ITR: Science on the Semantic Web: Prototypes in Bioinformatics

A proposal submitted to NSF 02-168
Information Technology Research

A medium scale, five year project

PI: Tim Finin (UMBC)
CO-PI: Jim Hendler (UMCP)
CO-PI: Jim Quinn (UCDavis),
CO-PI: Neo Martinez (SFSU)
CO-PI: John Schnase (NASA)

PARTICIPATING INSTITUTIONS

San Francisco State University (Martinez)
UMBC Computer Science and Electrical Engineering (Finin)
NASA Goddard Space Flight Center (Schnase)
UMCP Computer Science, MIND Laboratory (Hendler)
UC Davis Environmental Science and Policy (Quinn)

February 13, 2003
A. Project Summary

We will develop a framework for conducting science research and education on the semantic web, and will implement and evaluate prototype tools and applications for use in the biocomplexity and biodiversity domains. These capabilities include the ability to collaborate and convey meaning through the automatic and semi-automatic semantic annotation of web documents; to improve information retrieval using background knowledge and inference; and to extract and fuse information from multiple, heterogeneous sources in response to a query. Our major testbed for prototyping these capabilities will be the web portal of the National Biological Information Infrastructure (http://www.nbii.org).

Intellectual Merit. Our team brings together researchers from the forefront of the semantic web, data management, biodiversity, and ecoinformatics research communities. This is an excellent match -from the point of view of research in biodiversity, there is a growing consensus, amongst leading organizations such as the Global Biodiversity Information Facility (GBIF), on a semantic web approach. From the computer science point of view, the diverse user community and the semantic complexity inherent in biodiversity studies makes them the right choice for the initiation of large scale deployments of semantic web technologies to benefit scientific research and education.

Our framework will include specifications for ontologies, protocols, agents (for discovery, query, analysis, etc.), and tools for authoring, automated ingest, and annotation. These tools will leverage collaboratively constructed ontologies to bring diverse communities together, and to enable the community construction of scientific knowledge. In addition to the domain-specific ontologies needed to capture the conceptual space for the biodiversity testbed, we will develop general purpose ontologies to enable metadata about the contents and structure of databases and other knowledge repositories to be expressed in emerging knowledge markup languages such as RDF, DAML, and OWL. This will enable agents to both access and index the hidden web, and will also enable the data mining of diverse and distributed databases.

Broader Impact. Our goal is to allow knowledge from one community to be effectively used by another, even in cases where the communities do not normally interact. Such is the case for the disparate physical and biological science communities contributing to the inherently multi-disciplinary study of biodiversity. In addition to the community of professional scientists who study biodiversity, "K through gray" learners are prime examples of potential users who do not fall within any particular discipline, since the learner is not a member of any professional scientific community. In the biodiversity domain, the K-gray audience is also an important generator of new data, as the majority of discoveries of rare species and new invasives (in addition to many other important observations) are made by interested amateurs rather than by professional scientists.

We will prototype an education gateway into NB II which will engage learners directly in the scientific process, and will foster and encourage the publishing of their results in an on-line student Journal of Biodiversity. We will pursue three distinct approaches to infusing this new learning technology into the classroom.

By providing access to and integration of biodiversity data and information, our tools will be enabling technologies for the National STEM Digital Library (NSDL). We will exploit the close ties between our team and the NSDL to ensure that maximum synergy is achieved. These tools will also be valuable contributions to the Global Biodiversity Information Facility, of which the United States is the first national charter member (see http://www.gibif.org/). Finally, while our tools will be developed specifically for the field of biodiversity research, the ideas and technologies will be applicable in other areas of science, and in many other domains as well.
C Project Description

C.1 Introduction

Our objective is to transform the web from a universal file system that stores scientific information and simple programs to an active ecology of agents that produce, consume and act on information. To do this, we will develop, prototype, and evaluate systems that provide, find and manage scientific information and services on the Internet. Our approach is based on two concepts: the use of semantic web languages such as RDF, DAML and OWL that can be used to encode information in a machine-understandable form and the development of agent-mediated software frameworks that can be used to construct and provide dynamic services on the web.

