An Ontology Based Personal Exposure History

Stacy A. Doore
University of Maine
344 Boardman Hall
Orono, ME 04469
207 581-2188
stacy.doore@maine.edu

Kate Beard
University of Maine
344 Boardman Hall
Orono, ME 04469
207 581-2147
beard@spatial.maine.edu

Carol Bult
The Jackson Laboratory
600 Main St.
Bar Harbor, ME
207 288-6324
carol.bult@jax.org

ABSTRACT
Health monitoring and disease surveillance systems can benefit from the integration of multiple data sources and semantic web technologies The development and management of exposure histories is one particular area that requires integration of data from multiple sources. Exposure histories capture the spatial and temporal dimensions of possible disease exposure events and convey the dynamic factors within individuals' environments. Data sources for an exposure history might include numerous governmental, institutional and community records which document people's relationships with various locations. Existing monitoring networks and new wireless sensor networks also provide data on toxic agents in the environment. This paper describes the development of an ontology for a *personal exposure history (PEH)* that specifies explicit relationships between persons and locations and locations and putative environmental toxic agents. These relationships provide a foundation for making inferences about person to putative toxic exposure relationships. The PEH ontology defines a framework of concepts and relationships on which to integrate data from heterogeneous sources. The ontology does not capture a complete exposure profile as yet, but instead represents key spatial and temporal concepts and demonstrates how these can be queried using current semantic web technologies.

Categories and Subject Descriptors
J.3.3 [Computer Applications]: Life and Medical Sciences

General Terms: Design

Keywords: Exposure history, semantic web, ontology

1. INTRODUCTION
Many diseases of interest to epidemiologists, geneticists, and public health researchers have long latency periods (e.g. years). Analysis of possible relationships of long-latency diseases, such as cancer, to environmental risk factors becomes complicated by the reality of changing spatial and temporal factors in the population and the environment. In a highly mobile society, individuals are likely to have moved several times from the point in time of possible exposure to disease causing agents to the time of disease diagnosis. They may also have associations with several locations in any one time period. Similarly, disease causing agents may be mobile (e.g. a toxic cloud or plume) or transient with time (e.g. a pesticide applied over two years more than ten years ago). The dispersion of disease cases over space and time makes it difficult for researchers to determine possible environmental factors that may be related to the development of a disease with any certainty. In order to document the location and duration of possible exposures to harmful agents over an individual's lifetime, an analysis strategy must include a host of factors operating on disparate scales. A comprehensive exposure risk history to support such analysis should contain information regarding a person's behavioral, socioeconomic, environmental, and genetic risk factors. This paper describes the use of semantic web technologies for modeling personal mobility information and environmental factors as component parts of such a comprehensive exposure history.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
IHI'10, November 11–12, 2010, Arlington, Virginia, USA.
Copyright 2010 ACM 978-1-4503-0030-8/10/11...$10.00.

2. SEMANTIC WEB TECHNOLOGIES
Data used in many environmental health monitoring systems are often limited for a variety of reasons. Some monitoring systems suffer from low spatial or temporal resolution or from inaccurate or missing information. Additional problems may arise from changes in data measurement, collection practices, changes to the information system infrastructure, or how environmental data is represented over time and space. Current environmental health systems often have difficulty with the integration of longitudinal data about people and their relationships with everyday locations. As a result, there is a need for systems that can efficiently link new, diverse sources of environmental data from an emerging 'Sensor Web' [12] and be flexible enough to integrate data about individual human mobility over long periods of time [29].

The term *Semantic Web*, [6] has come to represent an extension of the World Wide Web where the meaning or semantics of heterogeneous data, information and knowledge can be easily understood, processed, linked, and exchanged by machines. The power of the Semantic Web is its ability to transform the underlying language and structure of data into something that machines can understand and process directly or indirectly through the use of specifically formatted metadata in web based content [7]. Semantic Web technologies are rapidly changing data integration practices and system structures in many fields such as business and bioinformatics. These technologies are ideally suited to support information integration from diverse sources due to the data annotation features and their ability to link new information to existing knowledgebases through the use of explicit ontological structures. The Semantic Web and its related technologies can be understood through its four primary components [37].

Uniform Resource Identifiers (URIs): The Semantic Web relies on URIs as a standardized format for naming resources consistently and uniquely. The link between a URI and its reference resource allows for easy identification, retrieval and linkage to any resource on the web.

Resource Description Framework (RDF): The RDF model has been adopted by the W3C as a formal way of describing and representing resources on the web [48]. It offers a general method for conceptual description of the information content of web resources identified by URIs. RDF uses the triple construct *(subject, predicate, object)* to express assertions. The subject of an RDF triple can be a URI or an anonymous blank node, a predicate must be a URI, and an object can be a URI, blank node, or literal (i.e. values such as strings or numbers) [31]. A set of triples or *RDF graph* is represented as a directed labeled graph with typed edges and nodes where the directed edge is the labeled predicate connecting subject to object nodes [41].

Ontologies: An ontology is commonly defined as "an explicit specification of a conceptualization" [18] or a "representation of real world entities using a sophisticated structure with components such as definitions, parts, functions, attributes, and rules of relationship" [14]. Ontologies provide standardized vocabularies and explicit conceptual frameworks that allow for common understandings of concepts and data across multiple domains. The Web Ontology Language (OWL) with its basis in Description Logics (DLs) has become a cornerstone of the Semantic Web for its use in the design of ontologies. OWL defines concepts, properties (relationships), and logical combinations of concepts (e.g., intersection, union, and complement) and provides an inference mechanism by defining restrictions and assertions about property types (e.g. transitive, symmetric, functional, inverseOf, and inverse functional). The foundation in DL provides precise and unambiguous meaning to descriptions of a domain and supports reasoning algorithms that can handle complex queries about the domain.

Simple Protocol and RDF Query Language (SPARQL):
SPARQL is an RDF query language for use with the Semantic Web [49]. SPARQL is syntactically similar to SQL however it allows the user to query RDF graphs via pattern matching. In SPARQL, queries behave like RDF in that a pattern consists of triples with variables substituted for subjects, predicates, or objects. An advantage of SPARQL is that it does not require explicit joins to specify the relationship between differently structured data. Data from different sources with unpredictable or unreliable structure can typically be easily mapped to RDF and then queried using SPARQL. Mappings can be performed on the fly, allowing exploration of heterogeneous data at a higher level than that of the native structure of the original data.

Researchers seeking access to data for environmental health research face barriers in finding the most current environmental monitoring data collected and disseminated through sensor networks. This aspect of conducting environmental health research can be difficult because of the rapidly changing nature of the deployment of these types of networks and the potential volume of data they generate. Many of these systems are now using web based interfaces, but health researchers could potentially spend unnecessary time updating and maintaining this type of data within their own databases. Automated pre- and post-processing of environmental monitoring data could be achieved through harnessing the semantic power of web based data, and as these types of information sources become more readily available, the development of an information system that can semantically integrate environmental exposure data becomes a viable concept.

