

Building a Seismology Workflow-Driven Ontology: a Case Study

Leonardo Salayandia¹ and Aaron Velasco²

leonardo@utep.edu, velasco@geo.utep.edu

¹Computer Science Department, ²Geological Sciences Department
University of Texas at El Paso

Technical Report: UTEP-CS-07-06

1. Introduction

The goal of cyberinfrastructure efforts such as Geosciences Network (GEON) is to allow earth scientists to discover, access, integrate and disseminate knowledge in distributed environments such as the Web, changing the way in which research is conducted. To support this effort, the earth sciences community has begun the complex task of capturing knowledge from specific disciplines through ontologies. Ontologies are a computer-science solution to the challenge of representing knowledge in a way that a computer can interpret and reason about. Ontologies in general, however, are difficult to develop because knowledge can be captured at different levels of abstraction, and knowledge can be interpreted through different perspectives than those intended by the ontology developer, ultimately leading to ontologies that are difficult if not impossible to leverage for the purposes of achieving the stated goals of cyberinfrastructure. The Workflow-Driven Ontology (WDO) approach is an approach that allows domain scientists (not computer scientists) to create ontologies with special properties that makes the ontologies more amenable to the goals of cyberinfrastructure. The WDO approach is useful not only to guide the scientist through the complex task of capturing knowledge from a specific discipline in an ontology, but the WDO approach also is useful for creating workflow specifications that serve as use cases for cyberinfrastructure.

A case-study was conducted to apply the WDO approach to the field of seismology. As a starting point, Dr. Velasco, the domain expert, defined concepts related to the field of seismology to create an initial WDO. Subsequently, the initial seismology WDO and a general seismology ontology produced by the 2004 Seismology Ontology Workshop held at Scripps Institution of Oceanography in San Diego, California, were compared. The comparison yielded that the general ontology developed by the workshop was more comprehensive and encompassed a wider terminology of the seismology field in general. In contrast, the seismology WDO had a narrower seismic application that described a specific task in the seismology field, i.e., to develop discriminants in the local to near-regional transition region. The specific application described by the seismology WDO facilitated the refinement of the hierarchical classifications defined in the general seismology ontology. The result of merging the seismology WDO and general seismology ontology resulted in an ontology that is more task-oriented and amenable to cyberinfrastructure efforts.

The rest of this document is organized as follows; Section 2 briefly presents the WDO approach; Section 3 presents the initial seismology WDO produced; Section 4 presents

the seismology ontology developed at the 2004 seismology ontology workshop held at Scripps Institution of Oceanography in San Diego, California; Section 5 presents a comparison of the seismology WDO, and the seismology ontology, and shows the result of merging both ontologies; Finally, Section 6 presents conclusions.

2. The WDO approach

The WDO approach is useful not only to guide the scientist through the complex task of capturing knowledge from a specific discipline in an ontology, but the WDO approach also is useful for creating workflow specifications that serve as use cases for cyberinfrastructure. Figure 1 shows the WDO upper-level ontology. This ontology is normally imported by an ontology that is developed using the WDO approach, where the upper-level WDO serves as reference for concepts (e.g., Raw Data) that are used to classify new domain knowledge in the developing ontology. The OWL document for the upper-level WDO can be found at the following URL: <http://trust.utep.edu/2006/12/wdo>.

