LA-UR- 09-04203

*Approved for public release; distribution is unlimited.*

*Title:* Towards an Ontology Framework Supporting the Integration of Geographic Information with Modeling and Simulation for Critical Infrastructure Protection


*Author(s):* John Ambrosiano, Russell Bent, Steve Linger



*Intended for:* 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems


Los Alamos NATIONAL LABORATORY EST.1943

## Towards an Ontology Framework Supporting the Integration of Geographic Information with Modeling and Simulation for Critical Infrastructure Protection

John Ambrosiano (ambro@lanl.gov), Russell Bent (rbent@lanl.gov), Steve Linger (spl@lanl.gov), Los Alamos National Laboratory, P.O. Box 1663, MS K488, Los Alamos, NM 87545

### Abstract

Protecting the nation's infrastructure from natural disasters, inadvertent failures, or intentional attacks is a major national security concern. Gauging the fragility of infrastructure assets, and understanding how interdependencies across critical infrastructures affect their behavior, is essential to predicting and mitigating cascading failures, as well as to planning for response and recovery. Modeling and simulation (M&S) is an indispensable part of characterizing this complex system of systems and anticipating its response to disruptions. Bringing together the necessary components to perform such analyses produces a wide-ranging and coarse-grained computational workflow that must be integrated with other analysis workflow elements. There are many points in both types of workflows in which geographic information (GI) services are required. The GIS community recognizes the essential contribution of GI in this problem domain as evidenced by past OGC initiatives. Typically such initiatives focus on the broader aspects of GI analysis workflows, leaving concepts crucial to integrating simulations within analysis workflows to that community. Our experience with large-scale modeling of interdependent critical infrastructures, and our recent participation in a DHS initiative concerning interoperability for this M&S domain, has led to high-level ontological concepts that we have begun to assemble into an architecture that spans both computational and "world" views of the problem, and further recognizes the special requirements of simulations that go beyond common workflow ontologies. In this paper we present these ideas, and offer a high-level ontological framework that includes key geospatial concepts as special cases of a broader view.

### 1. Motivation and Background

Protecting the nation's infrastructure from natural disasters, inadvertent failures, or intentional attacks is a major national security concern. Gauging the fragility of infrastructure assets, and understanding how interdependencies across critical infrastructures affect their behavior, is essential to predicting and mitigating cascading failures, as well as to planning for response and recovery.

Modeling and simulation (M&S) is an indispensable part of characterizing this complex system of systems, and anticipating its response to disruptions. Bringing together the necessary components to perform such analyses produces a wide-ranging and coarse-grained computational workflow that must be integrated with other analysis workflow elements. There are many points in both types of workflows in which geographic information system (GIS) services are required.

The GIS community has recognized their essential role in this problem domain as evidenced by past OGC initiatives. For example, interoperability for Critical Infrastructure Protection (CIP) has been discussed in an OGC draft interoperability program report [1]. Similarly, the role of GIS in providing decision support for disaster mitigation and response planning, including the need to integrate simulation model results, has also been addressed [2].

Typically such discussions focus on the broader aspects of integrating geographic data with decision analysis workflows, leaving concepts crucial to managing simulation workflows to the simulation community. However, we are finding ample evidence that appreciation is growing for the need to more tightly integrate these two types of workflows, with major points of contact occurring in the requirements for interoperable GIS services. A good example is the GALEON OGC interoperability experiment in collaboration with NCAR [3]. Another is the ongoing work at Los Alamos to develop a service-oriented architecture for integrating scientific simulations for critical infrastructure interdependency analysis [4].

Recently, Los Alamos National Laboratory participated with Sandia National Laboratories in a DHS interoperability initiative for CIP known as the Complex Event Modeling and Simulation Analyses project (CEMSA). This project was motivated by the need to handle the real possibility of overlapping disruptions that might result from multiple natural disasters, terrorist exploitation of a natural disaster, or simultaneous terrorist attacks against various infrastructures at multiple locations. The inherent complexity of such scenarios, which the project sought to address, inevitably drives corresponding demands for interoperability among diverse simulation and decision analysis tools and their related workflows [5]. Our work in this area has given us a perspective on analysis workflows that rely heavily on simulation, and on the role of GIS services for those whose problem domain spans some portion of the globe.