To validate our approach, we will build prototypes for incorporation into the National Biological Information Infrastructure (NBII -- see support letter from Gladys Cotter, the chief USGS official responsible for the NBII). Our main focus within NBII will be the detection and early warning of new invasive species, thought to be one of the two most important causes of declines and extinction of rare species (Schmitz and Simberloff, 1977; NRC, 2002). It is estimated that invasive species cost the U.S. economy over $130 billion per year (Pimentel, et al, 1999) and detection of invasive species has become a national priority. Invasives typically cannot be eradicated once they have infested more than several hundred hectares, so timely detection and immediate control are essential. Because of this heavy cost, much investment has already been made in creating a cyberinfrastructure to combat invasive species. Specifically, the USGS National Institute for Invasive Species Science, and the joint NASA/USGS National Invasive Species Forecasting System have been chartered to document, map, and predict the distribution of all non-native plants, animals, and diseases in the U.S. It is becoming increasingly clear that their mandate will be impossible to fulfill without the semantic web developments we are proposing. (See Figure 2 and letter of support from Tom Stohlgren, director of the National Institute for Invasive Species Science) These investments themselves stand on the national (NBII) and international (GBIF) commitments to biodiversity research, which have resulted in a large degree of standardization and coalescing around common metadata standards and shared vocabularies. In other words, much of the formalization necessary to build an interdisciplinary semantic web has already occurred in this domain, and the community is embracing a semantic web approach.

Environmental biocomplexity in general, and invasive species detection in particular, are highly interdisciplinary studies, requiring collaboration and data sharing amongst specialists in the fields of systematics, ecology, and evolution, each of which has its own partially shared vocabulary and way of seeing the world. This makes the invasive species problem an ideal candidate to be a scientific testbed for the semantic web paradigm. The data discovery, knowledge sharing, and collaboration problems faced by practitioners in the fields that compose environmental biology are exactly the problems that the semantic web is meant to address. It is likely that tools developed to address these problems will be broadly applicable across many scientific disciplines.

The same capabilities that enable scientists from diverse disciplines to communicate and share research will also enable new learning technologies. We will demonstrate this by prototyping an educational portal to NBII. This portal will include an on-line student Journal of Biodiversity. Dynamic tutorials will guide the learner through the process of formulating a scientific question. The learner will
then use NBII resources (including those developed as part of this effort) to answer that question, and to write a paper for review by teachers and peers. The goal is to increase scientific literacy, where that term is taken to mean not simply an understanding of scientific concepts, but also an understanding of the scientific process. Students will have the opportunity to contribute to the nation’s battle against invasive species, while at the same time constructing their own scientific understanding. Finally, tying all aspects of our project together, an annual competition will engage students nationwide in exploiting semantic web technologies for biodiversity research.

Our team brings together researchers from the forefront of the semantic web, data management, biodiversity, and ecoinformatics research communities who will work together to explore these ideas, build prototype tools and systems, and evaluate their effectiveness. Figure 1 describes the synergy among the different high level components of our project. Biological Science has requirements that give rise to Technical Requirements which are addressed by our Technical Objectives. We manifest and explore these by building Prototypes of tools and applications, which are used to support Biological Science and its researchers and students. Our goal is to view this not as a circle but as a continual development spiral, approached through rapid prototyping.

C.2 Intellectual Merit and Scientific Justification

The standards of the emerging semantic web are currently being established. This proposal derives from our intention that these standards be informed by the needs of the scientific community. The leadership role played by our technical team in the specification of semantic web languages ensures that science will be a driver for these languages. The leadership role played by our science team in national and international standardization and coordination efforts will be the vector for the promulgation of our tools throughout the biodiversity research community. Our chosen application domain – early detection of invasive species – is seemingly narrow. But success in this domain will produce capabilities that will empower broad classes of scientific communities to realize the full promise of the web:

- To present an integrated view of available information, agents will need to fuse information from different sources despite inconsistencies and ontological differences.
- Queries may require information not directly available; therefore some computational process (such as statistical calculations or knowledge-based reasoning) may need to be activated on a remote site.

An important outcome of our work will be the creation of paradigmatic examples of semantic web systems that researchers in other disciplines can look to for guidance in constructing their own systems.

