Provenance Aware Linked Sensor Data

Harshal Patni, Satya S. Sahoo, Cory Henson, Amit Sheth

Kno.e.sis Center, Computer Science and Engineering Department, Wright State University, Dayton, OH 45435 USA
{harshal, satya, cory, amit}@knoesis.org

Abstract. Provenance, from the French word "*provenir*", describes the lineage or history of a data entity. Provenance is critical information in the sensors domain to identify a sensor and analyze the observation data over time and geographical space. In this paper, we present a framework to model and query the provenance information associated with the sensor data exposed as part of the Web of Data using the Linked Open Data conventions. This is accomplished by developing an ontology-driven provenance management infrastructure that includes a representation model and query infrastructure. This provenance infrastructure, called *Sensor* Provenance Management System (PMS), is underpinned by a domain specific provenance ontology called Sensor Provenance (SP) ontology. The SP ontology extends the Provenir upper level provenance ontology to model domain-specific provenance in the sensor domain. In this paper, we describe the implementation of the Sensor PMS for provenance tracking in the Linked Sensor Data.

Keywords: Provenance Management Framework, provenir ontology, Provenance, Lineage, Linked Data, Semantic Sensor Web, Sensor Data, Sensor Web Enablement, Dataset Generation, Resource Description Framework (RDF)

1. INTRODUCTION

The first North American blizzard of 2010 was tracked from the state of California to Arizona, through northern Mexico, and across the continental United States. The storm produced historic snowfall levels in the Mid-Atlantic States, as well as extensive flooding and landslides in Mexico. During this time, a number of weather stations collected data from thousands of sensors deployed in the United States. Semantic Sensor Web[1] proposes to annotate this sensor data with semantic metadata to provide contextual information essential for situational awareness. Such semantic metadata data can be used to answer aggregate queries spanning both temporal and geographical areas.

Let us consider the following scenario. *We are interested in finding all the sensors which have observations related to a blizzard of interest. In order to accomplish this task, we would need to know the properties associated with a phenomenon to be classified as a blizzard, the time period for which the blizzard was active, the location where the blizzard occurred, and sensors deployed in this location during this time period.*

[1] http://wiki.knoesis.org/index.php/SSW

This is an example of a sensor discovery query. Sensor discovery has been identified as a top-priority use case by the W3C Semantic Sensor Network Incubator Group[2], which is tasked with development of sensor ontology. In the sensors domain, the capabilities of the sensor, observation location (spatial parameter), time of observation (temporal parameter), and phenomenon measurement (domain parameter) are important to answer discovery queries. This data related to the sensor is the provenance metadata about the sensor. Provenance describes the history or the lineage of an entity and is derived from the French word "*provenir*" meaning "to come from". Provenance information enables applications to answer the "what", "where", "why", "who", "which", "when", and "how" queries to accurately interpret and process data entities.

Provenance has been studied from multiple perspectives, including (a) workflow provenance and (b) database provenance as discussed in Tan [1]. Workflow provenance represents "the entire history of the derivation of the final output of" [1] a workflow. Davidson et al. [2] addresses issues related to provenance in workflow systems. In contrast, database provenance refers to the process of tracing and recording the origins of data and its movement between databases [3]. In Sahoo et al. [4], we introduced the notion of semantic provenance to define provenance information that incorporates domain semantics to closely reflect the knowledge of an application domain.

In this paper, we use the observations from the 20,000 sensors within the United States (Figure 1) in the context of a blizzard as a running example.

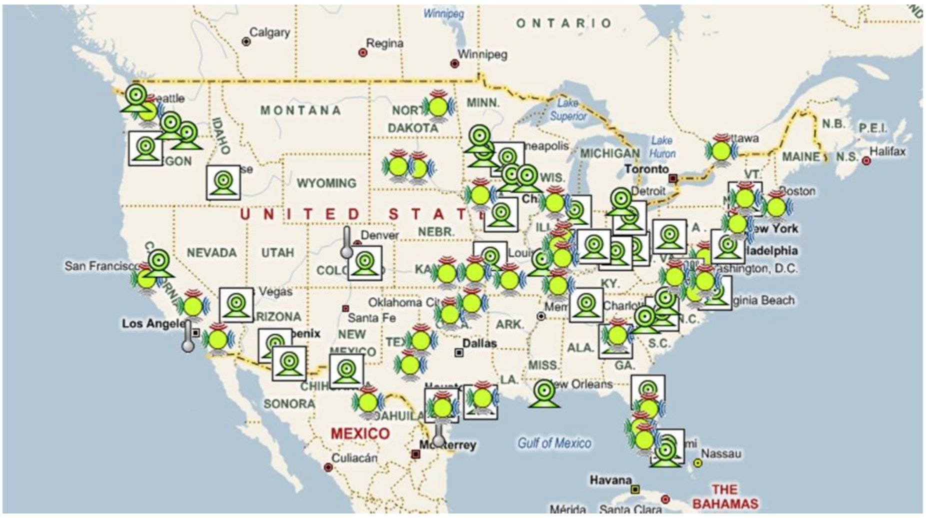

Fig.1. The distribution of 20,000 Sensors constituting the Semantic Sensor Web (SensorMap Image [5])

