and thus to provide algorithm developers and testers with an unprecedented level of completeness in covering the space of combinatorial variability available.

As a result of our effort we have developed a system, the Ontology-Driven Scenario Generator (ODSG) that generates a very large number of scenarios from a small amount of user input, using parameters that describe a particular process of searching for SNM. To accomplish this task we based software development on the concept of an ontology-driven information system (ODIS) (Guarino, 1998; Fonseca and Egenhofer, 1999; Fonseca, 2007). The ODSG automatically converts a concise description by the user into a large number of scenarios consistent with the user-defined requirements. The system is based on a geographical information system (GIS, Wikipedia 2014) utilizing both top- and domain-level ontologies to facilitate the framework. The scenarios are based on both real and synthetic data, and represent a realistic sequence of events in a realistic urban setting. The collection of scenarios can be used to benchmark and train search algorithms, and evaluate their performance with respect to a particular scenario type.

By using the ODIS development paradigm we addressed the problem of semantic complexity both at the developer and user level. A description of the problem in terms of numeric parameters of the setting alone limits the description to a particular narrow semantic frame. For example, to set the positions of trees, or the make and model of vehicles, such entities must be postulated as present or absent in the scene. The existence of these in the scenario must be provided by a semantic description. It is conceivable that one could include all possible objects and all possible relationships between objects in an over-arching model, and then use this model to create a strictly parametric description, but such a practice requires large numbers of constraints to be useful and such constraints would constitute a de facto ontology.

ontology-driven information systems (ODIS) and Ontology-Driven Architecture for Software Engineering (ODASE) have been discussed in the literature for at least a decade (Guarino, 1998; Fonseca and Egenhofer, 1999; Fonseca, 2007; Bossche et al., 2007; Akerman and Tyree, 2006; others). The distinctive property of these paradigms is the reliance on ontologies to accomplish multiple goals during system development and operation. Recent advances in ontologies and other kinds of semantics technology are turning ODIS and ODASE into practical tools capable of addressing a number of challenges of GIS development. Placing ontologies at the core of GIS and GIS development offers a promise of more effective geographic information integration, reduced costs of development, streamlined and coherent user interfaces, and, most important, adequacy of the information system to the users' conceptualization of the domain.

Ontologies played a pivotal role at all stages of the development of ODSG and are indispensable for its operation. During system development we relied on ontologies and ontology-based techniques to formalize application domain and expert knowledge. At the system configuration stage, ontologies are used to link disparate datasets and simulation models and to generate the web user interface (UI). At run time ontologies are used for database schemata translation, workflow planning, reasoning about missing data, and presentation of the results. The overall architectural approach can be described as mapping between ontologies and database schemata and traversing ontology categories to pull the most appropriate data from known datasources.

An important part of the process of developing software to generate scenarios is gathering expert opinion on the variables to be included in the simulated environment. The variables should be extensive enough to properly define the complexity of radiation detection

in urban environments accounting for the presence of legitimate sources such as construction materials, patients who have received medical isotopes for diagnosis or treatment, shielding, natural sources, and enhanced NORM in all likely geometric configurations.

We discuss existing solutions to automated generation of scenarios in Section 2. Our proposed solution, increasing the number of scenarios by using an ODIS framework, is presented in Section 3. Section 4 discusses results of example scenarios generated. Finally in Section 5, we conclude with the lessons learned and potential for application of ODSG in other subject areas.

2. Existing solutions for automated scenario generation

For purposes of examining the existing solutions for automated scenario generation, we define scenarios as *a time sequence of events describing an activity or process*. In this section we will review common approaches to automated generation of scenarios. Scenarios are used in many areas to generate synthetic sequences of events including training/planning, resource allocation, estimating risks, evaluating physical models, etc. Martin and Hughes (2010) describe various approaches to implementing automated scenario generation including: 1) seeded approaches that blend human-created elements and system-automated processes; 2) heuristic approaches that create scenarios and their subcomponents by randomly selecting components and then comparing the scenario against the heuristics (Grois et al., 1998); 3) enumeration approaches which are a variant of the heuristic approaches in which all possible scenarios are enumerated; and 4) procedural modeling. The last approach was first popularized in computer graphics to create visual models, textures, and/or animations where rules and symbols were used to represent the elements being modeled. Examples of the procedural rules encoding are "shape grammars", L-systems (Lindenmayer systems) or Functional L-systems (Martin and Hughes, 2010).

An example of the use of procedural rules is the generation of scenarios for training emergency rescue crews (Hullett and Mateas, 2009). The specific goal was to generate scenarios for collapsed structure rescue. To develop the domain model, the authors used a hierarchical task network (HTN), which was derived from procedural methods such as L-systems to display geometric models of plants and buildings. As in the earlier reference to work by Martin and Hughes, the approach of Hullett and Mateas builds concepts into rules that can quickly implement the conceptual objects as needed. While the goal was to generate a large number of "levels by randomly combining elements," the authors found that "some plans result in the same scenario by performing the same steps in different order." Furthermore, as the size of the scenario grows, "the combinatorial number of possible variations grows...". The researchers found that an approach of creating variation by "changing the specified goals" created a wider variety of scenarios than using random generation techniques. The authors also note that a "human author is responsible for the knowledge engineering work required to specify both methods (tasks) and operators" (Hullett and Mateas, 2009).

Grois et al. (1998) used a Bayesian-network based model for generating scenarios for crisis management training. The software developed, called *ScenGen*, was based on Noisy-OR Bayesian Networks (NOBNs). The authors then compared their software to three conventional alternative approaches: 1) manual generation of scenarios by a human expert; 2) randomly-instantiated scenarios created by selecting a random initial event (corresponding to a root node of the NOBN net and propagating the network forward); and 3) case-based stochastic perturbation, which is a hybrid of the first two approaches. In this third traditional approach, human subject matter experts create seeds for the network, perturbing them "so as to obtain variability." The authors comment that in using their new approach "the very selection bias that gives *ScenGen* its power also significantly limits its

accessibility to the many portions of the search space.” Compared to the traditional approaches, “it is the first to exhaust its space accessible scenarios” (Gros et al., 1998).

