

Extending PROV Data Model for Provenance-Aware Sensor Web

Peng Yue^(✉), Xia Guo, Mingda Zhang, and Liangcun Jiang

State Key Laboratory of Information Engineering in Surveying,
Mapping and Remote Sensing, Wuhan University,
129 Luoyu Road, Wuhan 430079, China
geopyue@gmail.com

Abstract. Provenance has become a fundamental issue in Sensor Web, since it allows applications to answer “what”, “why”, “where”, “when”, and “how” queries related to the consumption process, which finally helps to determine the usability and reliability of data products. This paper proposes how the W3C PROV Data Model (PROV-DM) [1] can be used for creating a lineage model for Sensor Web to support interoperability.

In the sensor domain, the observation values, phenomenon measurement, observation location, time of observation, and observation procedure are important parts to answer sensor discovery queries. They are categorized as “what”, “why”, “where”, “when”, and “how” queries in this paper. These five perspectives, together, construct the lineage for an observation. In this paper, OGC standards and W3C SSN ontology [2] are leveraged to provide comprehensive domain vocabularies for Sensor Web, which are later mapped into concepts in PROV-DM for modeling observation lineage. It is possible to adopt SSN solely as the vocabularies and map them to W3C PROV. However, we argue that the solution of adopting O&M basic observation model [3] as a core with its complements by SSN ontology could facilitate the extraction of PROV data from the large amount of existing observation data following the OGC Sensor Web standards.

The domain vocabularies adopt the base Observation model in the O&M, enriched by classes and properties in the SSN ontology (Fig. 1). 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 (GFI_Feature) is “an abstraction of real world phenomenon”. Result (Result-Data) records the value of an observation. featureOfInterest and observedProperty together sketch the phenomenon to be measured. The “what” and “why” questions can then be answered using these entities. Furthermore, responsible party information is added using the class CI_ResponsibleParty from ISO19115:2003 to record the person or party responsible for the observation. An observation often involves a procedure, which can be a sensor, a human observer, or a series of process steps. Thus, a procedure (OM_Process) has two subconcepts: Sensor and Process. Observers (Observer) and sensing devices (SensingDevice) are modeled as subconcepts of Sensor. The Process can be understood as a processing method with Input and Output to describe how observations were made. These definitions together, answer the “how” question in

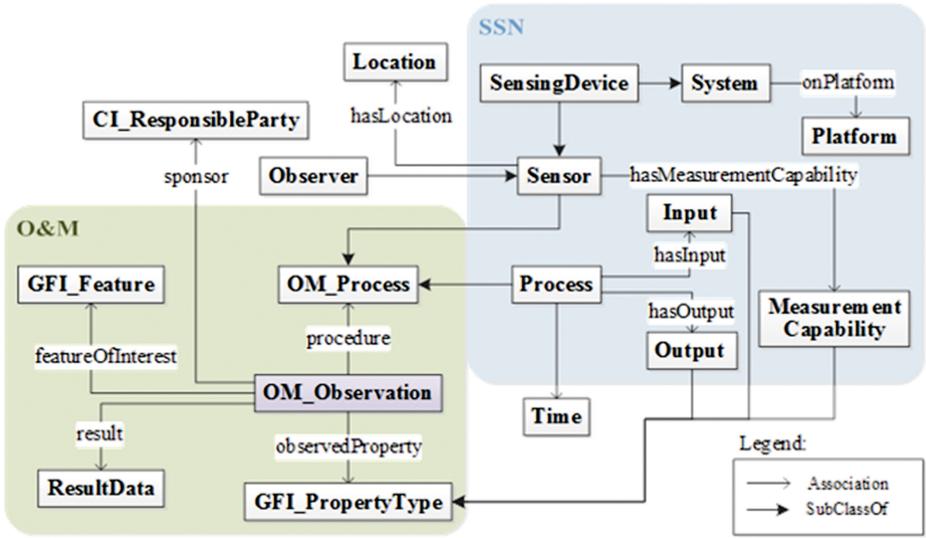


Fig. 1. Domain vocabularies for Sensor Web

provenance queries. Besides, the System and Platform are also added according to relationships among the sensor, system, and platform defined from the system perspective of the SSN ontology. Location and time, which are not described in the SSN ontology, answer “Where” and “When” questions, using associations with Sensor and Process respectively.