We use the definition of a blizzard provided by the NOAA[3], which describes it as:

BLIZZARD = High WindSpeed (exceeding 35 mph) AND Snow Precipitation AND Low Visibility (less than ¼ mile), for at minimum 3 hours.

Fig.2. Blizzard Composition

---

[2] http://www.w3.org/2005/Incubator/ssn/wiki/Main_Page

[3] http://www.noaa.gov/

A blizzard exists if the above conditions hold true for at least 3 hours within some geospatial region. Hence, the provenance of sensor observations describing the geospatial information of the sensors that record the observations, the time stamp of the observations, and the attributes of the sensor itself (for example, a motion sensor is not useful in context of a blizzard) are important for a sensor discovery query.
With a view of capturing the provenance information related to a sensor, the main objective of this paper is to implement a Sensor Provenance Management System (Sensor PMS). In this paper, we describe the creation of this infrastructure using the theoretical underpinning of the Provenance Management Framework (PMF) [4]. The key contributions of the paper are described below:

1. Implementing Sensor PMS to track provenance in the linked sensor data
2. Developing a domain specific ontology for Sensor PMS called Sensor Provenance (SP) ontology. The SP ontology uses concepts within the Provenir upper level ontology defined in PMF [4] to add provenance information within the sensors domain.
3. An evaluation of the Sensor PMS capabilities to answer provenance queries over the sensor datasets generated is provided.

The rest of the paper is organized as follows: Section 2 discusses background concepts. In section 3, we describe current infrastructure for generating sensor datasets and section 4 discusses the sensor datasets generated. Section 5 integrates the current infrastructure described in section 3 with the provenance management system and describes the architecture of Sensor PMS. Section 6 introduces the SP ontology and section 7 discusses the kind of queries that can be answered with the help of provenance information. Section 8 gives related work and section 9 concludes with summary and future work.

2. Background

In this section, we describe the resources used in our work including the Sensor ontology and the Linked Open Data initiative.

2.1 Ontology Model of Sensor Data – In computer science and information science, ontology is a formal representation of the knowledge by a set of concepts within a domain and the relationships between those concepts. It is used to reason about the properties of that domain, and may be used to describe the domain. [6] Our sensors ontology uses the concepts within the O&M standard to define sensor observations. Within the O&M standard, an observation (*om:Observation*) is defined as an *act of observing a property or phenomenon, with the goal of producing an estimate of the value of the property*, and a feature (*om:Feature*) is defined as an *abstraction of real world phenomenon*. (Note: *om* is used as a prefix for Observations and Measurements). The major properties of an observation include feature of interest (*om:featureOfInterest*), observed property (*om:observedProperty*), sampling time (*om:samplingTime*), result (*om:result*), and procedure (*om:procedure*). Often these properties can be complex entities that may be defined in an external document. For example, *om:FeatureOfInterest* could refer to any real-world entity such as a coverage region, vehicle, or weather-storm, and *om:Procedure* often refers to a sensor or system of sensors defined within a SensorML[4] document. Therefore, these properties are better described as relationships of an observation. Concepts described above and their relationships within the sensor ontology can be found in figure 2. The Sensor ontology can be found at [7]. Section 5 extends the Sensor Ontology with provenance related concepts found in the Provenir upper level ontology defined in the Provenance Management Framework (PMF) [4].