Reynolds (2001) takes a particularly novel conceptual approach described as a “framework for scenario generation.” The application area is finance and a typical objective is to “assess the credit riskiness of the portfolio on an annual basis for the next 10 years.” The framework for scenario generation includes four components:

1. **Blocks** are a group of risk factors with similar statistical properties. Blocks are created from the set of risk factors affecting the portfolio.
2. **Models** define the distribution or evolutionary process for a **risk factor**. Specifies a calibration method for obtaining all model parameters from historical or other data.
3. **Scenario Generator** is a fully calibrated model for generating scenarios that link **Models** and **Blocks** and defines the relationships among **risk factors**.
4. **Scenario Set Definition** specifies the details of creating the actual scenario set, e.g., number of scenarios, trigger times (future points of interest), and a description of the scenario set.

While no mention is made in Reynolds (2001), basing this approach on ontologies would provide the terminology needed for establishing the **Blocks** and the **Models** and ontologies could also provide the **relationships** among **risk factors** that are required for the scenario generator. An ontological basis for such a “framework for scenario generation” would generalize and strengthen this otherwise exceptional guideline for development of scenarios.

After a careful review of previous efforts on automated scenario generation, Martin and Hughes (2010) note that “often scenario generation systems us a ‘stovepipe’ approach, which focuses only on a single simulation system.” They propose that a better approach “would be to employ a robust data model that describes the rules and relationships required for each simulation component (Koziarz et al., 2003), allowing multiple simulation systems to be supported by a single scenario generation system and enabling the system itself to be extensible to multiple training domains.” It is just this goal that we address in our effort to build a flexible, extensible framework for scenario generation based on ontologies.

Thus while an analysis of the literature on automated scenario generators turned up many innovative approaches, these approaches lacked the ability to provide satisfactory coverage of most of the phase space of the variables, typically use a fixed set of variables, depended on frequent subject matter expert oversight, and incurred a high cost for adding alternative sets of scenarios.

3. Proposed solution: ontology-driven scenario generation

To overcome limitations of the existing scenario generation techniques we developed our own architecture built around ontology and semantics technologies. To reiterate, by scenario we mean a model of a sequence of events and a background setting. In the context of SNM detection a sequence of events includes the movement of a searcher with a detector and optionally the movement of other mobile objects such as cars or people in a setting. A setting is an area that can be covered by a searcher within some allocated time. For simplicity in the present version of ODSG is limited to urban areas in the conterminous US and the size of the search area was limited by the area that can be covered during a one-hour search on foot. In addition, the areas inside the buildings were also excluded from search.

The background radiation field in the setting varies due to a very large number of factors such as geographic location, geology, human activity, weather, and others and can be simulated based on the common known types of radiation sources, radiating and shielding

materials, and their geometric configurations. The sources include naturally occurring radioactive materials (soils, geologic formations, etc.), construction materials (granite, clay, concrete, etc.), anthropogenic sources (medical isotopes, industrial equipment, fertilizers, coal, etc.), and others. Realistic simulation was one of the major challenges of the project as simple arrangements of radiating and attenuating materials seldom produce realistic background radiation fields. In our system synthetic background data is generated by selecting a real-world location from several geodatabases. Location is chosen using a combination of user-defined criteria and a Monte Carlo method. In some cases, geodata about the selected location is augmented with details, e.g., placement of radiation sources, vehicles and pedestrians, and searcher paths, using simulation models and reasoning with ontologies. The radiation spectrum is simulated at each point where a detector can be placed in the scene and then the detector response is calculated using a model of the detector. Finally, the detector response is processed by an SNM detection algorithm and that result is compared with the ground truth in the scenario in order to evaluate the algorithm’s quality.

In the technique described in this paper we generate scenarios using a software approach based on a Special Nuclear Materials Detection Ontology (SNM DO). The user interface (UI) for the software will be generated using the SNM DO and the ontology will provide the backbone for interpretation of database terminologies that support the inferencing to generate multiple scenarios. This ontology-based approach could be employed to generate scenarios for other applications in which propagation of electromagnetic radiation plays an important role. For example, synthetic data can be used to model the quality of GPS signal or cell phone coverage on the complex terrain or in urban canyons (Klukas et al. 2003; Greenstein et al., 1997).

The top-level ODSG architecture is shown in Fig. 1. The system consists of two main components: ODSG run-time and ODSG configurator. ODSG configurator uses ontologies and other information like database schemata to generate an ODSG user interface and configuration files for the run-time component. The ODSG run-time is a web application that presents a user with a generated UI from which the user picks output variables and criteria for area selection. Using this input, ODSG run-time extracts the data from the database and executes simulation models performing necessary conversions. The resulting dataset is presented to the user through a web map interface and in a form resembling natural text. The synthetic data itself can be downloaded in an XML format that is transferable to other software to simulate background radiation. Below we discuss each component and functional step of the system in detail.

3.1. Representing expert knowledge in a computational ontology

The complexity of the source configuration requires reliance on expert opinion for the selection of physical locations that best represent background radiation cases, and selection and placement of radiating and shielding materials. Elicitation of expert knowledge is a common technique for addressing very complex and poorly formalized problems and has a long history in many research areas including environmental policy, ecology, and urban analysis to name a few (Fisher, et al., 2012; Price et al., 2011; Masser and Foley, 1987; Karray et al., 2011). In the context of this study capture of expert knowledge should be able to support frequent revisions and account for contradictory expert input.