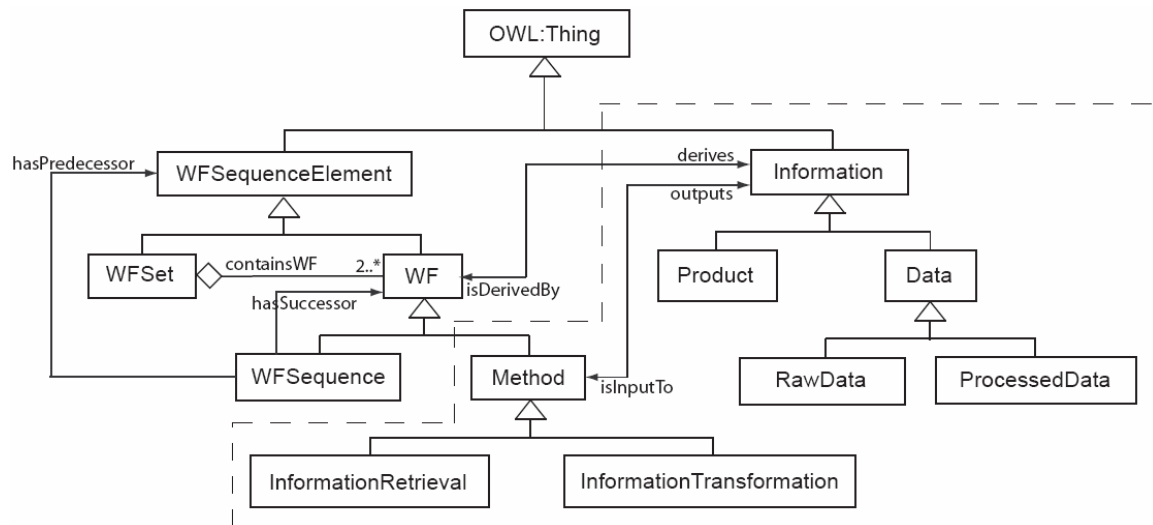


Figure 1. Upper-Level Workflow Driven Ontology

The concepts inside the dash-line box are pertinent to the capture of domain knowledge, while the concepts outside the dash-line box are used to capture workflow specifications derived from the captured domain knowledge. A more detailed discussion about the WDO approach can be found in [1], and the work presented in [2] discusses the methodology to extract Model-Based Workflow specifications from a WDO.

3. A Seismology WDO

As an initial step in this case study, the WDO approach was used to create an initial Seismology WDO that reflects the work that Dr. Aaron Velasco is proposing for the EarthScope initiative. This work is intending to data-mine the EarthScope platform to develop discriminants in the local to near-regional transition region. The development of the Seismology WDO included the following steps: 1) Leonardo Salayandia, a computer scientist with limited knowledge of the Seismology domain, parsed documentation that

described the proposed work of interest, and identified nouns from this documentation as possible concepts to use in the WDO; 2) Leonardo, with the aid of Aaron Velasco, the Seismology domain expert, filtered the initially identified concepts, and organized the concepts into the classification provided by the upper-level WDO ontology; 3) Aaron and Leonardo determined *isInputTo* and *outputs* relationships between the identified concepts in order to complement the Seismology WDO.

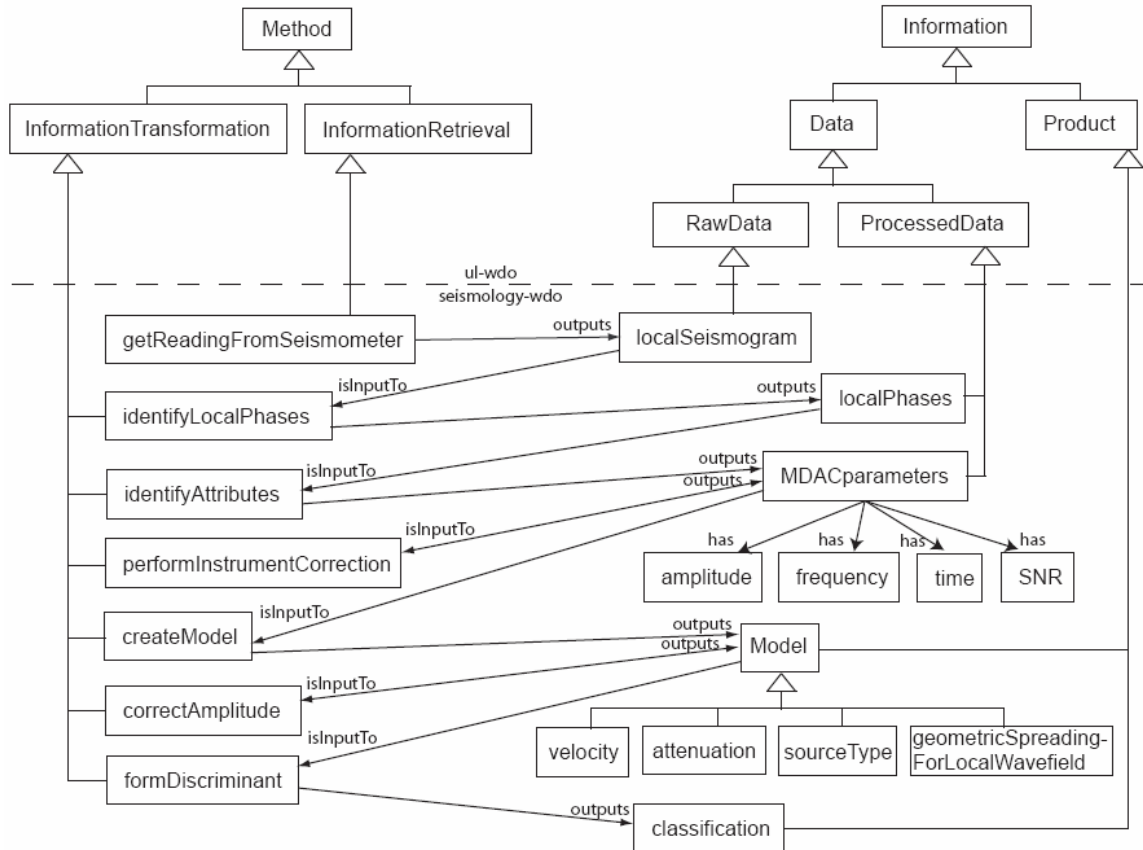


Figure 2. Seismology WDO

Figure 2 shows the diagram for the Seismology WDO. The concepts that appear above the dashed line are part of the upper-level WDO ontology, and are only showed in the diagram as reference. The Seismology WDO reflects the upper-level WDO reference through an import statement.