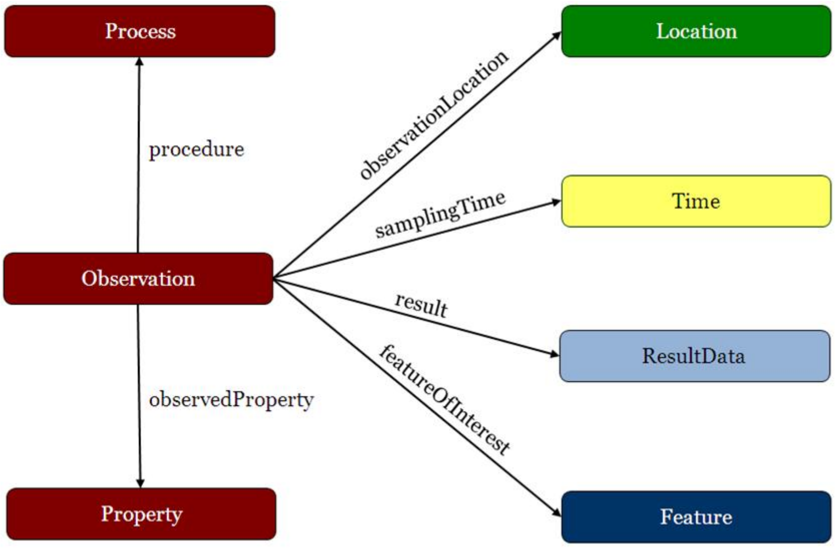

Fig.2. Concepts and their relationships within the Sensor Ontology

2.2 Semantic Web –The Semantic Web is an evolving development of the World Wide Web[5] derived from the World Wide Web consortium (W3C)[6] in which the meaning of information and services on the web is defined, making it possible for the web to understand and satisfy the request of people and machines that use the web content. [8] Resource Description Framework (RDF) is a publishing language within the Semantic Web, specially designed for data. RDF has now come to be used as a general method for conceptual description or modeling of information that is implemented in web resources, using a variety of syntax formats. [9]. It is also a standard model for data interchange on the web. [10] SPARQL[7] is a protocol and a query language for semantic web data sources. [8] In its usage, SPARQL is a syntactically-SQL-like language for querying RDF graphs. [11] Since Semantic Web is not just about putting data on the web but also linking the data, Linked Data is used to connect the Semantic Web[8]. Wikipedia defines Linked Data as "*a term used to describe a recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and RDF*." [12] Linked Data is a large and growing collection of interlinked public datasets encoded in RDF spanning diverse areas such as: life sciences, nature, science, geography and entertainment.

---

[4] http://www.opengeospatial.org/standards/sensorml

[5] http://en.wikipedia.org/wiki/World_Wide_Web

[6] http://www.w3.org/

[7] http://www.w3.org/TR/rdf-sparql-query/

[8] http://www.w3.org/DesignIssues/LinkedData.html

3. Current Infrastructure

The lifespan of sensor data starts as observable properties of objects and events in the real-world which are detected by sensors through observation. These observation values are then encoded in several formats of varying degrees of expressivity, as needed by applications that may utilize the data. The data generation workflow is comprised of four main parts, as shown in figure 3. The workflow begins with sensors deployed across the United States measuring environmental phenomena. Observations generated from these sensors are aggregated at MesoWest [13] which provides access to past sensor observations encoded as comma separated numerical values. These sensor observations are then converted to Observations and Measurements (O&M). O&M is an encoding standard and a technical framework that defines an abstract model and an XML schema encoding for sensor descriptions and observations. It is one of OGC[9] Sensor Web Enablement (SWE)[10] suite of standards that is widely accepted within the sensors community for encoding sensor observations. [14] In order to add semantics to the sensor descriptions and observations the O&M is converted to RDF. O&M is converted to RDF using the *O&M2RDF-Converter API* described in [15]. Two RDF datasets, *LinkedSensorData* and *LinkedObservationData* containing over a billion triples were generated. The datasets are described in the section 4. The RDF generated is then stored in a Virtuoso RDF knowledgebase [16]. The RDF datasets are made available on the Linked Open Data Cloud to provide public access. The data generation workflow is the main component of the Provenance Capture phase discussed in Section 5.

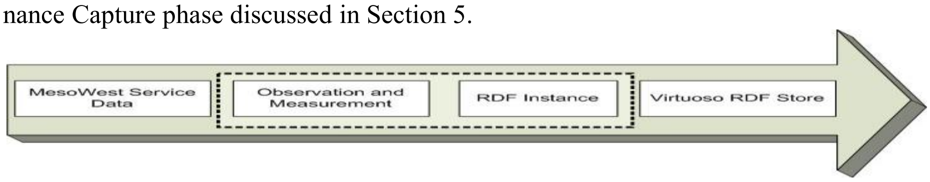

Fig.3. Data Generation Workflow. The O&M to RDF conversion (dotted portion) forms the main part of the workflow that uses the *O&M2RDF-CONVERTER API*.

3.1 Phase 1 – The first phase is comprised of querying MesoWest [13] for observational data and parsing the result. MesoWest provides a service to access past sensor data and returns an HTML page with the observational values encoded within a comma-separated list. The resulting HTML page is then parsed to extract the sensor observations.