This paper presents a conceptual framework for *a personal exposure history*. A personal exposure history represents spatio-temporal relationships between individuals, locations, and toxic agent levels that can be interactively queried [13]. The purpose of the framework is intended to help various users in the biomedical domain (epidemiologists, geneticists, public health researchers) detect noteworthy or significant interactions between person-location relationships and potentially harmful environmental events in order to more fully evaluate environmental health risks over the course of a person's lifetime. The approach builds upon earlier work of representing relationships between spatio-temporal events [5][13] specifically in a public health or epidemiological application. A key contribution of the PEH ontology is that it specifies different types of person-location relationships as events (e.g. person resides at residence) and differentiates these relationships, creating a richer semantic context for associated activities that can be inferred from them.

3. A PERSONAL EXPOSURE HISTORY

The term *exposure*, in the epidemiological domain, is defined as "the event when a person comes into contact with a toxic material**"** [30]. A person's behavior/lifestyle choices, geographic location, socio-economic conditions and genetic predispositions will impact an exposure outcome, as will the real world conditions of the contact event itself. In this definition of exposure, a toxic *agent* is an entity that contacts a *receptor*, which is the entity that receives an exposure or a dose (e.g., a human, human population, or a human organ). This paper focuses on *environmental exposures*, defined here as human exposure to pollutants originating from manmade sources or harmful levels of natural elements within a person's environment. Human beings, individually and collectively, are exposed daily to numerous harmful elements in both indoor and outdoor settings [42].

3.1 Representing Human Mobility

Exploring human movement in space and time can help health researchers to uncover patterns in disease and possible environmental agents. The widespread availability of location aware ubiquitous computing and communication devices has led to a proliferation of data and analysis concerning people as dynamic objects. This type of mobility data is often structured as a set of triples (o,p,t) recording the presence of an object o in position p at time t. The problem becomes how to extract information about trends and patterns on interactions within and between people and locations, locations and events at those locations, and the relationships that might be inferred about people and things that happen at locations at which they have been present. Hägerstrand's seminal work [21] discussed issues involving the study of individual movement and its ability to enhance knowledge of social and group large scale patterns of behavior. Odland [36] describes this type of timeline as a *lifeline*, the space-time points of an object's movement from sample point A to sample point B. From this distinction, a *geospatial lifeline bead* is derived [32][22] as a set of all possible spatial locations that an object could have occupied between two sample points given the maximum velocity at which the object moves between two points. A *geospatial lifeline* is defined as a time-stamped record of locations representing places a person has occupied over a period of time. Fine level spatio-temporal detail may provide new information and likewise, a coarser granularity or aggregation of spatio-temporal data will provide more general representations of object movements [24]. A coarsened granularity results in less temporal and spatial detail and less detail on speed of movement. However, these abstractions are useful for generalized views of movements based on very large databases of detailed geospatial lifeline data as might be needed to represent movement in a personal location history framework. This model is often used in representing residential tenure at a coarse granularity (months or years). A *location history* [22] is a record of an entity's location in space over an interval of time. Sinha and Mark [44] use residential history locations to represent discrete geospatial lifeline data to model clusters of people with similar exposure risk histories.

3.2 Representing Human Exposure

Exposure science and spatial epidemiology have a number of existing tools to assist in better understanding the spatial distribution of disease in human populations. The CDC's Epi Info [10] and IBM's Spatiotemporal Epidemiological Modeler (STEM) [15] provide epidemiologists, public health practitioners and researchers with tools for database construction, data entry, and analysis with epidemiologic statistics, maps, and graphs. These types of epidemiologic applications typically have spatial data, SQL query and HTML output capabilities. STEM provides spatial and temporal models of emerging infectious diseases and a number of mathematical models to represent multiple populations (species) and the transmission process between animal and human diseases in order to create a better understanding of interconnectedness between global disease transportation links.

Some systems have developed out of the field of exposure science. SHEDS (Multimedia v.3)[50] is a probabilistic aggregate residential exposure model with a capacity to address a variety of chemical classes and exposure scenarios such as particulate matter, air toxics and treated wood products. This modeling tool is used to simulate an individual's contact with environmental agents and estimate an exposure profile for multiple pathways. This system can also be used to predict ranges of exposure in a population. The SHEDS model has the capacity to link with toxicology source models, and pharmacokinetic (PBPK) models to quantify and reduce uncertainty in risk assessments. The data model consists of time series data for the simulated environmental toxin concentration within a residence, as well as fine scale human movement data. A limitation of this system's data structure is that each simulation will have tens of thousands of events for each simulated individual. While the results are automatically aggregated over time to produce simulated daily exposure estimations, the size of the datasets are problematic. Event-level exposure profiles are not saved as permanent output and only summary statistics on daily and longer periods are saved and retrieved for continuous usage.

In related work from the domain of geographic information science, Meliker et al. [33][34] have characterized the concept of an *exposure lifeline* as a temporally continuous data structure which allows aspatial attributes such as temporally varying exposure estimates to be visually represented. Jacquez, Greiling and Kaufmann [28] note that disease surveillance systems require spatio-temporal data structures that provide the ability to query on people, their disease status as well as the duration and indexing of possible exposure events. Their Space Time Information System (STIS) uses an object-based approach to model movement and attribute change. The PEH ontology based approach differs from the STIS data model in several ways: 1) it conceptualizes a personal exposure history as explicit event-event relationships between dynamic objects (personal location events) it uses event sequences at named locations, and 2) it employs of semantic web tools to explicitly represent spatial, temporal, and thematic semantics.

3.3 Semantic Technologies and Health Data

Applications of a Semantic Web approach are increasingly common in many domains including national security, business, regulatory compliance, and biomedical informatics [1][35][37]. This type of approach utilizes diverse data sources, ontologies and semantic metadata standards to help facilitate the aggregation and integration of complex information [39]. A recent application of this type of approach can be seen in the development of the cancer Biomedical Informatics Grid (caBIG™). Within this larger system, the Lymphoma Enterprise Architecture Data-system™ (LEAD™) [26] integrates pathology, pharmacy, laboratory, cancer registry, clinical trials, and clinical data from a wide variety of institutional databases. It utilizes the Cancer Common Ontological Representation Environment Software Development Kit (caCORE SDK) provided by National Cancer Institute's Center for Bioinformatics to establish a platform for data management, controlled vocabularies, and registered metadata to achieve semantic integration across multiple cancer databases. While this system provides a demonstration of semantic technologies in support of a wide range of clinical and research tasks, and integrates data from disparate systems into a single architecture, it is not designed to handle contextual information such as longitudinal data regarding human mobility patterns or environmental health monitoring data.