Subject matter experts (SMEs) for the study were selected from the personnel who had the appropriate background and field experience in radiation detection but did not have GIS know-how. These SMEs represented the potential user community for the system. Elicitation of expert knowledge started with a series of informal interviews with the SMEs. In addition to the interviews we also utilized other knowledge sources suggested by SMEs such

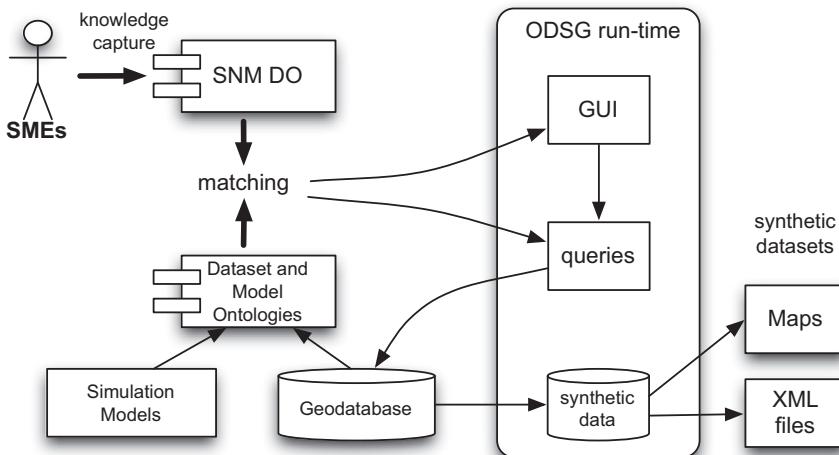


Fig. 1. ODSG system architecture.

as field guides and training manuals. All the sources were analyzed to identify the key concepts that, in opinion of the SME, are important to radiation detection.

Results of the analysis were represented in the Special Nuclear Materials Detection Ontology (SNM DO). A subset of SNM DO is shown in Fig. 2. The concept of a “detection environment” was identified as playing the central role for radiation detection. It is understood as a set of factors can affect readings from the detector. The “detection environment” consists of elements using a set-membership-like relation “element-of” that represent major groups of factors as perceived by SMEs. It is important to note that SNM DO does not include information about special nuclear materials themselves but rather it characterizes the factors that can affect SNM detection.

It is obvious that expert knowledge is based on a much larger body of implicit knowledge. For example, the concept of an isotope, critical to radiation detection and simulation models, implies a more general concept of a chemical element and nucleus. Additional sources such as textbooks, various reference materials, and existing ontologies were used to extend SNM DO to a wider knowledge base.

The use of external ontologies presented certain difficulties. For the areas pertinent to this study ontologies were either non-existent or incomplete for the purposes of our work. As a result, we mostly used external ontologies as terminologies for building our own ontologies. Additional terminologies have been drawn from OntoCAPE, SWEET and CityGML (Marquardt et al., 2010; Raskin and Pan, 2005; Kolbe et al., 2005).

Top-level ontologies have proved to be most useful for our work as they provide a set of unifying well-defined foundational categories and relations. For all ontologies developed in the project we followed best practices for ontology design outlined in such methodologies as ONTOCLEAN (Guarino and Welty, 2002) and Basic Formal Ontology (BFO) (Smith, 2010; Spear, 2006). Most of the categories of ODSG ontologies have a corresponding supercategory in BFO while relations generally follow definitions specified in Relation Ontology from the Open Biomedical Ontologies project (Smith et al., 2005).

The final version of SNM DO ontology that represents expert view on the subject of radiation detection contains around a hundred entities depicting physical (isotope, element, time distance, etc.), geographic (streets, buildings, soils, city furniture, etc.), and generic categories (time of day, season, weather, security assessment, etc.).

3.2. Dataset ontologies

The SNM DO ontology specifies a set of factors that can affect detector readings. Geographic information pertinent to these factors

has to be extracted from a large number of data sources. In ODSG retrieval of geodata is performed in several steps. First, the data source has to be found from the known sources or the lack of the appropriate data has to be recognized. Second, the data has to be retrieved from the source. This requires generation of a query, and then sending the query to the server, and finally retrieving the server response. Third, the database schema of the source has to be translated into ODSG internal database schema. In the end, the data is presented to the user and converted into output format.

Reliance on a computational ontology allowed us, with a relatively small development effort, to automate all stages of data extraction and at the same time made possible integration of a large number of very diverse sources of data and several simulation models. The main challenge for the use of ontology-based techniques at this development stage was the lack of ontologies describing the dataset content.

The following datasets were among the ones used in this study: Homeland Security Infrastructure Protection (HSIP) Gold 2010 (National Geospatial-Intelligence Agency 2010), Geographic Names Information System (GNIS) (U.S. Board on Geographic Names, 2012), Topologically Integrated Geographic Encoding and Referencing system (TIGER) 2010 (U.S. Census Bureau, 2010), building footprints (National Geospatial-Intelligence Agency, 2011), American Hospital Association (AHA) Annual Survey Database TM 2008 (AHA, 2008). Most of these datasets are provided with extensive documentation and metadata. Development of the dataset ontologies was based on the analysis of the internal structure of the datasets, accompanying documentation, and metadata. The extracted categories and relations were expressed in a computational ontology.

HSIP Gold 2010 is a database developed by the Department of Homeland Security and contains nearly 450 feature classes (data layers) and around 25 million features (records). It contains information on many elements of national infrastructure such as railroads, pipelines, electric generation and transmission lines, and others (National Geospatial-Intelligence Agency 2010). The layers are grouped by topics and subdatasets and each layer is provided with FGDC-compliant metadata.

As the first step for developing an ontology, we created a data dictionary containing all column names, unique values, and counts of unique values for all database tables. Using a data dictionary in conjunction with the standard metadata simplifies understanding of the dataset structure by the ontology developers. For example, data domains of the table columns and frequencies of unique values can be observed easily from the data dictionary. Columns belonging to different tables but having similar semantics can be easily identified even if the column names are different.

