

An Ontology-Driven Model for the Efficient Use of Provenance Information

Fred Fonseca

*The Pennsylvania State University
School of Information Sciences and Technology*

1. Introduction

In this research we combine the gathering of provenance information with upcoming trends in research on semantic models, namely ontologies. We will analyze all measured provenance information to research and develop an optimal ontology-driven model for the most efficient utilization of the gathered provenance information.

The proposed research builds on two main lines of work. First, it leverages NCSA current efforts on information gathering about decision processes using geospatial electronic records and medical images (Lee and Bajcsy, 2005; Bajcsy and Clutter, 2006). Second, it adds on my previous research on ontologies of geospatial phenomena and imagery (Fonseca *et al.*, 2002b; Fonseca *et al.*, 2002c; Fonseca and Sheth, 2002; Fonseca *et al.*, 2003; Fonseca and Martin, 2004; Fonseca and Martin, 2005). NCSA provides a unique opportunity for doing this research because of its high performance computing capabilities to deal with massive amounts of data and CPU intensive processes, as well as the availability of the software to gather the provenance information. This proposed research cannot reach the optimal ontology-based model without having all NCSA resources. Therefore, the use of NCSA computer infrastructure and the joint work with NCSA's research staff will lead to enhanced results both for NCSA's research and for the my own research on the use of ontologies for integration of geo-spatial information.

This follows up on Bajcsy and Clutter's work, "Information Gathering about Decision Processes Using Geospatial Electronic Records". There, they claim that a "future direction for this research could be to embed provenance into a workflow engine [...] The provenance data are represented currently as triples with (subject, predicate and object). More elaborate taxonomy and ontology representations, as well mechanisms for interfacing the provenance meta-data would be considered in the future". In this research we will focus on multiple types of data and process provenance information collected by previous NCSA research using computer-centric and human-centric information gathering mechanisms. We will create a hierarchical organization of provenance information in the form of ontologies.

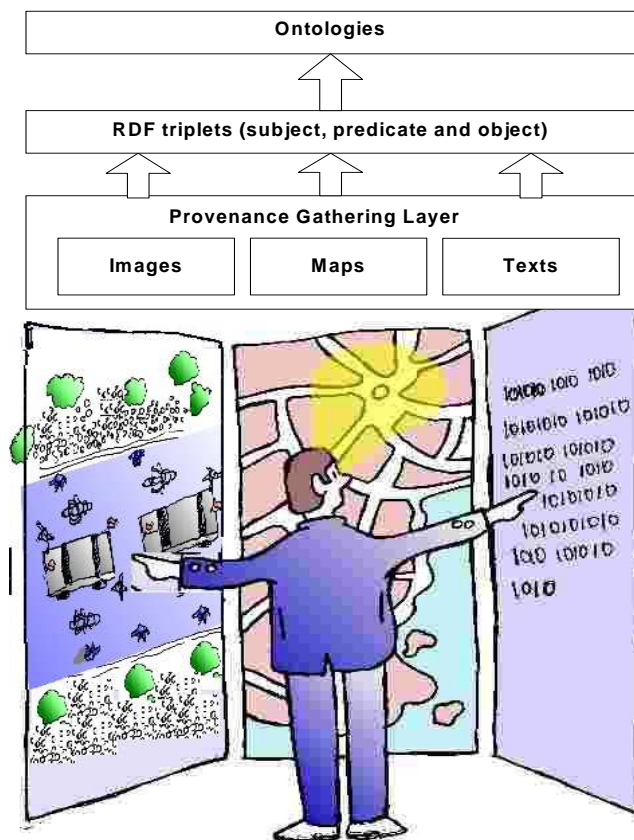


Figure 1- Gathering provenance information during the decision making process

2. References

- Bajcsy, P. and Clutter, D. (2006) Information Gathering About Decision Processes Using Geospatial Electronic Records. NCSA Cyber-environments Division, Urbana-Champaign, IL, Technical Report ISDA06-001.
- Fonseca, F., Davis, C. and Camara, G. (2003) Bridging Ontologies and Conceptual Schemas in Geographic Information Integration. *Geoinformatica* 7(4): pp. 355-378.
- Fonseca, F. and Egenhofer, M. (1999) Ontology-Driven Geographic Information Systems. in: Medeiros, C. B., (Ed.), 7th ACM Symposium on Advances in Geographic Information Systems, Kansas City, MO, pp. 14-19.
- Fonseca, F., Egenhofer, M., Agouris, P., and Câmara, G. (2002a) Using Ontologies for Integrated Geographic Information Systems. *Transactions in GIS* 6(3): pp. 231-257.
- Fonseca, F., Egenhofer, M., Davis, C., and Borges, K. (2000) Ontologies and Knowledge Sharing in Urban Gis. *Computer, Environment and Urban Systems* 24(3): pp. 232-251.
- Fonseca, F., Egenhofer, M., Davis, C., and Câmara, G. (2002b) Semantic Granularity in Ontology-Driven Geographic Information Systems. *AMAI Annals of Mathematics and Artificial Intelligence - Special Issue on Spatial and Temporal Granularity* 36(1-2): pp. 121-151.
- Fonseca, F. and Martin, J. (2004) Space and Time in Eco-Ontologies. *AI Communications - The European Journal on Artificial Intelligence* 17(4): pp. 259-269.
- Fonseca, F. and Martin, J. (2005) Toward an Alternative Notion of Information Systems Ontologies: Information Engineering as a Hermeneutic Enterprise. *Journal of the American Society for Information Science and Technology* 56(1): pp. 46-57.
- Fonseca, F., Martin, J. and Rodríguez, A. (2002c) From Geo to Eco-Ontologies. in: Egenhofer, M. and Mark, D., (Eds.), *Geographic Information Science-Second International Conference GIScience 2002*, Boulder, CO, pp. 93-107.
- Fonseca, F. and Sheth, A. (2002) *The Geospatial Semantic Web*. UCGIS - University Consortium for Geographic Information Science, Technical Report.
- Lee, Y.-J. and Bajcsy, P. (2005) An Information Gathering System for Medical Image Inspection. in: Ratib, O. M. and Horii, S. C., (Eds.), *SPIE Medical Imaging 2005: PACS and Imaging Informatics*, San Diego, CA, pp. 374-381.