A recent extension of semantic analytics, Geospatial Semantic Analytics (GSA) [4][38][40] provides a solution and the necessary infrastructure for the representation of contextual information through complex spatial, temporal, and thematic information analysis by using ontologies based on multimodal geographic information. This ontology based approach exploits the relationship-centric nature of semantic web data to model and query spatial, temporal and thematic data. GSA uses indirect relationships to achieve a mapping between thematic objects and spatial objects and uses these relationships to define a notion of *context* that can be queried using spatial and temporal properties. Our work builds on this approach to support the spatial, temporal and semantic integration of data specifically for the construction of personal exposure histories. The ontology creates the schema for resource descriptions and defines a vocabulary of labels for nodes (classes) and edges (properties) that are used to define key concepts and relationships of an exposure history. Relationships between persons and locations are defined as well as the time spent in those relationships. Independently, we define relationships between toxic events and locations and the duration of these relationships. Potential exposures are then identified by spatial and temporal overlaps in these relationships. The approach provides substantial flexibility in that the vocabularies can be easily extended to meet the description needs or conventions of one user group (i.e. toxicologists) while not compromising the autonomy in descriptions used by another community or group (i.e. healthcare workers, epidemiologists).

4. ONTOLOGY BASED PERSONAL EXPOSURE HISTORY FRAMEWORK

The conceptual detail for a personal exposure history ontology can be complex, and the intent of this paper is not to cover all these complexities but to address key spatial and temporal concepts. To adequately represent a personal exposure history, concepts are needed for entities that are static and entities that are dynamic. Rather than reinvent general concepts, upper level ontology concepts for space and time are used to ground the PEH. This section introduces definitions of general concepts from an upper level ontology and then specialize theses concepts within the PEH ontology.

4.1 Upper Level Ontology

General spatio-temporal concepts for the PEH utilize the concepts from the Basic Formal Ontology (BFO) [16]. The BFO provides both a static (snapshot) view and a process (lifespan) view of time and space through its SNAP and SPAN divisions. The BFO was developed for broad application in the biomedical domain and has been used in several related domain ontologies such as clinic-genomic trials on cancer [9][17].

The BFO is divided into two kinds of entities: substantive entities (continuants) and processual entities (occurrents). These entities provide the structure to represent spatial and temporal concepts as both abstract and concrete entities (Figure 1). A *Continuant* is classified as a SNAP entity in that it represents a substantive entity that exists and persists through time while maintaining its identity (examples: a person, a city, a building, a tumor). The *Continuant* entity is the high level concept for several persistent entities within the PEH ontology. An *Occurrent* is a processual entity that has temporal parts and represents a 'timespan' view of the world for entities that exist and are located within an interval of time, unfold through time, and then disappear [46]. SPAN entities have no realization in a snapshot of time, requiring some interval of time to establish a presence [45]. *Occurrent* entities are associated with a temporal setting [47] through a has_A relation. This *Temporal Setting* uses concepts and relations from the OWL-TIME ontology [23].

4.2 Middle Level Ontology

The middle level of the ontology specializes concepts found in the BFO. The primary *Continuant* subtype in this level is the *Dynamic Entity* and *Occurrent* subtypes include *Spatial Occurrents* and *Non-Spatial Occurrents* (Figure 1). A *Dynamic Entity* is an entity that changes its spatial behavior [38]. Like all continuants, a dynamic entity maintains a unique identity but this identity is not strongly associated with space. A *Spatial Occurrent* represents an event with a definite spatial setting for every time unit during which it exists. A spatial occurrent's temporal existence is defined by the interval of time for which it persists [45]. The *occurredAt* relation connects the *Spatial Occurrent* entity to a spatial setting. The *occuredAt* relation has an inverse relation, the *experiences* relation which expresses that a *Spatial Setting* can experience an event. These inverse relationships add flexibility for queries as described later and OWL supports inference on inverse relations. A *Non-Spatial Occurrent* has only a temporal setting (e.g. a patient treatment regime). In the PEH ontology, the non-spatial occurrent concept is not fully developed but provides a placeholder for future extensions.

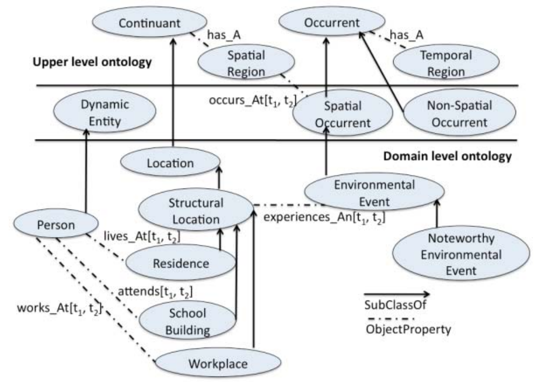

Figure 1. Subset of PEH ontology concepts and relations

4.3 PEH concepts and relations

Emerging work by Cohen Hubal [11] provides three high level concepts for an exposure model: a stressor, a biological receptor, and an outcome (Figure 2). The PEH focuses specifically on the second tier of this model (highlighted in the figure below), the level of the individual. An individual is the dynamic entity of interest and the PEH ontology focuses on capturing the dynamic behavior of individuals in space and time. The individual is also a core entity as individuals can be aggregated to various populations (e.g. genetic, geographic, ethnic, socio-economic) as well as being a carrier of tissues and cells. The PEH models relationships of individuals to locations and stressors to locations with the intent to find where these two types of relationships coincide in space and time. Behavioral, socioeconomic and genetic concepts are not specified here but would be components of future work in a comprehensive exposure model.

