

# SRI DAML Project Plan for 2001

Jerry R. Hobbs, Grit Denker, Patrick Lincoln, David Martin,  
Jose Meseguer, Srinu Narayanan, and Richard Waldinger  
Artificial Intelligence Center  
SRI International  
Menlo Park, California

In our work on the DAML project in 2001, we plan to develop several key ontologies, as well as techniques for articulating different ontologies and for doing inference with ontologies. In particular, we are collaborating in the development of the DAML language and ontology elements for the specification of services on the Web. We will examine a number of web resources from the standpoint of characterizing them in terms of the ontologies. As a useful example of this work, we plan to implement and “market” a tool that supports non-expert users in maintaining resource websites (collections of links), and transparently generates DAML markups, making use of these ontologies.

## Ontology Development

Web pages have content, capabilities, and modes of access. Therefore, key to characterizing web pages is a development of ontologies of the structure of documents, of processes and services, and of security, privacy and trust. We will work on all three of these.

1. **The Structure of Documents:** We have sketched out a core theory of the structure of information and the structure of documents, explicating for example various ways different parts of a document can be related to other parts and to external entities and information. We will examine a number of web pages and other documents within this framework and develop the ontology on this basis. This ontology will generalize and extend the ontology we have already implemented for publications.

2. **Processes and Services:** For describing the capabilities of active web pages, in conjunction with Sheila McIlraith of Stanford KSL, Katia Sycara of CMU, Mark Burstein of BBN, and Ora Lassila of Nokia, we have

defined an initial minimal ontology of processes and services. We expect to release it for feedback in the next few weeks to the larger DAML community interested in services. The ontology consists of simple events, means for composing events into more complex events, and a simple theory of time including instants and intervals. Future layers of the ontology will address the execution status of processes, an ontology of resources, temporal measurement and the clock and calendar, and the structure of particularly important subclasses of events. As we develop this ontology, we will use it in characterizing various web sites that we believe will push its limits. The ontology is based on a core theory of processes that we have already explicated.

**3. Security, Privacy, and Trust:** Many web sites require logging in, are restricted to particular sets of users, or have other entry properties. E-commerce and other sites enable the transmission of credit card numbers and other confidential information, and are protected by various security measures. Various levels of trust are to be attached to various kinds of information found on various sites. We will develop ontologies of security, privacy, and trust, and use them in describing the relevant properties of various web sites.

### **Articulation of and Inference with DAML Ontologies**

As DAML ontologies proliferate, it will be necessary to relate them to each other. Much of the work in this area heretofore has been based on graph matching. Our view is that sometime complex mappings must be defined between ontologies or that “articulation” axioms must be stated that specify how the elements of one ontology relate to the elements of the other. Moreover, a DAML query language will be the more useful the more inference capabilities it has. We will be investigating these issues using our highly developed inference systems, Maude and Snark. Specifically, we will investigate applying specification and deductive inference techniques as part of ontology building and ontology composition, as follows:

**1. Translation between Formalisms.** To take advantage of high-performance inference engines, we need to translate between DAML and other formalisms, including the first-order logic that is understood by Maude and Snark. Similar translations may be necessary between DAML and other, more domain-specific formalisms; for example, we have used Maude to translate from Bibtex into DAML.

**2. Mapping between Ontologies.** When different ontologies for the same topic have been specified by different groups, it is necessary to express

the relationship between them. For instance, several groups have developed different ontologies for publications. The facilities of Maude are well suited for this purpose.

**3. Modular Construction of Ontologies.** To make ontologies highly reusable, it is advisable to develop them by composing more elementary ontologies. For example, if there are ontologies for dates and for places, it is natural to compose them to develop an ontology for schedules of events. We will study ontology composition operations based on the module algebra of Maude and expressing articulation axioms in Snark.

We will also investigate applying these techniques to query processing and search in DAML.

**4. Inference for Task Handling.** When a complex task is posed, it is necessary to decompose it into more manageable subtasks. For instance, if we need to obtain a certain publication, we must identify it, find where it is located, and obtain a copy. This is an inference problem which we can investigate using Snark's capabilities, and moreover will exercise the service ontology we develop.

More generally, these tools will be used to test the viability and utility of the ontologies we or other sites build.

### **A Tool: A Resource Resource**

Some of the most useful web sites are "resource pages": topic-oriented collections of links to other web-available resources. There are many such Web pages out there, in many different domains - and when a good one is found, it's considered a valuable resource. (For instance, many if not most users have bookmarks for resource pages in such areas as medicine, hobbies, financial topics, travel information, humor, etc.) Resource pages of a range of sizes are maintained by both commercial and non-commercial interests.