## 2. Conceptual Framework

As a strategy for developing interoperability semantics for problems that combine simulations and decision analysis workflows, we assert that the problem can be decomposed into two related knowledge domains (see Figure 1): the *world* domain and the *computational* domain. These two domains arise naturally when entities in the real world are abstracted into concepts and relationships for the purpose of analysis and problem-solving, and then abstracted again to build computational automata (simulations and other computational tools). The process of computational modeling and analysis takes us from the real world, to our conceptualizations of it, to the computerization of those concepts. What emerges from the computer is then subjected to scientific scrutiny, interpretation, and inference. This takes us back through our conceptualizations to the real world again.

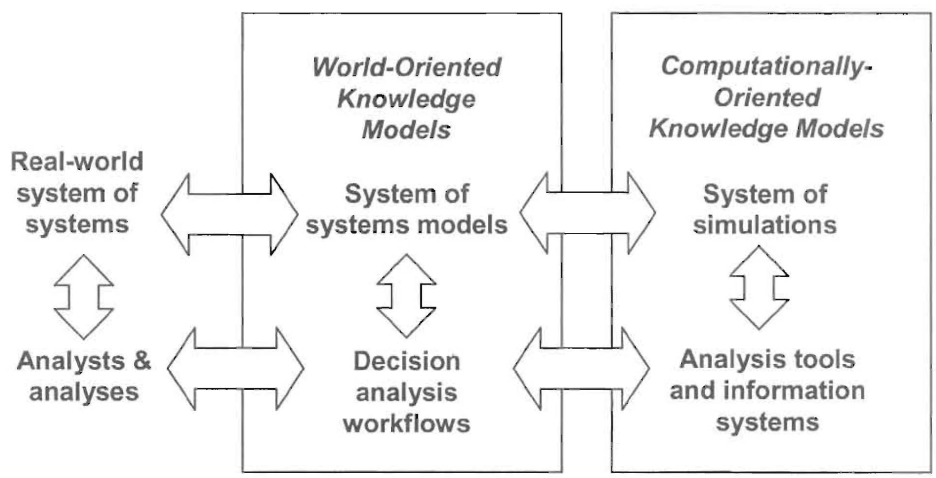


*Figure 1. A "high-concept" for interoperability based on "world-oriented" and "computationally-oriented" semantics for concepts related to simulation and decision analysis workflows respectively. Relationships linking the two domains are many-to-many in general.*

Mappings between world systems and computational systems are many-to-many in general. For example, a disaster planning analysis of dam breach and subsequent flooding involves several concepts in the real world such as the construction and geometry of the dam, the dimensions of the breach and the flow of water from it, leading to inundation of the downstream terrain. Each of these aspects of the problem has many potential representations in the computational world. Similarly, the shallow-water equations and their implementation in computational form, can be applied to many problems, not just the flooding problem as described. This is why both views are necessary.

Analysis workflows and the real-world systems of particular interest to analysts that employ simulations, dynamical systems, share a common metaphor. They are systems in which many of the key semantic relations are based on processes that occur over time. However, the interpretation of process semantics is different in each of the two.

In the case of *workflows*, we tend to speak of process terms of a *task* to be performed, the *actors* that perform it, its *artifacts*, some required to perform it and others produced by it, and the *resources* it consumes. Sometimes the task is performed by a computational actor, and in the interesting cases, by a simulation. The workflow aims overall to accomplish some goal, given the actors, artifacts and resources at its disposal, and does this by executing its series of tasks, each of which accomplishes a relevant sub-goal. In this regard, a workflow is a *plan* in the sense of artificial intelligence [6]. As such, AI planning algorithms have been proposed as a means of automatically generating workflows for enterprise systems [7].

For real-world dynamical systems, the metaphors employed in describing dynamical systems naturally apply. Some of the most widely used concepts come from control theory [8]. A system is said to have *inputs* and *outputs* with the latter related to the former by a *response*. Dynamical control systems may be described either as *continuous* or *discrete*. Simulations are computer surrogates for dynamical systems in the real world and as such are necessarily discrete. Well known, general frameworks for simulation have been developed on this idea [9].

