Argos: An Ontology and Web Service Composition Infrastructure for Goods Movement Analysis

José Luis Ambite[1], Genevieve Giuliano[2], Peter Gordon[2], Stefan Decker[1,3], Andreas Harth[1], Karanbir Jassar[1], Qisheng Pan[2,4] and LanLan Wang[2]

[1] Information Sciences Institute [2] School of Policy, Planning, and Development
University of Southern California
4676 Admiralty Way, Marina del Rey, CA 90292, USA

Many scientific problems can be modeled as computational workflows that integrate data from heterogeneous sources and process such data to derive new results. These data analysis problems are pervasive in the physical and social sciences, as well as in government practice. Therefore, techniques that facilitate the creation of such computational workflows are of critical importance. In our work on the Argos project we are developing techniques to automatically create computational workflows in response to user data requests. As a unifying paradigm, we represent both data access and data processing operations as web services.

We focus on a domain representative of many economic analysis problems: good movement analysis in a large metropolitan area. This domain is of great interest to both social scientists and government practitioners, since understanding the patterns of freight flow is crucial for urban planning and forecasting the economic development of a region. Previous approaches to this type of analysis, mainly surveys, are prohibitively expensive. In Argos we propose to integrate data from secondary sources to estimate the flow of commodities in the Los Angeles metropolitan area, following Gordon and Pan [2001]. The Argos approach presents several advantages. First, it is significantly more cost-effective since the problem is now reduced to automated data integration and processing, instead of labor-intensive surveys. Second, many more questions can be posed and answered, since different data sources and data operations can be composed to derive a variety of novel data for the given domain. Finally, our system produces results based on the latest data, which is obtained live from the data sources whenever possible.

In this system demo, we will showcase our initial design for automatic composition of web services for goods movement analysis. Our approach consists of three steps:

1. Define an ontology of the domain to model data processed throughout a workflow.
2. Describe data sources and operations using this ontology.
3. Automatically compose a workflow in response to a user request based on mediator techniques.

First, we have developed an ontology of the goods movement domain to describe the data provided by the sources and the data participating as input and output of the processing operations (see Figure 1). We consider each data item as a measurement that has values along a set of dimensions, such as geographic area (e.g., Long Beach), type of flow (e.g., imports), type of product (e.g., vehicles and parts), time interval (e.g., January 2001), value and unit (e.g., 1108 thousand short tons). Some of the values of the dimensions have a hierarchical structure, either *type* or *part-of* hierarchies. For example, California is a geographical area that is part of the US. The part-of hierarchies are used for value aggregation. For example, we could approximate a measurement for all of the US if we have data for its states and territories. To facilitate the knowledge acquisition task we used the Protégé ontology editor [Noy et al., 2001] (see Figure 1). As a common syntax for the ontology and the data we adopt the Resource Description Framework (RDF) [Lassila & Swick, 1999]. RDF represents labeled graphs (i.e., semantic networks). RDF also includes edge labels with predefined class/subclass semantics. We considered more

[3] Also with the Digital Enterprise Research Institute (DERI), Ireland.
[4] Currently at Texas Southern University.

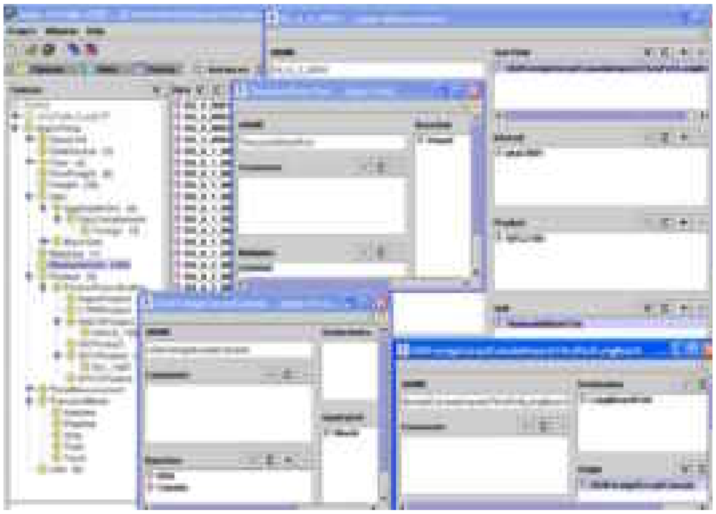

Figure 1. Ontology for Goods Movement Analysis

expressive languages, such as the description logic OWL [Dean & Schreiber, 2003], but we prefer RDF for its simplicity, flexibility, and the fact that much of our reasoning involves aggregation in part-of hierarchies for which OWL does not provide support. So far, the representational power of RDF has been sufficient for the object-oriented model of our domain.

Second, we have modeled the data sources and operations uniformly as web services that produce or transform measurements. To describe the semantics of these web services, we use Triple [Sintek & Decker, 2002], which is a logic programming language that adapts F-logic [Kifer et al., 1995] to process RDF data. We used Triple logic rules built with terms from the ontology to accurately specify the inputs and outputs of sources and operations.

Third, we are developing an approach to web service composition based on mediator techniques [Thakkar et al., 2003]. Given a user data request, our system uses the Triple service descriptions (sources and operations) to generate a logic program that answers the request. Our work differs from [Thakkar et al., 2003] in that we use a richer representation language with object-oriented features as opposed to their flat relational model. Moreover, we model more complex transformations, not just relational algebra operations. In fact, the operations can be arbitrary algorithms or Triple programs.

For the system demonstration, we will show our goods movement ontology, the description of a subset of the sources and operations in this domain using the ontology, the implementation of these sources and operations as web services that exchange RDF data, examples of our design for automatic web service composition in response to user requests, and the execution of the composed web services based on a BPEL4WS [2003] workflow.

References

BPEL4WS (2003), Business Process Execution Language for Web Services Version 1.1, 05 May 2003,

Dean, M., Schreiber, G., Eds., (2003). OWL Web Ontology Language Reference *W3C Proposed Recommendation*, 15 December 2003. http://www.w3.org/TR/owl-ref/.

Gordon, P, Pan, Q. (2001) "Assembling and Processing Freight Shipment Data: Developing a GIS-Based Origin-Destination Matrix for Southern California Freight Flows." Final Report, Project METRANS National Center for Metropolitan Transportation Research. (www.metrans.org).

Kifer, M., Lausen, G, Wu, J. (1995). Logical foundations of object-oriented and frame-based languages. Journal pf the ACM 42, 741-843. July 1995.

Lassila, O, Swick, R. (Eds) (1999). Resource Description Framework (RDF) Model and Syntax Specification, *W3C Recommendation* 22 February. 1999. http://www.w3.org/TR/RECrdf-syntax/

Noy, N.F., Sintek, M., Decker, S., Crubézy, M., Fergerson, R., Musen, M.A (2001). Creating Semantic Web Contents with Protégé-2000. In *IEEE Intelligent Systems*, 16 (2), 60-71. Special Issue on the Semantic Web. March/April 2001.

Sintek, M., Decker, S (2002). TRIPLE - A Query, Inference, and Transformation Language for the Semantic Web. *International Semantic Web Conference (ISWC*), Sardinia, Italy, June 2002.

Thakkar, S, Knoblock, C. A., Ambite, J. L. (2003). A View Integration Approach to Dynamic Composition of Web Services *Workshop on Planning for Web Services*, The 13th International Conference on Automated Planning & Scheduling (ICAPS 2003), Trento, Italy, 2003.