3.2 Phase 2 – The second phase consists of converting the raw textual data retrieved from MesoWest into O&M. The sensor observations parsed from the HTML page in phase 1 are fed to an XML parser. We used the SAX (Simple API for XML) parser[11] to generate the O&M. Here we also query GeoNames [17] with the sensor coordinates to get GeoNames location that is closest to the sensor. The O&M generated in this phase is the input for the *O&M2RDF-Converter API*.

[9] http://www.opengeospatial.org/
[10] http://www.opengeospatial.org/projects/groups/sensorweb
[11] http://www.saxproject.org/

3.3 Phase 3 – The third phase consists of converting sensor observations encoded in O&M to RDF. Since both O&M and RDF have XML syntax, XSLT is used to convert O&M to RDF. XSLT is a language for transforming XML documents into other XML documents [18]. The XSLT performs the conversion for our *O&M2RDF-Converter* API.

3.4 Phase 4 - The fourth phase consists of storing the RDF in Virtuoso RDF store. Virtuoso RDF is a native triple store available in both open source and commercial licenses. It provides command line loaders, a connection API, support for SPARQL and web server to perform SPARQL queries and uploading of data over HTTP. It has been tested to scale up to a billion triple. A more detailed description of the data generation workflow can be found in [15].

4. Sensor Dataset Description

The data generation workflow described in section 3 lead to the generation of 2 RDF datasets *LinkedSensorData* and *LinkedObservationData* containing over a billion triples described in detail below.

4.1 **Linked Sensor Data** - LinkedSensorData is an RDF dataset containing expressive descriptions of ~20,000 weather stations in the United States. The data originated at MesoWest, a project within the Department of Meteorology at the University of Utah that has been aggregating weather data since 2002. [13] On average, there are five sensors per weather station measuring phenomena such as temperature, visibility, precipitation, pressure, wind speed, humidity, etc. In addition to location attributes such as latitude, longitude, and elevation, there are links to locations in Geonames [17] near the weather station. The distance from the Geonames location to the weather station is also provided. The data set also contains links to the most current observation for each weather station provided by MesoWest [13]. This sensors description dataset is now part of the LOD.

4.2 **Linked Observation Data** - LinkedObservationData is an RDF dataset containing expressive descriptions of hurricane and blizzard observations in the United States. The data again originated at MesoWest. [13] The observations collected include measurements of phenomena such as temperature, visibility, precipitation, pressure, wind speed, humidity, etc. The weather station's observations also include the unit of measurement for each of these phenomena as well as the time instant at which the measurements were taken. The dataset includes observations within the entire United States during the time periods that several major storms were active -- including Hurricane Katrina, Ike, Bill, Bertha, Wilma, Charley, Gustav, and a major blizzard in Nevada in 2002. These observations are generated by weather stations described in the LinkedSensorData dataset introduced above. Currently, this dataset contains more than a billion triples. The RDF dataset for each of the above storms is available for download in gzip format at [19]. The statistics for each of the storms can also be found at [19]

5. Sensor Provenance Management System

The Sensor PMS infrastructure uses the data generation workflow described above (section 3) and addresses three aspects of provenance management as identified by [20]. See Figure 4 for an architecture of Sensor PMS.

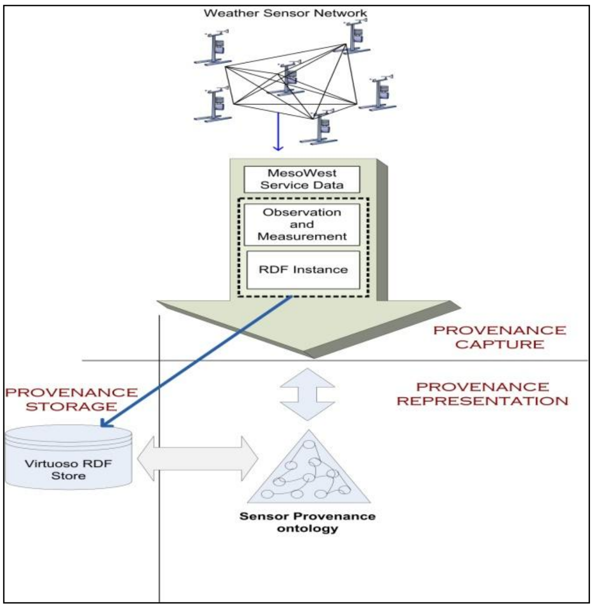

Fig.4. The architecture of the *Sensor* PMS addressing Three aspects of provenance management