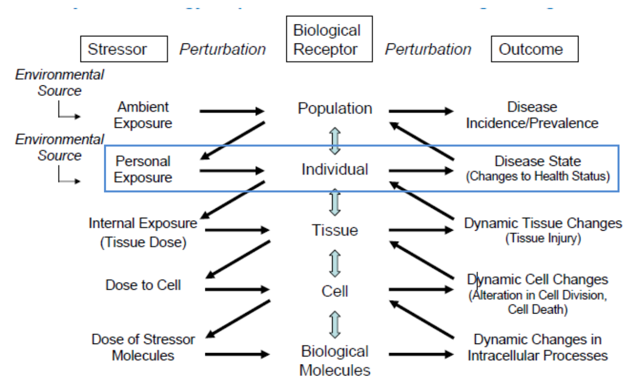

Figure 2. Cohen Hubal's exposure-response processes (2009)

A *Person* is the central class in the domain level of the PEH ontology representing an individual as the biological receptor. The class description for *Person* includes several basic properties necessary for epidemiological research, as well as other applications. General axioms for *Person* include:

Declaration(Class(*:Person*))

SubClassOf(*:Person, :BiologicalReceptor*)

FunctionalDataProperty(*:hasID*)

FunctionalDataProperty(*:lastName*)

FunctionalDataProperty(*:birthDate*)

FunctionalDataProperty(*:gender*)

For spatial epidemiological purposes, however, the modeling challenge lies in representing a person's dynamic behaviors. A typical person has associations with many locations, some for long periods of time and others only very briefly. Current location aware technologies can provide nearly continuous traces of people's short-term movements and obtain fine-grained location histories for individuals. Representing longitudinal location data that captures movement flow and spatio-temporal relationships between individuals and the locations where they spend longer intervals of time are less common. Rather than near real-time location tracking, the PEH models a person's relationships to locations where they tend to spend larger blocks of time: home, school, work, or military base.

The second key class in the PEH is *Location*. A *Locati*on is defined as an identifiable geographic place [27] that may or may not have well defined physical boundaries such as those delineating a mountain, lake, valley, marsh or fiat boundaries as delineating a parcel, census block, or town. A location must be identified by one or more spatial references, which defines a location's spatial setting. A spatial reference can be a coordinate based geospatial object such as a point, line or polygon, a linear reference (position along a pathway) or a geoidentifier (e.g. placename, zipcode). Some axioms for *Location* and *Spatial Reference* include:

Declaration(Class(*:Location*))

Declaration(Class(:*SpatialReference*)

SubClassOf(*:GeoObject, :SpatialReference*)

SubClassOf(:*GeoIdentifier, :SpatialReference*)

ObjectProperty(:*locates*)

ObjectPropertyAssertion(*:locates, :SpatialReference, :Location*)

InverseObjectProperty(*:locates :locatedBy*)

A *Person* has a relationship to a *Location* through an OWL object property *occupies* and its inverse *hasOccupant*. The two axioms,

ObjectPropertyAssertion(*:occupies :Person, :Location*)

InverseObjectProperty(*:occupies :hasOccupant*)

allow us to deduce that when arbitrary Person A and Location B, are linked by the *occupies* property, that B and A are interlinked by the *hasOccupant* property. A *Structural Location* is a subclass of *Location,* which defines a human constructed and physically bounded space. Structural locations define confined indoor spaces that can have different types of daily exposure risks for a person than unbounded geographic locations. A *Structural Location* is subclassed to types of structures that include *Residence*, *SchoolBuilding* and *Workplace.* The PEH ontology represents

explicit semantics for relationships between *Person* and these *Structural Location* subclasses. These relationships are subproperties that specialize the general *occupies* property between *Person* and *Location*. For example, a *Person* has the specific relationship *livesAt* with the *Structural Location* subclass *Residence*.

A *Residence* entity is defined conceptually as a building used as a home or dwelling. It is the place where a person lives on a permanent basis for some interval of time as distinguished from a place of temporary settlement. A person may have many *livesAt* relationships with class *Residence* as needed to model their residential history. Each of these relationships is associated with a time interval indicating the period of residency. Treating the concept of 'residency' as a relationship provides semantic flexibility as it can have a specialized inverse. The *hasResident* subproperty specifies the relationship of *Residence* to *Person* and represents an occupant tenure point of view of a residence. The relationships between person *livesAt* residence, and its inverse *hasResident* allow for inferences about instances such that if *a livesAt b* we can infer that *b hasResident a*. The inverse relationship shares the same temporal duration as its inverse.

In a similar manner, a *Workplace* is defined as a structure where a person participates in labor activities for wages and is the physical space where a person's place of business or other wage paying organization exists. A *Person* is related to a *Workplace* through the property, *worksAt* which captures the perspective of the person. The inverse property, *hasEmployee* captures the perspective of the workplace. A home-based location of a self-employed person can be accommodated as a person can have both the relationship *worksAt* and *livesAt* with *Residence*. A person can have multiple *worksAt* relationships with *Structural Locations,* which in the aggregate, describe a person's employment location history. Likewise, a *SchoolBuilding* is a structural location that contains an institution/organization that provides instruction and where a person engages in classes or educational activities. While a collection of school buildings can be viewed as a campus, this entity represents a single building occupying a bounded spatial region. A person has a specialized relationship with *SchoolBuidling* through the object property *attends* only if they are enrolled in classes at the *SchoolBuilding*. School staff would not be identified with this relationship as they have a *worksAt* relationship with the structure. A home-based private educational program can be accommodated as a person may have a relationship, *homeschoolsAt* with *Residence*. A selected set of axioms for these classes and relationships would include:

SubClassOf(*:StructuralLocation :Location*)

SubClassOf(*:Residence, : StructuralLocation*)

SubClassOf(*:School, : StructuralLocation*)

SubClassOf(*: Workplace, : StructuralLocation*)

SubPropertyOf(*: livesAt, : occupies)*

SubPropertyOf(*: worksAt, : occupies)*

SubPropertyOf(*: attends : occupies)*

InverseObjectProperty(*:livesAt :hasResident*)

ObjectPropertyAssertion(*:livesAt :Person, :Residence*)

ObjectPropertyAssertion(*:homeSchoolsAt :Person, :Residence*)

ObjectPropertyAssertion(*:worksAt :Person, :Residence*)

ObjectPropertyAssertion(*:attends:Person, :School*)

ObjectPropertyAssertion(*:workAts:Person, :Workplace*)

The third key class is *Agent*. An *Agent* is a subclass of Cohen-Hubal's *Stressor* entity responsible for creating both an *Environmental Event* and a *Noteworthy Toxic Event.* A *Noteworthy Toxic Event* (NTE) occurs when an *Agent* of interest exceeds a user defined threshold for a specific interval of time.