Although the WDO approach is intended to produce ontologies that are task-oriented and amenable to cybeinfrastructure goals, it is expected that the knowledge captured using the WDO approach will require several refinement iterations before executable workflows can be derived from the WDO. For example, the seismology WDO illustrated in Figure 2 represents an initial iteration of the WDO approach. Consider the concepts *Model* and *formDiscriminant*, and the *isInputTo* relationship between them. It is expected that there exist multiple techniques to form discriminants, and that the technique to use in a particular instance will depend on the type of model to be used. As a result, after

determining the resources available (i.e., the type of model to be used, and the more specific technique to use to form the discriminant), a second iteration of the seismology WDO would result in the refinement of the *formDiscriminant* concept with more specific child concepts that identify the particular techniques to form discriminants. In addition, the *isInputTo* relationship between the *Model* and *formDiscriminant* concepts would be also refined in the WDO to specify a relationship between the specific type of model and the specific technique that is used for that model type.

4. A Seismology Ontology

The ontology presented in this section is the result of a seismology ontology workshop held at Scripps Institution of Oceanography in San Diego, California in February 2004. The participants of the workshop included experts in the areas of seismology and information technology¹. Figure 3 shows a portion of the diagram of this ontology.

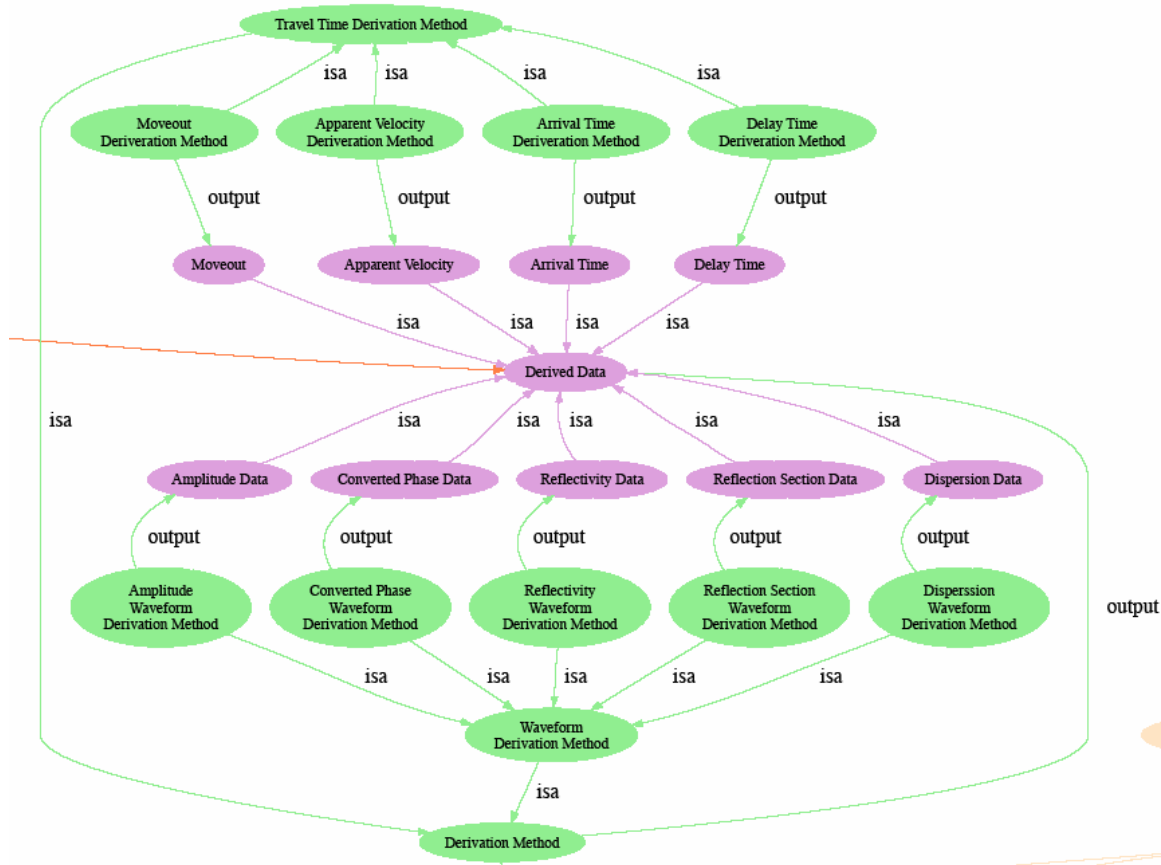


Figure 3. Part of the Seismology Ontology

¹ Randy Keller and Ann Gates, University of Texas at El Paso; Bertram Ludaescher, Dogan Seber, Chaitan Baru, and Kai Lin, San Diego Supercomputer Center; Gabi Laske and Frank Vernon, Scripps Institute, University of California at San Diego; Tim Ahern, IRIS; Colin Zelt, Rice; Matt Fouch, Arizona State; John Hole, Virginia Tech; David James, Carnegie Institute of Washington; Bill Pike, Penn State

The seismology ontology contains more than 70 concepts and is predominantly classificatory in nature, containing mostly relationships of the type *is_a* and *has*. The seismology ontology contains also relationships of the type *output*, *generates*, and *apply to* that are more in accordance to the relationships that need to be established between cyberinfrastructure resources in order to compose workflow applications; this relationships however, appear at a very abstract level in the hierarchy, and may be hard to establish a mapping between the ontology concepts and actual cyberinfrastructure workflow applications. For example, consider the relationship (*Derivation Method* → *outputs* → *Derived Data*) that appears in the portion of the diagram of the Seismology Ontology that is shown in Figure 3. As discussed in the previous section, these types of relationships may be correct at a general level of abstraction, but practical usage of cyberinfrastructure through the development of workflows require these types of relationships to be more specific, i.e., the specification of what type of *Derivation Method* *outputs* what type of *Derived Data* is needed before a workflow