As a demonstration of the utility of our ontology building and inference efforts, and as a way to catalyze the creation of a large set of DAMLized web sites, we plan to develop a tool, a "resource resource", that will make the construction and maintenance of these resource pages more convenient, and will as a by-product produce a DAML encoding of the web site, thus making it available for a DAML-based search engine. The tool would provide a user interface that would reflect the underlying ontologies but would hide the ontologies and DAML itself from the user.

**An Ontology of Resource Pages:** In our view, a description of a resource page should consist of at least three things:

1. Its topic or topics. This could, in some cases, come from well-established taxonomies of topics, such as Yahoo's, or the ACM computer science topic categories, or similar taxonomies in medicine. In other cases, the most useful topic taxonomies might be far more collaborative, distributed entities maintained by interest groups of users.

2. A list of items. For resource lists, the items will generally be web pages or other Web-accessible documents, and in some cases including Web-based services. The description of the individual items will thus tap into the ontologies we develop for the structure of documents, for processes and services, and for security. For specific domains (see below), these will be augmented by domain-specific ontologies of varying degrees of expressiveness.

3. An intensional description of the list. This is the description of the kind of entities the list is intended to contain and the coverage it aims at, independent of what it actually contains at any given time. If the list is of hotels, for example, the description of the list should indicate how completely the set of hotels in the region are covered. Is it all, or only the most expensive, or only those belonging to a particular chain? This description will depend on an ontology of lists or sets, together with the ontology of the individual items.

Some resource sites generate their lists of resources on the fly, in response to user queries. For these sites there would be no extensional list of items, only the intensional description of the lists the site can produce.

**Other Features of the Resource Resource Tool:** The ideal, long-term version of the tool would include other useful features. It could provide a variety of ways to assist the user in creating and editing resource entries. It could also provide a variety of community-building and community-leveraging aids. First, the tool could support the user in identifying a dynamic collection of other similar resource pages, with some indication of their relevance and value relative to the current resource page. Given that, the tool could notify page-maintainers (and ultimately users) of a resource page of modifications made to similar lists. With DAML-encoded semantics provided by this (and other) tools, these features would be relatively easy to implement, and far more effective than current approaches to online community-building.

The initial version would contain only the minimal subset of these features and may be applicable only to a relatively narrow domain. But we will place a high priority on getting something useful implemented and dispersed in the relevant communities.

Marketing Strategy. Once a DAML search engine is available and an initial version of our resource tool is implemented, then in collaboration with the search engine developers and with the imprimatur of DARPA and W3C, we would make the tool available to as many of the webmasters of the appropriate resource lists as possible, where the inducements are greater convenience in maintaining the lists, ready access to similar resource lists, greater visibility to the appropriate communities of users, and more hits.

## A Domain

It is necessary for our research and development to be grounded in a concrete domain, simply because reality has a way of presenting you with challenges and solutions you can't anticipate in the abstract. The criteria for a good domain are a very large potential user population and a great diversity in the kinds of content, capabilities, and modes of access that must be represented. The diversity drives the research and the large population makes it useful.

We have chosen the domain of travel, broadly construed. Virtually everyone in the world travels or wants to travel. Making travel arrangements is one of the most utilized and most visible functionalities on the web. Insofar as we succeed in making the navigation of travel-related sites convenient, the potential user population is enormous.

It is closely related to other domains of similar complexity and ubiquity. People not only arrange flights, rental cars and hotels. They also buy books about places they travel, They look at maps. They schedule other events, such as the theatre and concerts. They register for conferences. They upgrade their cameras. They chat about their experiences. Building ontologies of travel domains leads naturally to ontologies of these other domains, making travel a good domain for initiating the "bobsled effect".

There is a great deal of variety in the capabilities of travel sites, from simply exploring a set of options to making purchases. Moreover, the task of planning a trip provides a good research locus for the automatic composition of services.

At the same time, there are many travel resources for the same topics – hotels, flights – so that relatively small ontologies could have a big impact early on.

In the coming year, we will examine several categories of travel sites and develop the domain ontologies that these require. We expect in subsequent years of the project to begin to build tools that would make the construction of these ontologies easy even for those who don't know that DAML exists.

The travel domain will provide a good opportunity for collaboration with other researchers, including Sheila McIlraith at Stanford KSL, Katia Sycara at CMU, and others. We have not yet decided upon a scenario, but it is likely to involve the automatic composition of travel-related services, probably including our resource resource, using the tools and the DAML encodings of sites that various research groups provide.