Declaration(Class(*a:Agent*))

SubClassOf(*:Agent :Stressor*)

Declaration(Class(*:EnvironmentalEvent*))

SubClassOf(*:EnvironmentalEvent, :SpatialOccurent*)

SubClassOf(*:NTE : EnvironmentalEvent*)

An NTE is established by the detection of the presence (or absence) of a toxin as a binary value, or when a sensor reports a value that exceeds (or falls below) an established safety threshold or health standard. An NTE is a subclass of *Environmental Event*. *Environmental Events* represent a significant change in a value measured by one or more sensor nodes. For example, a significant increase or decrease of arsenic in well water between two sampling dates would qualify as an *Environmental Event*. This *Environmental Event* becomes an arsenic NTE if the arsenic level in the well exceeds a specified maximum contaminant limit (MCL). *Agents* are linked to *Environmental Events* and NTEs through the *hasAgent* relationship and its inverse, *agentOf such that:*

ObjectPropertyAssertion(*:agentOf, :Agent, :EnvironmentalEvent*)

InverseObjectProperty(*:agentOf, :hasAgent*)

In this work, *Environmental Events* and NTEs may be detected and extracted explicitly from a sensor data stream. An *Environmental Event* or NTE might represent an event in a data stream associated with an air-based agent (e.g. carbon monoxide), a water based agent (e.g. arsenic), or a soil-based agent (e.g. dioxin or pesticide). To indicate sources pathways for *Environmental Events* and NTEs, these events have a relationship *hasSourcePathway* to a *Pathway*. These events as subclasses of a *Spatial Occurrent* inherit the *occursAt* relationship to a spatial setting and by definition as an event they have a temporal setting. An *Environmental Event* and NTE share the location of a sensor platform which provides their spatial setting. An instance of a *Spatial Occurent*, and by extension an *Environmental Event* or NTE, can occur at only one location and time. The *occursAt* relationship is therefore modeled by the functional object property. This relationship between an event and a location has an associated temporal interval. In the case of an NTE, the associated temporal interval indicates the period during which an *Agent* exceeded recommended concentration levels. The inverse of *occursAt* is not a functional property as a *Location* can be associated with many NTEs that have occurred over time and may be generated by the same or different *Agent*. The *experiences* property between a *Location* and a *Spatial Occurrent, Environmental Event* or NTE is defined as the inverse of *occursAt* but not as a functional property such that:

FunctionalObjectPropertyAssertion(*:occurAts,:SpatialOccurent :Location*)

InverseObjectProperty(*:occursAt, :experiences*)

ObjectPropertyAssertion(*:experiences,:Location,* :*SpatialOccurent)*

As a subclassOf property, any event in the *Spatial Occurent* hierarchy inherits the *occursAt* relationship to a location. This relationship is not specialized by location subclasses thus any *Spatial Occurrent* and by extension an *Environmental Event* or NTE can have an *occursAt* relationship with any location in the *Location* class hierarchy. Given the axioms:

ObjectPropertyAssertion(*:occursAt,:NTE, :Location*)

InverseObjectProperty(*:occursAt, :experiences)*

Declaration (Named Individual (*:NTE211*))

Declaration (Named Individual (*:B34*))

ClassAssertion (:*Location*, :B34)

ClassAssertion (:NTE, :NTE211)

an instance of the relationship *NTE211 occursAt B34,* it can be inferred that *B34 experiences NTE211*.

Environmental events and NTEs form a set of events for a specific location that, when aggregated, define the concept of a location's *environmental profile.* Exposure risks at a particular location can be evaluated by specific types of toxic agents (e.g. carbon monoxide). We might find that a particular location had multiple toxic agents above the recommended threshold levels at the same time or low level concentrations over long intervals of time. This type of framework allows detection at a specific location during a defined interval of time. For example, this approach would be able to represent complex situations where several agents in the air (i.e. ozone and particulate matter) may have reached unhealthy levels at the same time as arsenic was measured above the threshold in the drinking water and dioxin presented a high level of risk in the location's soil**.**

4.4 Exposure Events

The previous sections described relationships between entities which allow for the representation of person-location events and NTE-location events. An *Exposure Event* is defined as an instance of contact between a NTE and a *Person* in time and space. In this ontology, a putative exposure event occurs *if and only if* a person's location event intersects in space and time with a NTE. A putative exposure event is discovered by matching the locations of person-location relationships with locations at which NTEs occurred and determining if the temporal intervals of these events intersect. If an intersection is discovered, an *Exposure Event* is created with location determined by the shared location of *Person* and NTE and the time period defined by the interval of temporal overlap. For example consider the following facts for instances of *Person* and NTE:

ObjectPropertyAssertion(*:livesAt, :Person101, :Residence22)*

[January 20, 1984 : December 20, 1990].

ObjectPropertyAssertion(*:experiences,:Residence22, :NTE12387)*[September 12, 1988: January 27 1991]

In this scenario, Person101 lives at Residence 22 from January 1984 to December 1990. An NTE occurs at Residence 22 from September, 1988 through January 27^{th} 1991. The temporal intervals associated with the relationships overlap according to Allen's interval relations [2]. Given the presence of an overlap relation, an inference can be made that a putative exposure event has occurred. Such an exposure event is represented in the ontology through the relationship *exposedTo* between a *Person* and an NTE along with an explicit inverse of an NTE *hasExposed* a *Person*. An *Exposure Event* as a subclass of Spatial Occurrent inherits the *occursAt* relation to the shared location. The duration of this exposure is the intersection of the location-person relationship's temporal interval and the NTE temporal interval, such that:

ObjectPropertyAssertion(*:exposedTo, :Person101, :NTE123872)*

[September 12, 1988 : December 20, 1990].

5.0 The RDF Store

The ontology was created in Protégé (v.3.4) [43], converted to ntriples and imported into an Allegrograph (v.3.2) RDF store [2].

The ontology essentially serves as a schema for the RDF store. Relationships specified by the ontology (e.g *livesAt,*) directly correspond to RDF properties connecting subject and object. In the ontology, certain relationships are characterized as having a temporal interval indicating the duration for which the relationship holds. Following work described in Perry [38] and Gutierrez et al. [20], temporal information is incorporated by labeling relationship instances with their valid times. These time intervals are grounded within a discrete, linearly ordered timeline. A temporal context is established for a relationship through RDF reification. The RDF reification construct allows one to make statements about statements. In this case, a statement asserts that a given RDF statement has a valid time. The temporal structure for representing the valid time of an RDF statement employs many concepts and relations from the OWL-Time ontology [23].

6.0 Using SPARQL on the PEH

Once populated with data from various sources, the resulting RDF store can be queried. SPARQL is the W3C standard for querying RDF data [49]. A SPARQL query is comprised of the following elements: 1) prefix declarations for abbreviating URIs, 2) dataset definitions that state what RDF graph is being queried, 3) a result clause that identifies what information to return from the query, 4) a query pattern that specifies what to query in the underlying graph dataset, and in some cases 5) query modifiers for slicing, ordering, and rearranging query results. The results of SPARQL queries can be returned in a number of formats including: XML, JSON, RDF, and HTML. A simple SPARQL query example that returns all persons (subject) and names (objects) connected by the property/relationship (lastName) is as follows:

```
SELECT ?Person ?name
WHERE { ?Person exposure:lastName ?name}
```

To explore a set of person-location relationship instances (residence locations) for a single individual along with associated temporal context (residence tenure) for a specific instance of location, the natural language expression might be phrased *"How long did person p reside at location l? (Person 8, Residence 10).*

The SPARQL query result provides a conceptual building block visualized as an RDF graph (Figure 3) using Allegrograph (v. 3.2) and Gruff (v.1.4.1) [2][19].