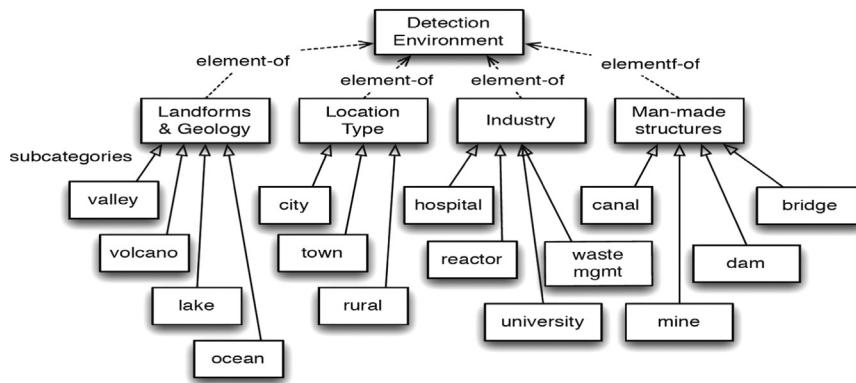


Fig. 2. A portion of the special nuclear materials detection ontology (SNM DO).

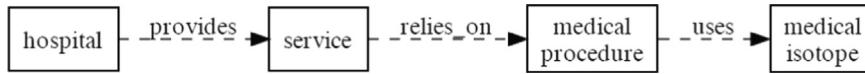


Fig. 3. Ontology diagram for relating hospitals, services, medical procedures and uses of isotopes in procedures.

The resulting ontology for HSIP Gold 2010 has a hierarchical structure with the root category combining all HSIP categories. The next levels of the hierarchy are subsequently represented by infrastructure sectors and feature classes. Each database record is considered to represent an instance in the ontology. Additional hierarchies were created for categorization of features that in the original datasets were represented as repeating values in the text columns. Each entity in the ontology was annotated with a SQL statement to retrieve corresponding instances from the database.

Other datasets used in the study have much simpler structures. GNIS is the official national repository for domestic feature names. It contains over two million features categorized into 65 feature classes. More specific categories are listed in the feature class definitions and can be extracted from the feature names. For example, most of the feature names of the feature class "Dams" contain one of the following terms: breakwater, dike, or jetty. The resulting GNIS ontology contains 307 categories and is organized into three hierarchical levels. The ontology for TIGER 2009 was developed based on the documentation that accompanies the dataset with each feature category identified using MAF/TIGER Feature Class Codes (MTFCC) organized into a hierarchical structure.

In cases when the data was not available we have used a simulation model to substitute for them. The models were used, for example, to create outlines for pavement and sidewalks along street centerlines. The searcher path across the scene was simulated using an A* shortest path algorithm implemented in a discrete event simulation framework (Hart et al., 1968; Luke et al., 2005). Ontologies describing model inputs and outputs were also used for automatic planning of the executable workflow and linking models with the data sources and other models.

In some cases it was possible to deduce new data using inferencing from ontologies. The additional ontologies that were created or adapted as necessary from available sources included an ontology that supports inferencing (see Fig. 3) from the existence of hospitals within the search area to the presence of anthropogenic sources in individuals being detected by the searcher. For example, if a hospital in the search area provides oncology services which rely on ventilation/perfusion (V/Q) procedures (pulmonary perfusion that uses Tc-99m and pulmonary ventilation that uses Xe-133), patients exiting this hospital might be carrying these specific medical isotopes.

3.3. Data integration

In ontology-driven systems, data integration, i.e., transparent use of heterogeneous datasources into a single framework, amounts to integrating the dataset ontologies. This can be achieved by either merging ontologies into a single coherent ontology, mapping between ontologies, or including a single ontology into another using bridging axioms (Buccella et al., 2009). For purposes of this study mapping through the mediator ontology was chosen as the preferred approach (Fig. 4). The use of a mediator ontology reduces the number of matches that have to be maintained with either SNM DO or dataset ontologies updates and the mediator ontology was compiled from the common categories that are found in all ODSG ontologies. Unlike other approaches, mapping between ontologies minimizes the changes in the ontologies being integrated.

Changes in dataset ontologies are often problematic due to a number of reasons. First, altering a dataset ontology to fit another ontology may cause mismatches with the dataset feature definitions. Second, dataset ontologies have to be updated in parallel with new dataset releases and these updates have to be propagated to the rest of the system. In our approach only mappings between the updated and mediator ontologies have to be modified. Third, multiple mappings between the categories of mediator, dataset, and model ontologies are possible. For example, the data pertinent to a certain SNM DO category may be contained in multiple datasets and it should be possible to easily switch between preferred mappings to support system flexibility. Finding matches between categories in different ontologies was done semi-automatically. As the first step, candidate matches were automatically found by scoring similarity between category names and textual descriptions. Then candidate matches which had the highest score were manually selected and the matches were verified using dataset documentation and by comparing the features of the matched categories.

3.4. Ontology-driven GUI generation

One of the advantages of ODIS is the ability to automate generation of the user interfaces. (Section 3.2.2 in Guarino, 1998) Knowing all user-defined categories and their connections to the data sources

allows developers to keep the UIs up to the user expectations and, at the same time, to support the best UI integration with the underlying data. The ODSG UI is implemented as several HTML forms where users can specify the location selection criteria (ranges of variable values) and choose the categories for generating data (Fig. 5). Generation of the UI is done by traversing all SNM DO categories and then using associated information for placing the UI elements on the appropriate HTML pages or parts of pages. For example, subcategories of the same SNM DO category can be placed in a single group on the UI (callout A on Fig. 5). For example, the SNM DO category "Region in US" may have subcategories: "Middle Atlantic", "Midwest", "New England", "South", "southwest", and "West" depending on the particular underlying database of regional names, and the user choices will produce scenarios in multiple, differing regions of the U.S. All these subcategories are grouped together on the same line as the parent category (callout B on Fig. 5).