implementation over can be devised. The WDO approach provides domain experts with a clear understanding of purpose for the creation of the ontology, i.e., the development of executable workflows, and as a result, the WDO approach provides ‘next steps’ to follow towards the refinement of the knowledge in order to achieve the intended purpose. In contrast, the development of ontologies without the WDO approach produces a general-purpose ontology, and as a result, it is more flexible, and harder to refine while maintaining it at the general purpose level.

5. Seismology WDO vs. Seismology Ontology

The comparison between the Seismology Ontology and the Seismology WDO yielded that the general ontology developed by the workshop was more comprehensive and encompassed a wider terminology of the seismology field in general. In contrast, the Seismology WDO had a narrower application that described a specific task in the seismology field, i.e., to develop discriminants in the local to near-regional transition region. The specific application described by the seismology WDO facilitated the detection of inconsistencies in the hierarchical classification of concepts defined in the general seismology ontology. After correcting such inconsistencies, the general seismology ontology was refined to include the seismology WDO created in the case-study, hence resulting in a refined ontology that is more task-oriented and amenable to cyberinfrastructure efforts. Figure 4 shows the full diagram of this combination ontology.

6. Conclusions

In the field of seismology, the data gathered in a seismogram is used to either analyze the structure of the earth, or to analyze the propagation of an event, i.e., an earthquake or an explosion. If the seismogram is used to analyze the structure of the earth, then assumptions are made about the characteristics of an event, and visa versa. Correspondingly, domain knowledge in the field of seismology can be gathered in a way that is more amenable to the study of the earth structure, or more amenable to the study of events. The general seismology ontology was developed with the first objective in mind, while the seismology WDO was developed with the latter objective. In spite of the intended use of the ontology produced, the WDO approach allows the refinement of domain knowledge towards a specific task. The resulting combination of the general seismology ontology and the seismology WDO represents a refinement of domain knowledge that was initially gathered with the objective describing the seismology field for the purposes of studying the structure of the earth, and has made the ontology more task oriented towards the different objective of studying the propagations of events, i.e., to develop discriminants in the local to near-regional transition region.

References

- [1] Leonardo Salayandia, Paulo Pinheiro da Silva, Ann Q. Gates, and Flor Salcedo. "Workflow-Driven Ontologies: An Earth Sciences Case Study," *In Proceedings of the 2nd Intl Conference on e-Science and Grid Computing*, Amsterdam, Netherlands, December 2006
- [2] Leonardo Salayandia, Paulo Pinheiro da Silva, Ann Q. Gates, and Alvaro Rebellon. "A Model-Based Workflow Approach for Scientific Applications," *In Proceedings of the 6th OOPSLA Workshop on Domain-Specific Modeling*, Portland, Oregon, October 2006