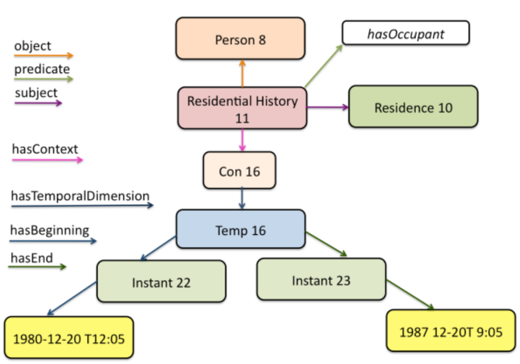

Figure 3. Instance of person-residence relationship

An example of a full residence history query for a specific instance of person, *"Where are the locations person p has resided?"* would translate into the following SPARQL query and resulting graph (Figure 4).

```
SELECT ?Person ?Residence
WHERE {?Person <exposure:PID> "P008".
       ?Person <exposure:livesAt> ?Residence}
```

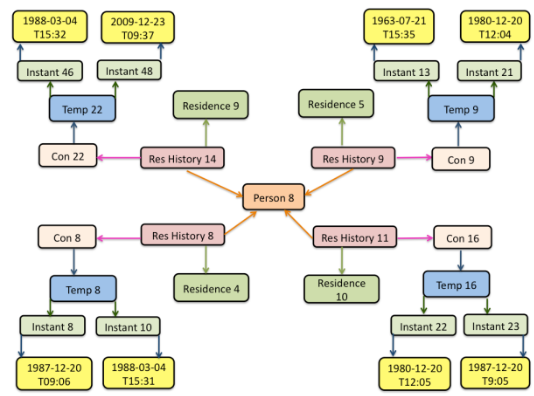

Figure 4. Person-location history (residence with tenure)

These query examples use person-residence relationships but any other person-location relationship (work or school) would provide the same type of information. Although person-person relationships are not explicitly represented in this preliminary stage of the ontology, researchers may also want to explore relationships between people and the locations they may have in common. For example, the user might want to know what people might have relationships to one another as occupants of a residence at any interval in time or what schools did members of the same family attend. These types of person-person relationships can be explicitly added to the framework later depending on the availability of the data and the needs of the user. Similarly, relationships between environmental events, NTEs and locations can be explored. Within the domain of exposure science, users will be particularly interested in detecting when toxic agents have been measured above a certain threshold by sensors at associated locations. The set of these NTE events for a single location can be thought of as a location's *toxic event profile*. An agent can also be queried by the locations where it is present at user specified levels (Figure 5).

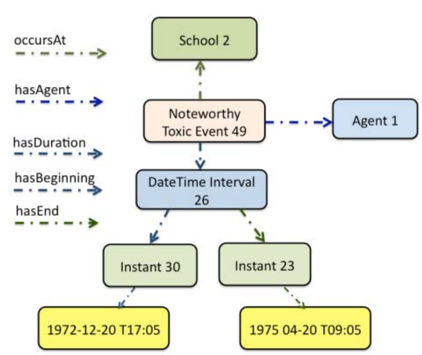

Figure 5. Instance of location-NTE relationship

Researchers need to determine if people have come into contact with toxic agents in order to investigate the health related outcomes of those interactions. In the PEH, a single intersection of a person-location event and an NTE is represented as an *Exposure Event* (Figure 6). These person-toxic event relationships form the basis for inferences about a person's contact with toxic agents in single events.

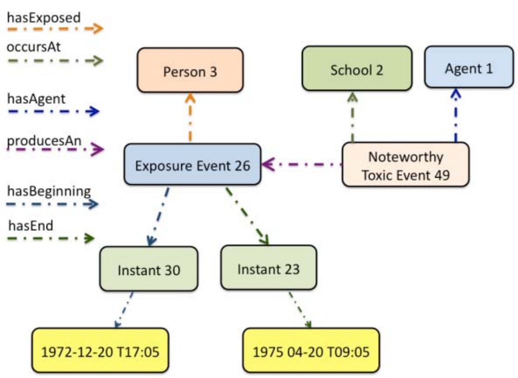

Figure 6. Instance of single exposure event

Ultimately, the knowledgebase needs to represent the aggregate of all exposure events for a specific instance of person thus presenting a researcher with a PEH for a single individual (Figure 7). The results provide the user with information on the person's exposure to a number of different toxic agents at different locations along with the temporal context of the exposure periods. Personal exposure histories, like other aggregate concepts (i.e. person-location history and location environmental profile), can be generated from this type of query and then added as new RDF statements to update the knowledgebase with this derived information for later use and modeling.

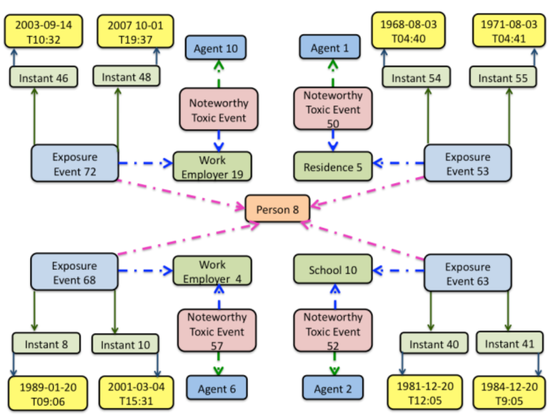

Figure 7: PEH for single individual

7. CONCLUSIONS

This preliminary work on the PEH ontology is intended to provide a conceptual framework for the identification of event-event intersections between data on human mobility over long intervals of time and environmental event data at specific locations. The primary purpose of the ontology is to define important concepts and relationships that contribute to an individual level exposure history. Concepts in the framework are developed based on the supposition that people have specific relationships with locations (residency, employment, attendance) that convey more specific semantics about individual's differing behaviors and associations with these locations. These explicit spatio-temporal event-event relationships can provide critical information for the modeling of toxic exposures of individuals to environmental health risks from multiple sources. The RDF store developed from the ontology allows for direct queries on these relationships using the SPARQL query language.

This framework provides semantic clarity about the specific type of relationship a person has with an associated location and the types of behaviors that can be inferred from that relationship (i.e. sleeping, showering, eating, drinking, working, etc.). This increased inference capacity has implications for understanding what an individual's exposure risks might be in one setting over another depending on the relationships between a person and the location. The capacity to represent exposure risk over time for individuals also presents the opportunity to aggregate common locations among groups of people based on shared relationships with locations in their past (i.e. shared residence, shared workplace, school building cohort). It then becomes possible to identify and model risk groups based on their relationships to specific locations.