Also the UI generation algorithm utilizes information that can be extracted from the data underlying SNM DO categories. Each SNM DO category was mapped (see Section 3.3) to a category in the mediator ontology and then to the dataset or model ontology. This mapping is used to build a query to retrieve the corresponding data from the database. The same query can also be used to retrieve other

information for automatic selection of appropriate UI elements. For example, for geographic data types the system automatically chooses UI elements that allow the user to specify the proximity search criteria (callout C on Fig. 5). The sets of the unique values or ranges retrieved from the database are used to create drop-down boxes and groups of checkboxes (callout D on Fig. 5). Category descriptions from the ontologies or metadata can be used for on-line help in the appropriate UI elements.

Additional UI interface information was manually associated with ontology categories and then utilized for UI generation. For example, map styles for presenting the results are built from styling information associated with the categories of the mediator ontology. Associating styling information with common categories and making it independent from a particular dataset improves UI consistency and interchangeability of the data sources.

4. Results and discussion

The system was tested by generating a few hundred scenarios for various user-defined conditions. The maps of four typical scenarios are shown on Fig. 6. Fig. 6A shows one of the scenarios in which the user

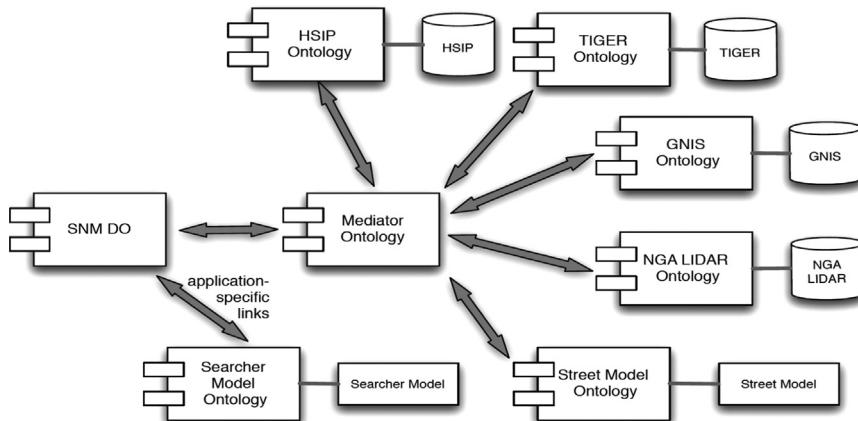


Fig. 4. Mapping between ontologies with a mediator ontology.

ODSG _User Input Variables	
Start:	Save Changes
Instructions:	Reset to Defaults
Main	
Edit Parameters:	
Setting	
Climate	
NORM	
Anthropogenic	
Situation	
Searcher	
Detector	
Post Editing:	
Display Parameters	
Generate Scenarios	

Search for Setting Parameters

General Setting:

User	Parameter:	Value Options:
Region in US	Middle Atlantic, Midwest, New England, South, Southwest, West	
Type of Setting	Metropolis, City, Town, Rural, Residential	
Zoning	Downtown, Commercial, PublicPark	
Population Density/km ²	0-1000, 1001-5000, 5001-10000, above 10000	
Elevation (m)	0-300, 301-600, 601-900, 901-1200	

Local radiation detection infrastructure

User	Iterates:	No closer than:
1	0	3
1	0	3
1	0	3

Industry in the Area

User	Iterates:	No closer than:	No further than:
Hospital	1	0	3
University	1	0	3
Manufacturing	1	0	3
Coal Plant	1	0	3
Nuclear Reactor	1	0	3
Nuclear Industry	1	0	3

Fig. 5. ODSG input form showing a portion of screen for capturing Setting parameters.