As a last, key abstraction in the proposed semantic architecture, we introduce the idea of a *field*. For many applications of computing in science and engineering, including simulations, a principle artifact in the workflow is the representation of a function that associates a property or quantity with a point in some domain or base space. GIS services routinely define an entity called a *coverage*, a GIS, a spatial dataset containing a uniform set of geographic features such as arcs, nodes, polygons and points, and associates with them a table of attributes. A coverage usually represents a *theme* such as soils, streams, roads or land use. Reflecting their importance, the OGC has produced a web coverage standard [10].

In the broader perspective of scientific and engineering computing, this is sometimes called a field. This must not to be confused with the term "field" as it is applied when labeling an attribute in a database table. Rather, the connotation is like that in physics where one speaks of an "electromagnetic field" such as that of the Earth. While the idea seems simple enough, representing a "field quantity" can become quite challenging when, for example, it is vector-valued, or mapped onto some complicated topological structure or coordinate manifold, or when values are non-uniformly represented across the domain. To handle these sorts of challenges, researchers have applied a number of sophisticated mathematical structures including fiber bundles and sheaves [11, 12]. The domains for such structure have been generalized from a discrete topology of isolated points or simple grids to cell complexes built up by gluing a collection of cells together to make topological spaces, including those that support measures of distance [13].

### 3. Semantic Architecture

The concepts discussed in the previous section can be used to develop an architecture for the semantics of decision analysis workflows that rely on simulations, and that exploit complex data sets. In this architecture, GIS data is a special case of a broader class of data sets that span many scientific and engineering applications. Figure 2 is a sketch of a semantic architecture in which the most coherent sets concepts are gathered in to sub-ontologies. The figure, which only shows groups of related concepts, is not intended to offer a formal ontology, which is beyond the scope of this paper, but only to convey the essentials of the approach. Also, it is not presented as a layered architecture, because the dependency relationships among the concepts are not strictly ordered. However, the figure has been arranged so that more basic concepts are near the bottom. There are a number of important relationships connecting these component ontologies that cannot be easily represented in the diagram and that we will now describe.

First consider (2a) which covers the essential semantics of the workflow. The workflow is made up of tasks performed by actors who use or produce artifacts. These tasks are purposeful and informed by a plan that seeks to achieve a particular goal. Usually the goal is to obtain new information by transforming the information given through the process of analysis and in particular through simulation. The plan (2b) is composed of actions that are intended to achieve the goal, and which transform the conditions that enable them into new states. These are their effects. Actions are equivalent to tasks in the overall workflow. They are also equivalent transitions in the basic workflow semantics (2f).

Turning our attention now to (2d), we consider that the new information we may wish to obtain is related to our hope of predicting the behavior of some dynamical system in the world. The ontology casts the dynamical system in as a control system with inputs driving outputs via the system response. The workflow contains tasks intended to model the dynamical system through simulation (2e), and therefore interprets the simulation as an actor in the workflow. The simulation is itself a program, which is a process, (2f) again, but one whose semantics are enhanced by considering regarding it as a process in a discrete control system, and more specifically as a discrete event system which applies notions of events in time and state durations to the system. To run the simulation (23) we need data, which for the more interesting problems is a field of information mapped to some portion of the problem domain. For many of the problems of interest in, say disaster response planning, these fields are GIS coverage data sets. However, they could be data sets with arbitrarily complex relationships to the problem domain. Field concepts that employ more general domain representations such as cell complexes, and more sophisticated mappings like fiber bundles and sheaves, still fit within the ontology.

Finally, as the simulation completes its assigned task, we are left with additional artifacts (2a), new information obtained from the simulation. Given good representations of the data as fields, the challenge of relating the data back to the real-world domain of dynamical systems becomes more manageable, since the problem semantics contained in the domain representation and mapping are there to aid interpretation. The GIS community already knows that this approach works. A coverage, which is just a GIS version of a field, knows how to map, say the variable "rainfall" onto the relevant portion of the problem. GIS operations, like intersecting the boundaries of a drought-stricken farming region with the rainfall coverage field gives an immediately accessible interpretation of how the data relates to the problem.