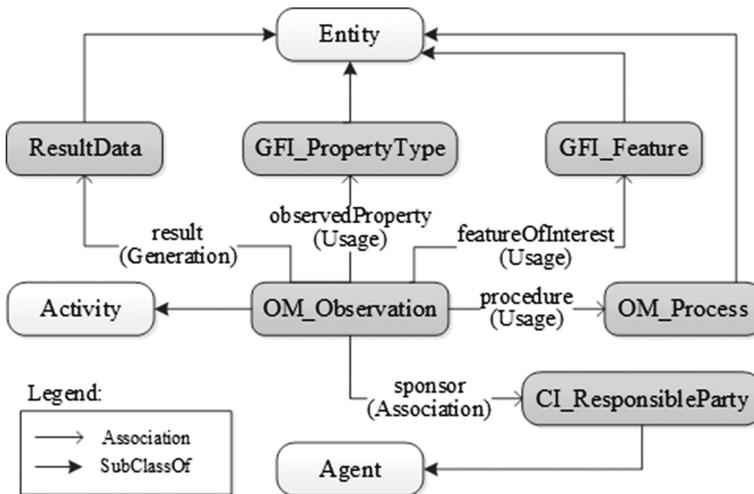


Fig. 2. Observation lineage model by extending PROV-DM

Mapping domain vocabularies from the Sensor Web to PROV-DM can facilitate interoperable provenance modeling. Figure 2 shows domain-specific extensions of W3C PROV by creating mapping from the domain vocabularies in Fig. 1 to PROV-DM.

GFI_PropertyType and GFI_Feature are entities that together can be used to describe the phenomenon observed by an observation. An observation (OM_Observation) is then a kind of Activity. The relationships observedProperty and featureOfInterest are categorized as the relation concept Usage. The procedure associates OM_Observation and OM_Process, and thus can be also seen as a kind of Usage. The relation result follows the relation concept Generation as the results are generated by observation activities. CI_ResponsibleParty is responsible for the observation, and then the relation sponsor is categorized as the relation concept Association. The subconcepts of OM_Process, such as Process, Sensor, and SensingDevices are all entities.

In the paper, an ontology is used to represent the observation lineage model which are extended to PROV-O, using rdfs:subClassOf and rdfs:subPropertyOf relationships. The observation lineage model is then enriched according to the specific sensors and database schema when dealing with the PM cases. A relational database is used to store datasets including observation values and corresponding metadata. The datasets content is then mapped into RDF, which is published as linked data and thereby allows to be browsed and searched.

Three typical query examples, including temporal and spatial provenance filtering, are conducted and showed through a prototype system. One of the query examples following the SPARQL syntax is showed in Table 1. Specially, to demonstrate the potential of interoperability, a query example is also conducted with combining generic concepts in PROV-O and domain-specific concepts in observation lineage ontology. Using OWL reasoners, concepts and relations that are mapped to PROV can be discovered by applying PROV-related queries.

Table 1. The query example

```

SELECT ?sensor ?geonames ?latitude ?longitude
WHERE {
  ?Observation prov-pmsw:procedure ?sensor.
  ?sensor prov-pmsw:type "sensingdevice".
  ?Observation prov-pmsw:resultTime ?instanceTime.
  ?instanceTime prov-pmsw:xsdDateTime ?xsdDateTime.
  FILTER(?xsdDateTime >= "2013-06-16T08:00:00+08" && ?xsdDateTime <= "2013-
06-16T08:00:00+08").
  ?sensor prov-pmsw:hasLocation ?location.
  ?location prov-pmsw:geonames ?geonames.
  ?location prov-pmsw:wgs84_lat ?latitude.
  ?location prov-pmsw:wgs84_long ?longitude.
  FILTER(?geonames="Nanjing").
}
    
```

Acknowledgements. We are grateful to the anonymous reviewers for their constructive comments. The work is supported jointly by National Basic Research Program of China (2011CB707105), National Natural Science Foundation of China (41271397), and Program for New Century Excellent Talents in University (NCET-13-0435).

References

1. Moreau, L., Missier, P.: ROV-DM: The PROV Data Model. WWW document (2013). <http://www.w3.org/TR/prov-dm/>
2. W3C 2005 Semantic Sensor Network Ontology. WWW document. <http://www.w3.org/2005/Incubator/ssn/ssnx/ssn>
3. Cox S 2007 Observation and Measurements Implementation Specification, Version 1.0. Open Geospatial Consortium document No. OGC 07-022r1