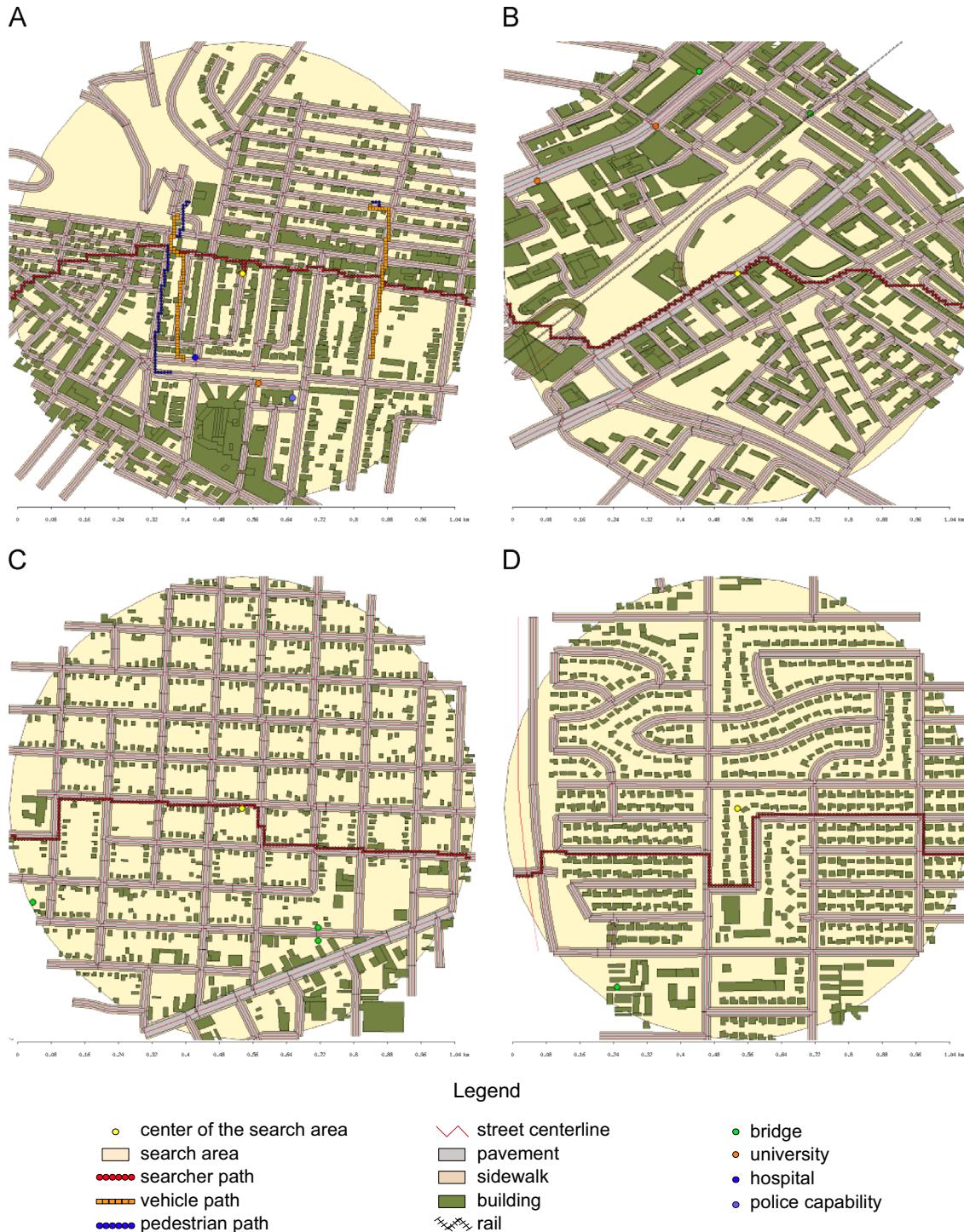


Fig. 6. Examples of the scenarios generated using ODSG.

has made the following choices: (1) two pedestrians and (2) two vehicles are moving inside the search area, and (3) a hospital and (4) a university are located within a 3 km distance from the center of the search area. Fig. 6B shows a scenario that was generated for the following inputs: (1) a single searcher, (2) mixed industrial and commercial area, and a (3) rail line crossing the search area. Fig. 6C and D contains scenarios for similar searches in predominantly residential areas. Each scenario shows a search area and its center,

the path of the searcher, and a number of other objects like street centerlines, extent of the pavement and sidewalks, building footprints, and objects of interest.

One of the main advantages of ODIS as a framework is the ability to manage very semantically diverse datasets like HSIP Gold and utilize information that is contained therein more effectively. For example, many tables in HSIP have columns with codes from external classifications, such as the North American Industry Classification

System (NAISC) ([U.S. Census Bureau, 2011](#)). The ability to automatically generate a query that would retrieve features with the same NAISC codes from multiple tables and at the same time take into account hierarchy of NAISC categories is a definite asset for the system users and maintainers. ODSG relies on mapping between ontologies and database schema and then using ontologies to configure the application and generate the application UI. Our approach is close to the one described in ([Guarino, 1998](#)). Unlike other approaches we did not use ontologies to describe software artifacts such as files, database tables, or map between programming objects ([Fonseca et al., 1999; Knublauch, 2004; Fonseca et al., 2003](#)). For most blocks of the system we have used readily available open-source GIS software components.

Providing the system with the right ontologies was one of the major challenges of the project. Most of the readily available ontologies have demonstrated limited suitability for this type of application and required a great deal of additional effort to incorporate these into the system. We found the ODIS approach to be useful for recognizing mismatches between the concept structures of specific datasets and users. For example, SMEs have identified the category “University” as a kind of “Industry” ([Fig. 7](#)). This is a justifiable decision in the context of SNM detection as universities often have industrial-scale research equipment. At the same time, in HSIP Gold ([National Geospatial-Intelligence Agency 2010](#)) universities were grouped under the category “Education” and in GNIS ([U.S. Board on Geographic Names, 2012](#)) under the category “Schools”. In both datasets “Industry” was represented by categories separate from both “Schools” and “Universities”. The structures of the conceptualization in regards to “University” are also different between HSIP Gold and GNIS: GNIS uses the term “Schools” for a number of educational institutions including “Universities” while in HSIP Gold the term “Schools” refers to secondary schools.

Another example of mismatched definitions is the concept of a street as understood by SMEs and as it is represented in the datasets. In the datasets, streets are represented either as centerlines (e.g., TIGER) or as driving lanes (e.g., OpenStreetMap). However, in the context of SNM detection, street is most important as a place where certain object (e.g., cars, pedestrians) and materials (e.g., asphalt, curbstone, concrete) are present. In that sense SME’s conceptualization of the street is closer to the “Transportation Complex” in CityGML ([Groger et al., 2008](#), see Fig. 36). In some cases users’ specifications were not (or possibly cannot be) exact. For example, specifying U.S. regions (e.g., South, North West, New England) is straightforward at the intuitive level. However, it implies one of a number of regionalizations that users do not explicitly state. To address this problem we support multiple regionalizations that can be selected by a user. An

alternative solution would be to use vague regions and utilize associated probabilities for the Monte Carlo selection of a suitable location.

5. Conclusion

Like other systems based on the ODIS paradigm, ODSG takes advantage of a number of features of ODIS both in terms of user and developer experience. The ODSG application presents a UI that utilizes categories that are familiar to end-users and faithfully represents conceptualization of the users’ domain. At the same time users are not exposed to the underlying mechanics of data retrieval and other computational details. A sophisticated system for generating a large number of scenarios, the ODSG was created within a time frame that would not be feasible with the traditional approaches relying on hard-coded UIs and datasets. The ODSG demonstrates one of the main advantages of ODIS, its flexibility to accommodate new datasets and modifications in user requirements. For the majority of cases adding a new dataset requires a minimal programming effort with most of the work spent on ontology development and mapping between ontology categories. Simple switching between the datasets with similar semantics (e.g., TIGER or OpenStreetmap) allows users to select datasets best suited to their specific tasks.