## 4. Summary and Directions for Future Work

Motivated by the need to analyze complex event scenarios related to potential disasters and other kinds of incidents, we have looked into the problem of supporting complex decision analysis workflows in critical infrastructure protection. This has led to our encounter with the more general problem of scientific analysis workflows with simulations as embedded actors. To remove some of the complexity, we adopted a strategy that considers the problem domain from two related viewpoints, the world view and the computational view. We next examined the relationship between simulations and the overall workflow and have appreciated that the semantics of processes are generic to both the workflow and its systems of interest in the real world, as well as to simulations as computational actors in the computational view. From this point we considered that workflows and simulations have important semantic interpretations related to two different process metaphors. For the workflow, it is the semantics of planning to achieve a goal by performing a series of tasks, and have adopted concepts from the AI planning community for these. Simulations are most closely related to the semantics of dynamic control systems. The simulations are themselves readily interpreted as discrete event systems in terms of their behavior. Their effects, from the workflow planning perspective, are best described as consuming and producing data sets. However, in order to successfully preserve meaning, many of these data sets are best represented as fields that map information to some part of the problem domain. GIS information systems do this very well, and yet GIS data is just one kind of problem field representation. The encapsulation of data sets as fields aids the overall workflow, since as workflow artifacts they are self describing in terms of the problem semantics.

We believe the results of this conceptual exploration may be helpful in several ways. First of all, we see this work as just one more piece of evidence suggesting there is much to be gained from a fusion of efforts relating web-based enterprise system development and scientific workflows. We believe the ideas presented are in good alignment with concepts emerging from both communities. Next we think it is helpful to have an overall semantic architecture that provides a landscape in which to place concepts like simulation data management and automated workflow planning so that they can be seen in reasonable proximity. We hope to have made a start on that with this architectural sketch. Finally we believe there are great opportunities for the GIS community to provide expertise in solving interoperability challenges like those that have been discussed here. The idea of fields is natural to GIS analysts, and many of the mathematical tools that are required for practical implementations, such as computational geometry and topology, and the routine use of coordinate charts and atlases, as well as the use of cell-based domains. We also suspect that as simulations become more commonplace, GIS services may need to be better informed about how simulations and geographic data object may be required to interoperate in decision analysis workflows. In our own work we anticipate expanding our simulation frameworks in this direction, particularly in assessing infrastructure fragility.

## References

1. "Critical Infrastructure Collaborative Environment Architecture: Computational Viewpoint," Viewpoint Specification OGC 03-063r1.

2. John Radke, Tom Cova, Michael F. Sheridan, Austin Troy, Lan Mu, Russ Johnson, "Challenges for GIS in Emergency Preparedness and Response," ESRI White Paper, May 2000.

3. The GALEON OGCnetwork, see http://www.ogcnetwork.net/galeon.

4. Russell Bent, Tatiana Djidjeva, Birch Hayes, Joe Holland, Hari Khalsa, Steve Linger, Mark Mathis, Sue Mniszewski and Brian Bush, "Hydra: A Service Oriented Architecture for Scientific Simulation Integration, Proceedings of the 42nd Annual Simulation Symposium (ANSS), San Diego, CA, March, 2009, and Los Alamos Technical report LA-UR 08-07830.

5. "Needs Assessment for Complex Event Modeling, Simulation and Analysis," Sandia Technical Report SAND-8029P, or Los Alamos Technical Report LA-UR-08-6510, September, 2008.

6. Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, $2^{nd}$ ed., pp 375-461, Prentice Hall 2003.

7. Hilmar Schuschel and Mathias Weske, "Integrated Workflow Planning and Coordination," *Database and Expert Systems Applications*, LNCS 2736, pp 771–781, Springer 2003.

8. John Dorsey, *Continuous and Discrete Control Systems: Modeling, Identification, Design, and Implementation*, McGraw Hill, 2002.

9. Bernard Zeigler, Herbert Praehofer, Tag Gon Kim, *Theory of Modeling and Simulation: Integrating Discrete Event and Continuous Complex Dynamic systems*, $2^{nd}$ ed., pp 460-467, Academic Press, 2000.

10. "Web Coverage Service (WCS) Implementation Standard Open Geospatial," Arliss Whiteside and John D. Evans eds., document reference 07-067r5, OGC Consortium Inc., 2008.

11. D. M. Butler, and M. H. Pendley, "A Visualization Model Based on the Mathematics of Fiber Bundles," Computers in Physics, Volume 3, Issue 5, pp.45-51, September 1989.