1. **Provenance Capture** – The provenance information associated with the sensor is captured within the data workflow as described in section 3. The time related information (temporal parameter) is obtained from MesoWest [13] and location related information (spatial parameter) is obtained by querying GeoNames [17] with the sensor coordinates.
2. **Provenance Representation** – The Sensor Provenance ontology (SP) is used to model the provenance information related to the sensor. The SP ontology extends the Provenir upper level provenance ontology defined in PMF [4] to support interoperability with provenance ontology in different domains.
3. **Provenance Storage** – The provenance information is stored in the Virtuoso RDF store. Virtuoso RDF is an open source triple store provided by OpenLink Software.[16] The Virtuoso RDF store currently contains over a billion triples of sensor observational data. Virtuoso RDF provides a SPARQL endpoint to query these dataset discussed in section 4, which can be found at [21]. More information about querying the dataset can be found at [19].

6. Sensor Provenance Ontology

In this section we discuss the Sensor Provenance Ontology that forms the key component of the Sensor PMS. As discussed above, provenance information includes the location of the sensor, the time when the observations were taken by the sensor and the sensor observation values. Since SP ontology extends the provenir ontology, we discuss the provenir ontology in section 6.1 followed by SP ontology in section 6.2

6.1 Provenir Ontology - Provenir ontology is a common provenance model which forms the core component of the provenance management framework. [4] This modular framework forms a scalable and flexible approach to provenance modeling that can be adapted to the specific requirement of different domains. Use of Provenir ontology as the reference model to built domain-specific provenance ontologies ensures (a) common modeling approach, (b) conceptual clarity of provenance terms, and (c) use of design patterns for consistent provenance modeling

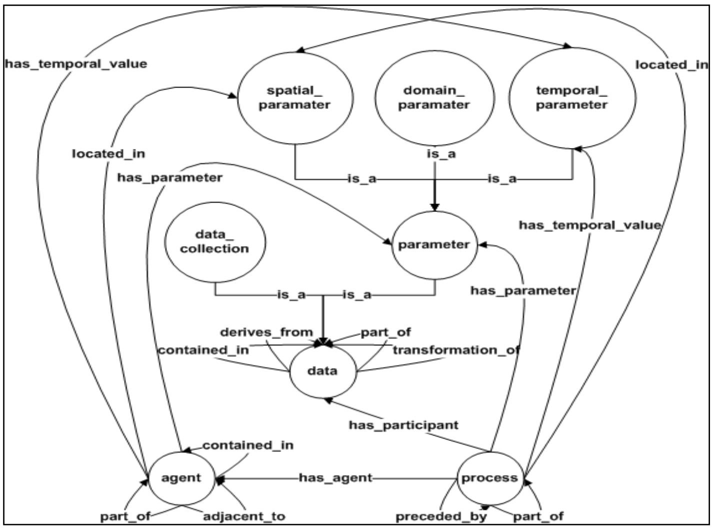

Fig.5. Provenir Upper Level Ontology [4]

The ontology defines three base classes *data*, *agent* and *process* using the well defined, primitive concepts of *occurent* and *continuant*. [22] *Continuant* is defined as "entities which endure, or continue to exist, through time while undergoing different sorts of changes, including changes of place" [22] while *Occurrent* is defined as "entities that unfold themselves in successive temporal phases". [22]. The two base classes, *data* and *agent* are defined as specialization (sub-class) of *continuant* class while the third base class *process* is a synonym of *occurent*. The *data* class has two sub-classes, *data_collection* -- that represents the datasets that undergo modification during an experiment -- and *parameter* -- that influences the execution of an experiment. The *parameter* class has three sub-classes representing the spatial, temporal, and thematic (domain-specific) dimensions, namely *spatial_parameter*, *temporal_parameter*, and *domain_parameter*. Instead of defining a new set of properties, the ontology reuses and adapts properties defined in the Relation ontology (RO)[12] from the Open Biomedical Ontologies (OBO) Foundry[13] such as *part_of*, *contained_in*, *preceded_by*, and *has_participant*. The Provenir ontology is defined using OWL-DL[14] that is complaint with the DL profile of OWL2[15], with an expressivity of

[12] http://www.obofoundry.org/ro/
[13] http://www.obofoundry.org/
[14] http://www.w3.org/TR/owl-features/

$\mathcal{ALCH}$; further details of the ontology can be found at [23]. Figure 5 shows the Provenir ontology schema obtained from [4].