From a simple set of user inputs selected from the ODSG UI, the system was able to create over 100 different scenarios for multiple locations in the continental U.S. The SNM DO and other supporting ontologies provide a critical backbone for interpretation of database terminologies associated with the large number of databases that support the ODSG system, including HSIP, GNIS, and TIGER. The versatility of the results and the simplicity of modification of the overall design and construction of the ODSG testify to the design principles behind ODIS and speak to their effectiveness for this type of application. One of major advantages of ODIS is the ability to generate UI based on SNM DO, the SNM-specific ontology created from subject matter expertise. The UI can be easily altered by addition, deletion, or expansion of SNM DO. Mapping of SNM DO categories to the categories in the mediator ontology and then to the associated data source ontologies can be used to set checks and provide default values for user inputs.

One of the main challenges of the project was creation of the ontologies. Direct use of subject-specific large-scale ontologies such as OntoCAPE or SWEET has proved to be problematic because of the difficulties of connecting of these ontologies to SNM DO.

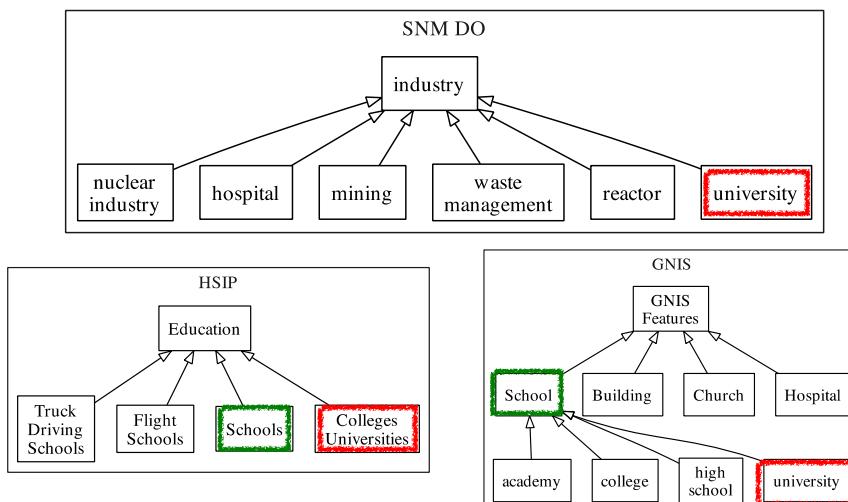


Fig. 7. Mismatches in ODSG Ontologies: highlighted categories show different structure of subcategory relations in different ontologies.

However, upper-level ontologies were helpful for developing links among SNM DO and dataset ontologies. Nonetheless, the ODSG has proven that a robust system can be built on ODIS design principles and that it can be used to effectively generate a very much larger number of scenarios given just a few user input variables. Further enhancements will only add to this capability and provide for a more-refined approach to generating scenarios, extending significantly the methods previously discussed for automatic generation of scenarios to support modeling and simulation and testing and validation of algorithms for SNM detection. Several software modules and ontologies that were developed or used as a part of this project were released under open source licenses and are available on GitHub repositories <https://github.com/geo-ont/GooGK> and <http://sorokine.github.com/SOFT>.

Acknowledgments

Research sponsored by DOE/NNSA NA-22 Simulation, Algorithms and Modeling Program under Contract: OR10-Ontology Demo-PD06. This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