This ontology based approach links people with locations through a specific relationship to those locations within a valid temporal context. It provides a novel way to evaluate environmental health risks beyond the traditional person to location layer approach used in many existing health information systems. The design explicitly allows for expansion and incorporation of additional ontologies representing sensor networks, toxic agent-genetic interactions, and other related medical or environmental health knowledgebases. Future work will include a data driven evaluation process to test inference and query performance with a variety of data (i.e. tumor registry data, genetic profiles, school enrollment registries, sensor network data). This process will also identify and address potential issues that might arise with various user roles [8]. The integration of diverse data sources raises an important limitation of this preliminary work regarding the evaluation of data quality and the ways in which users might more clearly weight data based on source reliability and uncertainty measures. Likewise, this preliminary framework does not yet account for critical bioinformatics issues such as accounting for data provenance. However, this approach does have the potential to provide health researchers with an efficient and cost saving solution for the widespread problems associated with the integration and analysis large and diverse datasets and presents structures to evaluate and model toxic exposures at the individual level over space and time.

8. ACKNOWLEDGMENTS

This work was supported by the Department of Defense's Maine Cancer GIS project (DoD Grant # W81XWH-07-2-0116) and the National Science Foundation's GK-12 Sensors! program (NSF-DGE grant # DGE-0538457).

9. REFERENCES

[1] Aleman-Meza, B., Nagarajan, M., Ramakrishnan, C., Ding, L., Kolari, P., Sheth, A.,Arpinar, I. B., … and Finin, T. Semantic analytics on social networks: Experiences in addressing the problem of conflict of interest detection. In L. Carr, D. De Roure, A. Iyengar, C. A. Goble, & M. Dahlin,

(Eds.). *15th International World Wide Web Conference* (Edinburgh, Scotland 2006), ACM, 407–416.

[2] AllegroGraph,(Version 3.2), Oakland, CA: Franz, Inc. http://www.franz.com/products/allegrograph

[3] Allen, J.F. (1983). Maintaining knowledge about temporal intervals. In *Communications of the ACM, 26* (pp. 832-843). New York: ACM Press. ISSN 0001-0782 Retrieved from http://www.isi.edu/isd/kr/ALLEN.pdf

[4] Arpinar, B., Sheth, A., Ramakrishnan, C.,Usery, E.L., Azami, M., and Kwan, M. Geospatial Ontology Development and Semantic Analytics. *Transactions in GIS*, *10*, (4) 551–575.

[5] Beard, K., Deese, H., and Pettigrew, N.R. A framework for visualization and exploration of events. *InformationVisualization*, *7*, 133-151. DOI= doi:10.1057/palgrave.ivs.9500165

[6] Berners-Lee, T. *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor*. HarperCollins, (San Francisco, 1999).

[7] Berners-Lee, T., Hendler, J., and Lassila, O. The semantic web. *Scientific American*, *184*, (5). 2001 34-43.

[8] Brewster, C. Alani, H., Dasmahapatra, S. and Wilks, Y.,. Data driven ontology evaluation. Proceedings of International Conf. on Language Resources and Evaluation, Lisbon, 2004.

[9] Brochhausen, M., Weiler, G., Martín, L., Cocos, C.,Stenzhorn, H., Graf, N., Dörr, M., Tsiknakis, M., and Smith, B. In *Lecture Notes in Computer Science: Vol. 5333*. Applications of the ACGT Master Ontology on Cancer. (Berlin/Heidelberg, 2009): Springer-Verlag, 2009 1046-1055. DOI = doi:10.1007/978-3-540-88875-8_132

[10] Centers for Disease Control. Epi Info Version 3.5.1.Retrieved September 10, 2008 from http://www.cdc.gov/epiinfo/

[11] Cohen Hubal, E. A. Biologically relevant exposure science for 21st century toxicity testing. *Toxicological Sciences. Society of Toxicology*, 111(2), 226-232.

[12] Delin, K.A. The Sensor Web: A macro-instrument for coordinated sensing, *Sensors*, *2*, 275.

[13] Doore, S.A. Modeling a personal exposure history through event-event relationships. Masters Thesis. University of Maine. 2010

[14] Fonseca, F., Egenhofer, M., Davis, C., and Câmara, C. Semantic granularity in ontology-driven geographic information systems. *Annals of Mathematics and Artificial Intelligence, 36*, (1-2), 121-151.

[15] Ford, D.A., Kaufman, J.H. and Eiron, I. 2006.An extensible spatial and temporal epidemiological modeling system. *International Journal of Health Geographics*, 5, (4). doi:10.1186/1476-072X-5-4

[16] Grenon, P. Spatio-temporality in Basic Formal Ontology: SNAP and SPAN, Upper Level Ontology, and Framework for Formalization Part I. IFOMIS Reports. University of Leipzig (Leipzig, Germany, 2003).

[17] Grenon, P., Smith, B., and Goldberg, L. Biodynamic ontology: Applying BFO in the biomedical domain. In D. M. Pisanelli, (Ed.), *Ontologies in Medicine, of Studies in Health Technology and Informatics*, 102. IOS Press, (Amsterdam, 2004) 20–38.

[18] Gruber, T.R. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5,(2), 199-220.

[19] Gruff (version 1.4.1), Oakland, CA: Franz, Inc. http://www.franz.com/agraph/gruff

[20] Gutierrez, C., Hurtado, C., & Vaisman, A. Introducing time into RDF. *IEEE Transactions on Knowledge and Data Engineering*, 19, (2), 207–218. DOI= doi:10.1109/TKDE.2007.34

[21] Hägerstrand, T. What about people in regional science? *Papers of the Regional Science Association*, *24*, 1–21. DOI = doi: 10.1007/BF01936872

[22] Hariharan, R. & Hornsby,K. Modeling intersections of geospatial lifelines. *1st International Conference on Geographic Information Science (*Savannah, GA, October 28-31, 2000), GIScience'00.

[23] Hobbs, J. & Pan, F. An ontology of time for the semantic web. *ACM Transactions on Asian Language Information Processing*, *3*,(1), 66-85.DOI doi.acm.org/10.1145/1017068.1017073

[24] Hornsby, K. Shifting granularity over geospatial lifelines, *AAAI Workshop on Spatial and Temporal Granularity*, (Technical Report WS-00-08), AAAI Press (Menlo Park, CA,2000).

[25] Hornsby, K. & Egenhofer, M. Modeling moving objects over multiple granularities, *Annals of Mathematics and Artificial Intelligence*, 36,(1-2) 177-194.