5.2 Sensor Ontology - Extending Provenir Ontology

The Provenir ontology has been extended to create the Sensor ontology that models the domain-specific provenance information for the sensor domain. The Sensor ontology extends the relevant Provenir ontology terms using the *rdfs:subClassOf* and *rdfs:subPropertyOf* relationships to create appropriate classes and properties. For example, the *sensor:ResultData* (representing the observation value) is a subclass of *provenir:data_collection*, the *sensor:Location* class (representing the geographical location) is defined as a subclass of *provenir:spatial_parameter*. Similarly, *sensor:samplingTime* is defined as a subproperty of *provenir:has_temporal_value*.

The sensor ontology has been defined in OWL-DL and consists of 89 classes, 53 properties with a DL expressivity of $\mathcal{ALEHIF}+(\mathcal{D})$. By extending the Provenir ontology, the sensor ontology ensures coherent modeling of concepts, consistent use of provenance terminology, and compatibility with other existing domain-specific provenance ontologies. For example, the Trident ontology extends the Provenir ontology to model provenance information in the Neptune oceanography project [24]. In the next section, we describe the queries that utilize the provenance information modeled in the sensor ontology.

7. Provenance Queries

Two classes of Provenance queries have been categorized by PMF [4]. Corresponding queries in the sensors domain that could not be answered without provenance information have been provided.

1. **Query for provenance metadata:** Given a data entity, this category of queries returns the complete set of provenance information associated with a data entity. **Example:** "*Given an observation value, give me the provenance information about the all the sensors that recorded this observation*"

```
SELECT ?sensor ?ID ?geonamesLocation ?geonamesDistance
            ?geonamesDistanceMeasure ?latitude ?longitude
            ?observedProperty ?XSDTime
WHERE
   {?sensor om-owl:generatedObservation ?generatedObservation .
    ?generatedObservation om-owl:observedProperty ?observedProperty .
    ?generatedObservation om-owl:result ?measureData .
    ?measureData om-owl:floatValue ?value .
    FILTER(?value = "78.0"^^xsd:float) .
    ?generatedObservation om-owl:samplingTime ?timeInstant .
    ?timeInstant owl-time:inXSDDateTime ?XSDTime .
    ?sensor om-owl:ID ?ID .
    ?sensor om-owl:hasLocatedNearRel ?locatedNear .
    ?locatedNear om-owl:hasLocation ?geonamesLocation .
    ?locatedNear om-owl:distance ?geonamesDistance .
    ?locatedNear om-owl:distanceUOM ?geonamesDistanceMeasure .
    ?sensor om-owl:processLocation ?sensorLocation .
    ?sensorLocation wgs84:lat ?latitude .
    ?sensorLocation wgs84:long ?longitude .
    }
```

[15] http://www.w3.org/TR/owl2-profiles/

2. **Query for data using provenance information:** An opposite perspective to the first category of query is, given a set of constraints defined over provenance information retrieve a set of data entities satisfying some set of constraints. Example: ***"Find all the sensors which have observations related to a blizzard occurring in Nevada on 24th August 2005 at 11 AM"***
To solve this sensor discover query, provenance information describing the spatio-temporal and thematic aspects of sensor observations and sensors can be analyzed. Figure 6 describes the multiple steps followed in identifying the appropriate sensor. In Step 1, sensors located in the "Nevada" region are identified (from a pool of 20,000 sensors located across the United State). In Step 2, the sensors that were active during the blizzard are identified, and finally in Step 3 provenance information describing the capabilities of a sensor help identify the observations that are relevant for the blizzard under study (for example, a wind speed sensor is considered relevant while a motion sensor is not considered relevant.)