12. Rene Heinzl, Michael Spevak, Philipp Schwaha, and Siegfried Selberherr, "A High Performance Generic Scientific Simulation Environment," *Applied Parallel Computing: State of the Art in Scientific Computing*, LNCS 4699, pp 996-1005, Springer 2007.

13. Allen Hatcher, *Algebraic topology*, Cambridge University Press 2002.

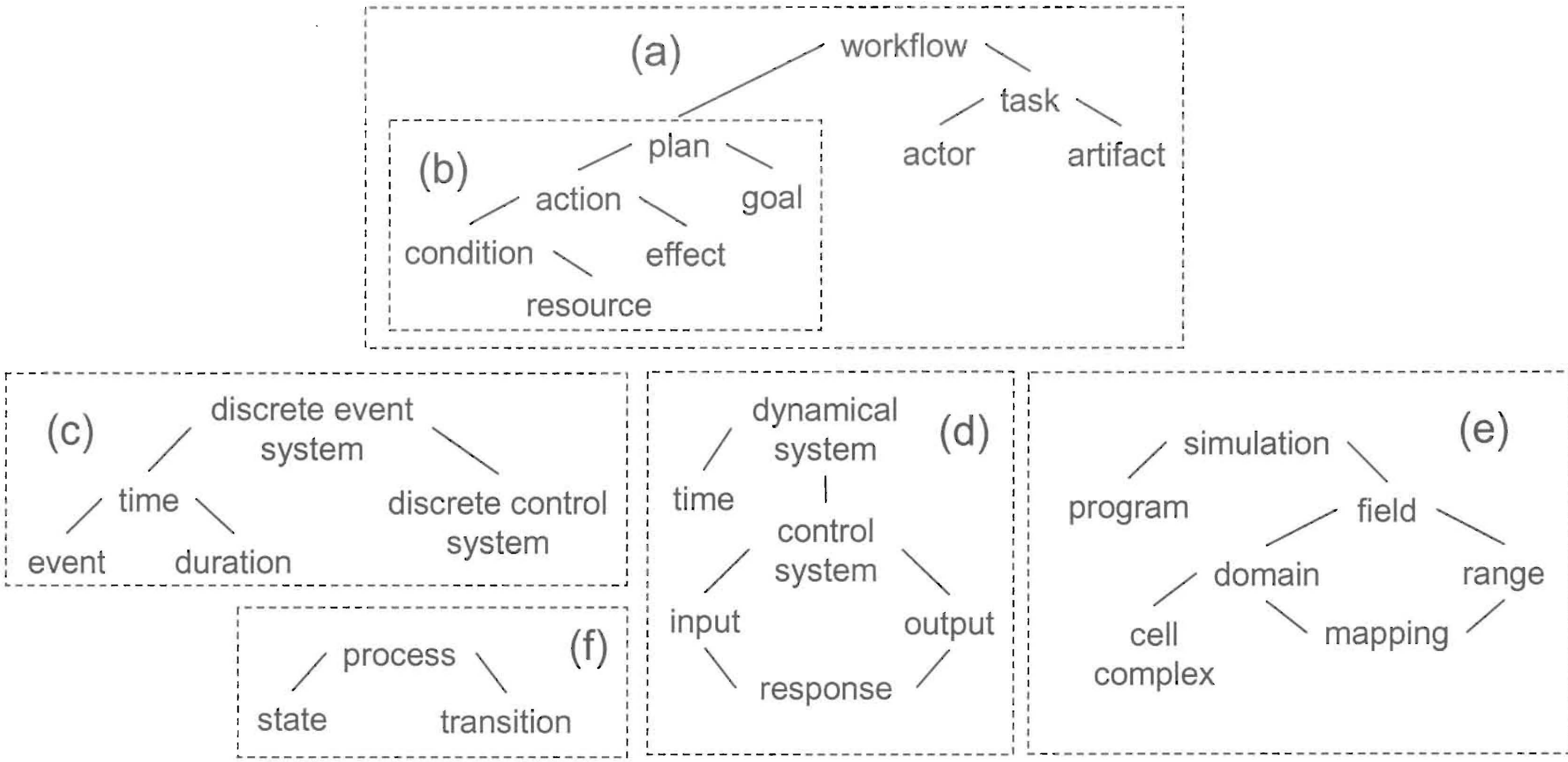


Figure 2. The conceptual building blocks of an interoperable architecture for decision analysis workflows involving simulations.