References

- American Hospital Association: AHA annual survey database TM, 2008.
- Akerman, A., Tyree, J., 2006. Using ontology to support development of software architectures. *IBM Syst. J.* 45 (4), 813–825.
- Bossche, M., Ross, P., MacLarty, I., Van Nuffelen, B., Pelov, N., 2007. Ontology driven software engineering for real life applications. In: Proceedings of the 3rd International Workshop on Semantic Web Enabled Software Engineering.
- Buccella, A., Cechich, A., Fillottrani, P., 2009. Ontology-driven geographic information integration: a survey of current approaches. *Comput. Geosci.* 35 (4), 710–723.
- Fisher, R., O'Leary, R.A., Low-Choy, S., Mengersen, K., Caley, M.J., 2012. A software tool for elicitation of expert knowledge about species richness or similar counts. *Environ. Model. Softw.* 30, 1–14.
- Fonseca, F.T., Egenhofer, M.J., 1999. Ontology-driven Geographic Information Systems, in: Proceedings of the 7th ACM International Symposium on Advances in Geographic Information Systems, GIS '99. ACM, New York, NY, USA, pp. 14–19. <http://dx.doi.org/10.1145/320134.320137>.
- Fonseca, F., Davis, C., Câmara, G., 2003. Bridging ontologies and conceptual schemas in geographic information integration. *GeoInformatica* 7 (4), 355–378. <http://dx.doi.org/10.1023/A:1025573406389>.
- Fonseca, F., 2007. The double role of ontologies in information science research. *J. Am. Soc. Inf. Sci. Technol.* 58 (6), 786–793.
- U.S. Board on Geographic Names: Geographic names information system, 2012. (<http://geonames.usgs.gov/>). (accessed 29.05.12.).
- "Geographic Information System." Wikipedia, the Free Encyclopedia, August 16, 2014. (http://en.wikipedia.org/w/index.php?title=Geographic_information_system&oldid=621498407).
- Greenstein, L.J., Erceg, V., Yeh, Y.S., Clark, M.V., 1997. A new path-gain/delay-spread propagation model for digital cellular channels. *IEEE Trans. Veh. Technol.* 46 (2), 477–485.
- Groger, G., Kolbe, T.H., Czerwinski, A., Nagel, C., 2008. OpenGIS city geography markup language (CityGML) encoding standard. Tech. Rep. 08-007r1, Open-geospatial Consortium.
- E. Grois, W.H. Hsu, M. Voloshin, D.C. Wilkins, 1998. Bayesian Network Models for Generation of Crisis Management Training Scenarios. In: AAAI '98/IAAI '98 Proceedings of the fifteenth national/tenth conference on Artificial intelligence/ Innovative applications of artificial intelligence. American Association for Artificial Intelligence Menlo Park, CA, USA, pp. 1113–1120.
- Guarino, N., 1998. Formal ontology in information systems. In: Proceedings of the 1st International Conference on Formal Ontology in Information Systems (FOIS'98). IOS Press, Amsterdam, pp. 3–15.
- Guarino, N., Welty, C., 2002. Evaluating ontological decisions with OntoClean. *Commun. ACM* 45 (2), 61–65.
- Hart, P.E., Nilsson, N.J., Raphael, B., 1968. A formal basis for the heuristic determination of minimum cost paths. *IEEE Trans. Syst. Sci. Cybern. SSC4* 4 (2), 100–107.
- National Geospatial-Intelligence Agency: Homeland security infrastructure protection (HSIP) gold 2010, 2010.Tech. rep., The Department of Defense, National Geospatial-Intelligence Agency, Office of Americas, North American and Homeland Division.
- Hullett, K. and Mateas, M., 2009. Scenario generation for emergency rescue training games. In: Proceedings of the 4th international Conference on Foundations of Digital Games (Orlando, Florida, April 26–30, 2009). FDG '09. ACM, New York, NY, pp. 99–106.
- Incident and Trafficking Database (ITDB). (<http://www-ns.iaea.org/security/itdb.asp>) (accessed 18.08.14.).
- Karray, M.H., Chebel-Morello, B., ZerHouni, N., 2011. A formal ontology for industrial maintenance. In: Proceedings of the Terminology & Ontology: Theories and Applications, TOTH Conference, Annecy, France.
- Klukas, R., Lachapelle, G., Ma, C., Gee, G.I., 2003. GPS signal fading model for urban centres. *Microwaves, antennas and propagation. IEE Proceedings* 150 (4), 245–252.
- Kolbe, T., Gorger, G., Plumer, L., 2005. CityGML - interoperable access to 3D city models. In: Proceedings of the First International Symposium on Geo-Information for Disaster Management GI4DM. Springer Verlag, pp. 21–23.
- Koziarz, W.A., L.S. Krause, L.A. Lehman, 2003. Automated scenario generation, Enabling Technologies for Simulation Science VII. In: Sisti, Alex F.; Trevisani, Dawn A. (Eds.), Proceedings of the SPIE. Vol. 5091, pp. 374–382.
- Knublauch, H., 2004. Ontology-driven software development in the context of the semantic web: an example scenario with Protege/OWL.
- National Geospatial-Intelligence Agency, 2011. LiDAR dataset for major U.S. cities.
- Luke, S., Cio-Revilla, C., Panait, L., Sullivan, K., Balan, G., 2005. MASON: a multi-agent simulation environment. *Simulation*, 81; , pp. 517–527.
- Marquardt, W., Morbach, J., Wiesner, A., Yang, A., 2010. OntoCAPE: a re-usable ontology for chemical process engineering. Springer Verlag.
- G.A. Martin and C.F. Hughes, 2010. A scenario generation framework for automating instructional support in scenario-based training. In: SpringSim'10 Proceedings of the 2010 Spring Simulation Multiconference Article no. 35. Society for Computer Simulation International, San Diego, CA.
- Masser, I., Foley, P., 1987. Delphi revisited: expert opinion in urban analysis. *Urban Stud.* 24 (3), 217–225.
- Medalia, J., 2010. Detection of Nuclear Weapons and Materials: Science, Technologies, Observations, Congressional Research Service Report for Congress 7-7500, pp. 10–24.
- Murauskaitė, E., 2014. The scientist and the nuclear smuggler: unexpected connection. *Bull. At. Sci.*, 2014 (<http://thebulletin.org/scientist-and-nuclear-smuggler-unexpected-connection7372>).
- U.S. Census Bureau: North American industry classification system (NAICS), 2011. (<http://www.census.gov/eos/www/naics/>).
- Price, J., Silbernagel, J., Miller, N., Swaty, R., White, M., Nixon, K., 2012. Eliciting expert knowledge to inform landscape modeling of conservation scenarios. *Ecological Modelling, Model. Hum. Decis.* 229, 76–87. <http://dx.doi.org/10.1016/j.ecolmodel.2011.09.010>.
- Raskin, R.G., Pan, M.J., 2005. Knowledge representation in the semantic web for earth and environmental terminology (SWEET). *Comput. Geosci.* 31 (9), 1119–1125.
- D. Reynolds, 2001. A Framework for Scenario Generation in ALGO Research Quarterly.
- Runkle, R.C., Smith, L.E., Peurung, A.J., 2009. The photon haystack and emerging radiation detection technology. *J. Appl. Phys.* 106 (4), 106–127.
- Smith, B., 2010. Basic Formal Ontology, (<http://www.ifomis.org/bfo>).
- Smith, B., Ceusters, W., Klagges, B., Khler, J., Kumar, A., Lomax, J., Mungall, C., Neuhaus, F., Rector, A., Rosse, C., 2005. Relations in biomedical ontologies. *Genome Biol.* 6 (5), R46.
- Spear, A.D. 2006: Ontology for the twenty first century: an introduction with recommendations. Tech. rep., IFOMIS, Saarbrücken, Germany, <http://ifomis.uni-saarland.de/bfo/documents/manual.pdf>.
- U.S. Census Bureau, 2010. Geography Division: TIGER data.
- Wald, C., 2014. Medical isotopes confound nuclear test monitoring. *Science* 345, 126. <http://dx.doi.org/10.1126/science.345.6193.126>.