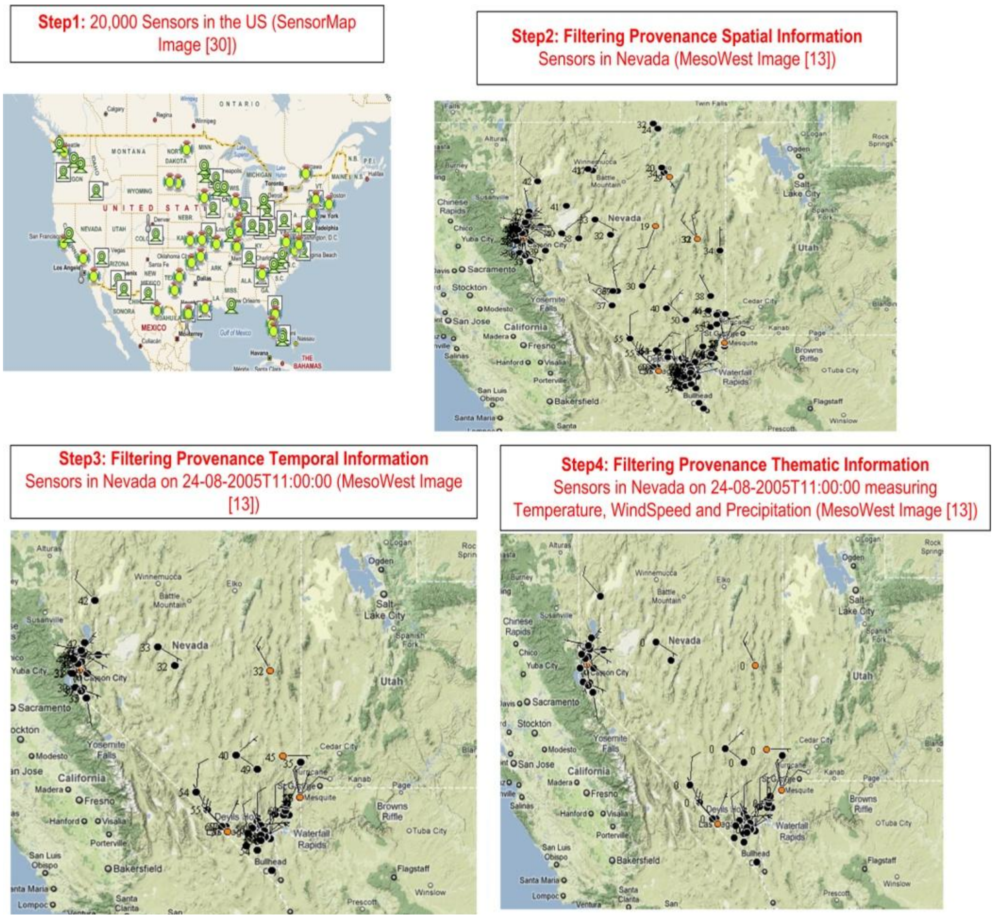

Fig.6. Answering a sensor-discovery query using spatio-temporal, and thematic provenance information

8. Related Work

Although this is the first attempt to develop an infrastructure for Sensor Provenance Management, there have been successful attempts to do the same in the domain of e-science. Within the sensors domain, provenance has been addressed from the storage point of view.

Provenance management within the eScience community has primarily been addressed in the context of workflow engines [25] while provenance management issues have been surveyed by Simmhan et al. [26]. The database community has also addressed the issue of provenance and defined various types of provenance, for example "why provenance" [27] and "where provenance" [27]. A detailed comparison of PMF (that underpins the *Sensor* PMS) with both workflow and database provenance is presented in [4].

The Semantic Provenance Capture in Data Ingest Systems (SPCDIS) [28] is an example of eScience project with dedicated infrastructure for provenance management. In contrast to the *Sensor* PMS, the SPCDIS project uses the proof markup language (PML) [29] to capture provenance information. The Inference Web toolkit [29] features a set of tools to generate, register and search proofs encoded in PML. Both *Sensor* PMS and the SPCDIS have common objectives but use different approaches to achieve them, specifically the *Sensor* PMS uses an ontology-driven approach with robust query infrastructure for provenance management.

In the Sensors community, Ledlie et al. [30] show how provenance addresses the naming and indexing issues related to sensor data storage. Park et al. [31] explore the need for data provenance in *Sensornet Republishing,* a process of transforming online sensor data and sharing the filtered, aggregated, or improved data with others.

9. Conclusion

This paper introduces an in-use ontology-driven provenance management infrastructure for Sensor data called *Sensor* PMS. We have developed a domain specific sensor provenance ontology by extending the provenir ontology. Due to this extension, SP ontology can interoperate with other domain-specific provenance ontologies to facilitate sharing and integration of provenance information from different domains and projects. We also show how provenance information can help answer complex queries within the sensors domain.

Acknowledgments. This work is funded in part by NIH RO1 Grant# 1R01HL087795-01A1 The Dayton Area Graduate Studies Institute (DAGSI), AFRL/DAGSI Research Topic SN08-8: "Architectures for Secure Semantic Sensor Networks for Multi-Layered Sensing.".

References

[1] W. C. Tan. Provenance in Databases: Past, Current, and Future. *IEEE Data Engineering Bulletin*, 30(4):3–12, Dec. 2007.

[2] S. B. Davidson, S. C. Boulakia, A. Eyal, B.Lud¨ascher, T. M. McPhillips, S. Bowers, M. K.Anand, and J. Freire. Provenance in Scientific Workflow Systems. *IEEE Data Engineering Bulletin*, 30(4):44–50, Dec. 2007.