[26] Huang, T., Shenoy, P.J., Sinha, R., Graiser, M., Bumpers, K, W. and Flowers, C. R. Development of the Lymphoma Enterprise Architecture Database: A caBIG™ Silver level compliant system. Cancer Informatics 8, 45-64.

[27] International Organization for Standardization (2003). Geographic information -- Spatial referencing by geographic identifiers (ISO 19112:2003) Geneva, Switzerland: Retrieved from http://www.iso.org/iso/catalogue

[28] Jacquez, G.M., Greiling, D.A. and Kaufmann, A.M. Design and implementation of a space-time intelligence system for disease surveillance. *Journal of Geographical Systems*, 7, 7-23. DOI = doi: 10.1007/s10109-005-0147-6

[29] Lioy P., Leaderer, B., Graham, J., Lebret, E., Sheldon, L., Needham, L., Pellazzari, E., and Lebowitz, M. The Application of Exposure Assessment to Environmental Health Science and Public Policy F What has been Accomplished and What Needs to Happen before Our 25th Anniversary in 2014. *Journal of Exposure Analysis and Environmental Epidemiology, 15*, 121–122. DOI = doi:10.1038/sj.jea.7500417

[30] Lioy, P., Lebret, E., Spengler, J., Brauer, M., Buckley, T., Freeman, N., Jantunen, M., …, and Zmirou-Navier, D. Defining exposure science. *Journal of Exposure Analysis and Environmental Epidemiology, 15*, 463. DOI = doi:10.1038/sj.jea.7500463.

[31] Manola, F. and Miller E., (Eds). 2004. RDF Primer. W3C Recommendation, 10 February 2004. http://www.w3.org/TR/2004/REC-rdf-primer-20040210/.

[32] Mark, D. Egenhofer, M. Bian, L. Hornsby, K. Rogerson, P. and Vena, J. Spatio-temporal GIS analysis for environmental health using geospatial lifelines In *2nd International Workshop on Geography and Medicine*, (Paris, France, November 22-23, 1999) *GEOMED'99.*

[33] Meliker, J.R., Slotnick, M.J., AvRuskin, G.A., Kaufmann, A., Jacquez, G. M., and Nriagu, J.O. Improving exposure assessment in environmental epidemiology: Application of spatio-temporal visualization tools. *Journal of Geographical Systems*.7:49–66.DOI = doi: 10.1007/s10109-005-0149-4

[34] Meliker, J.R., Slotnik, M.J., AvRuskin, G.A., Schottenfeld, D., Jacquez, G.M., Wilson, M.L., Gooverts, P., Franzblau, A. and Nriagu, J.O. Lifetime exposure to arsenic in drinking water and bladder cancer: a population based case-control study in Michigan, USA. Cancer Causes and Control 21. 745-757.

[35] Mukherjea, S. and Bamba, B. BioPatentMiner: An information retrieval system for biomedical patents. In M. A. Nascimento, M. T. Ozsu, D. Kossmann, R. J. Miller, J. B. Blakeley, and K. B. Schiefer, (Eds.), *Proceedings of the 30th International Conference on Very Large Data Bases,* (Toronto, Canada August 31- September 3, 2004) VLDB Endowment. 1066–1077.

[36] Odland, J. Longitudinal analysis of migration and mobility spatial behavior in explicitly temporal contexts. In M. Egenhofer & R. Golledge (Eds). *Spatial and Temporal Reasoning in Geographic Information Systems.* Oxford University Press. New York, 1998. 238-259.

[37] O'Hara, K., Berners-Lee, T., Hall, W. and Shadbolt, N. Use of the semantic web in e-research. In Dutton, W. H. and Jeffreys, P.W. (Eds.) World Wide Research: Reshaping the Sciences and Humanities. MIT Press, Cambridge, ME. 2010.

[38] Perry, M. *A Framework to Support Spatial, Temporal and Thematic Analytics over Semantic Web Data*. Doctoral Theis, Wright State University, 2008.

[39] Perry, M., Hakimpour, F., and Sheth, A. Analyzing theme, space and time: An ontology-based approach. In R. A. de By and S. Nittel, (Eds.), *14th ACM International Symposium on Geographic Information Systems*, (Arlington, VA November 10-11, 2006) Springer-Verlag, 147–154.

[40] Perry, M., Sheth, A., Arpinar, I.B & Hakimpour, F. Geospatial and temporal semantic analytics, In H. Karimia (Ed.), *Encyclopedia of Geoinformatics,* Idea Group.2008,

[41] Perry, M., Sheth, A.P., Hakimpour, F., and Jain, P. What, Where and When: Supporting Semantic, Spatial and Temporal Queries in a DBMS Technical Report: (KNOESIS-TR-2007-01) Kno.e.sis Center Technical Report, Dayton, OH:Wright State University. 2007.

[42] President's Cancer Panel. *2008-2009 Annual report: Reducing Environmental Cancer risk: What we can do now*. National Cancer Institute. National Institutes of Health. U.S. Department of Health and Human Services. May 2010.

[43] Protégé (Version 3.4.1). Stanford, CA: Stanford Center for Biomedical Informatics Research at the Stanford University School of Medicine. http://protege.stanford.edu

[44] Sinha, G. and Mark, D.M. Measuring similarity between geospatial lifelines in studies of environmental health. *Journal of Geographic Systems, 7*, 115–136 DOI = doi: 10.1007/s10109-005-0153-8

[45] Spear, A, D. Ontology for the Twenty First Century: An Introduction with Recommendations. Saarbrücken, Germany. Retrieved January 3, 2007 from http://www.ifomis.org/bfo/manual.pdf

[46] Worboys, M.F. & Duckham, M. (2006). Monitoring qualitative spatiotemporal change for geosensor networks. *International Journal of Geographical Information Science*, *20*, 1087-1108.

[47] Worboys, M., and Hornsby, K. From objects to events: GEM, the geospatial event model. In M.J. Egenhofer, C. Freksa, and H.J. Miller (Eds.), *Lecture Notes in Computer Science: Vol. 3234. GIScience 2004 (*Berlin,Heidelberg, Germany, October 20-23, 2004) Springer-Verlag. 327-343*.*

[48] World Wide Web Consortium. RDF syntax. Retrieved March 13, 2007 from http://www.w3.org/TR/PR-rdf-syntax

[49] World Wide Web Consortium. SPARQL Query language for RDF. Retrieved June 5, 2009 from http://www.w3.org/TR/rdf-sparql-query/

[50] Zartarian VG, Özkaynak H, Burke JM, Zufall MJ, Rigas ML, and Furtaw Jr. EJ. (2000). A Modeling Framework For Estimating Children's Residential Exposure and Dose to Chlorpyrifos Via Dermal Residue Contact and Non-Dietary Ingestion. *Environmental Health Perspectives* 108(6): 505-514.