[3] P. Buneman, S. Khanna, and W. C. Tan. Data Provenance: Some Basic Issues. In *Proceedings of the 20th Conference on Foundations of Software Technology and Theoretical Computer Science (FST TCS)*. Springer, Dec. 2000.
[4] S.S. Sahoo, R.S. Barga, J. Goldstein, A.P. Sheth, K. Thirunarayan, "Where did you come from...Where did you go?" An Algebra and RDF Query Engine for Provenance Kno.e.sis Center, Wright State University; 2009.
[5] SensorMap. http://atom.research.microsoft.com/sensewebv3/sensormap/, Retrieved March 22 2010
[6] Wikipedia Article on Ontology. http://en.wikipedia.org/wiki/Ontology_(information_science), Retrieved March 21 2010
[7] Sensor Data Ontology Model. http://knoesis.wright.edu/research/semsci/application_domain/sem_sensor/ont/sensor-observation.owl, Retrieved March 21 2010
[8] Semantic Web Wikipedia. http://en.wikipedia.org/wiki/Semantic_Web, Retrieved March 15 2010
[9] Wikipedia Article on RDF. http://en.wikipedia.org/wiki/Resource_Description_Framework, Retrieved March 19 2010
[10] Resource Description Framework. http://www.w3.org/RDF/ Retrieved March 15 2010
[11] SPARQL Protocol and Language: Frequently Asked Questions. http://www.thefigtrees.net/lee/sw/sparql-faq#what-is, Retrieved March 15 2010
[12] Linked Open Data Cloud, http://linkeddata.org/, Retrieved March 20 2010
[13] MesoWest. http://mesowest.utah.edu/index.html, Retrieved March 20 2010
[14] Observation and Measurements (O&M). http://www.opengeospatial.org/standards/om, Retrieved March 18 2010
[15] H. Patni, C. Henson, A. Sheth, 'Linked Sensor Data,' In: Proceedings of 2010 International Symposium on Collaborative Technologies and Systems (CTS 2010), Chicago, IL, May 17-21, 2010.
[16] Openlink Software. http://www.openlinksw.com/, Retrieved March 12 2010
[17] GeoNames. http://www.geonames.org/, Retrieved March 12 2010
[18] XSLT. http://www.w3.org/TR/xslt, Retrieved March 12 2010
[19] SSW Dataset. http://wiki.knoesis.org/index.php/SSW_Datasets, Retrieved March 12 2010
[20] S. S. Sahoo, D. B. Weatherly, R. Mutharaju, P. Anantharam, A. P. Sheth, R. L. Tarleton, "Ontology-Driven Provenance Management in eScience: An Application in Parasite Research." OTM Conferences (2) 2009: 992-1009
[21] Virtuoso SPARQL endpoint. http://harp.cs.wright.edu:8890/sparql, Retrieved March 14 2010
[22] B. Smith, W. Ceusters, B. Klagges, J. Kohler, A. Kumar, J. Lomax, et al., "Relations in biomedical ontologies." Genome Biol 2005;6(5):R46.
[23] Provenir Ontology. http://knoesis.wright.edu/library/ontologies/provenir/provenir.owl, Retrieved March 13 2010
[24] S. S. Sahoo, A. Sheth, "Provenir ontology: Towards a Framework for eScience Provenance Management", Microsoft eScience Workshop, USA, Oct 2009
[25] Provenance Challenge Wiki. http://twiki.ipaw.info/bin/view/Challenge/WebHome, Retrieved March 11 2010
[26] Y.L. Simmhan, B. Plale, D. Gannon, "A survey of data provenance in e-science" SIGMOD Rec. 2005;34(3):31 - 36
[27] P. Buneman, S. Khanna, W.C. Tan, "Why and Where: A Characterization of Data Provenance." In: 8th International Conference on Database Theory; 2001; 2001. p. 316 - 330
[28] SPCDIS. http://spcdis.hao.ucar.edu/, Retrieved March 11 2010
[29] Inference Web. http://iw.stanford.edu/2.0/, Retrieved March 11 2010
[30] J. Ledlie, C. Ng, D. A. Holland, K.-K. Muniswamy-Reddy, U. Braun, and M. Seltzer. "Provenance-aware sensor data storage." In *NetDB 2005*, April 2005.
[31] U. Park, J. Heidemann. "Provenance in Sensornet Republishing." In Proceedings of the 2nd International Provenance and Annotation Workshop , pp. 208-292. Salt Lake City, Utah, USA, Springer-Verlag. June, 2008