**Background on Biological Challenges.** Species that are introduced into ecosystems in which they are not aboriginal are classified as non-native or exotic. Invasives are the small subset of non-native organisms that, through uncontrolled spreading, damage or displace native species, disrupt ecological processes and productivity, or threaten human health. Famous invasives include West Nile virus, Chinese Snakehead fish, and Mad Cow disease; not so famous invasives include sudden oak death, leafy spurge,
and innumerable algae. Several thousand weeds, crop pests, plant diseases, disease-vector insects, exotic predators, etc. are of active policy concern in the U.S. The invasive species problem is growing, as the number of pathways of invasion (ship ballast water, airplane wheel wells, highways, disease vectors, human agents, etc.) increases. Information needs for challenging invasive species include:

- Real-time harvesting of highly distributed information – agricultural inspections, Park Service surveys, 4th grade classroom projects on bullfrog counts, etc. – each with its own semantics and degree of trustworthiness.
- Construction of potential population distributions – spatial representations of the likely presence of invasives throughout a region. The reliability of these distribution estimates increases massively with the number of surveys that go into their construction. Unfortunately, so does the complexity of semantic integration.
- Prediction of how an ecosystem is likely to respond to the introduction of a new species. This requires an understanding of the food web (“who eats whom”) [Williams00] of the invaded ecosystem, as well as the ability to compare this web to that of a similar ecosystem to which the invasive species is native. Predictive ecology has traditionally been hampered by the extreme difficulty of comparing field work on similar ecosystems in different parts of the country.

An analyst needs to quickly assess the risk of releasing the algal species in nearshore waters. The sequence of inquiries might be:

1. **What species is it?** The "worst invaders" list of the Global Invasive Species Programme and the Smithsonian Marine Invasive Species database both provide identifying features and photos from the many participating countries. Several photos are possible matches and these lead her to images of herbarium voucher specimens (to be) compiled under the Global Biodiversity Information Facility, annotated by several experts, with recent and historical nomenclature for those species, genetic information, and similar species.

2. **What could it do?** Once identified as belonging to the genus *Caulerpa*, which includes highly invasive species, the analyst accesses the alga's ecological information, including methods of reproduction and dispersal, ecosystems and habitats invaded by the pest in the past, and types of risk. Transport models serving the area and comparative oceanographic conditions in source countries are identified from the clearinghouse at the San Diego Supercomputer Center. The data are passed to both the Species Analyst model at the University of Kansas, the BioClim model at CSIRO and the National University in Australia, and the invasive species co-Kriging model at USGS/Colorado State to predict the potential range of infestation. Subsequently, fisheries models are used to predict the impact of the loss of groundfish habitat if the alga became established.

3. **What can be done about it?** The analyst accesses information about control methods and preventative measures for related species from fact sheets developed by USGS, the Hawaii Ecosystems at Risk project, Monsanto, and CSIRO. Experts are identified from NBII and IABIN experts registries, and a taskforce is recruited to develop screening tests, a protection plan, and ballast water exchange regulations for ships from infested points of origin.

Figure 3: Consider a hypothetical example in which the Coast Guard detects an unidentified alga in the ballast of an inspected ship in San Diego harbor.

C.3 Research Agenda

With all its changes, the web today has the same audience that it had ten years ago: people. In this section, we describe the capabilities that we will develop to enable a new audience: software agents. The need for this change, vis-à-vis science, is simple: there are not enough graduate students in the World to analyze the available data. Just as the web was created to serve the needs of the worldwide community of physicists, so do we hope that by serving the needs of ecologists, the semantic web will ultimately change all of science. Realizing the semantic web will require addressing three central problems: (1) where will
the semantic markup come from; (2) how will it be used; and (3) how can it be managed and maintained. We will address fundamental aspects of all three. Here are the activities we will pursue.

**C.3.1 Ontology Formulation.** One approach to the problem of ontology heterogeneity in an interdisciplinary domain is through the creation of a global schema that serves as an interlingua for human and software agents. This is not our approach. We do not believe that the conceptual space occupied by the biodiversity domain (or any other domain of active, interdisciplinary research) can be captured by a single, consistent ontology. Therefore, the construction of a ‘global schema’ is not our goal. Rather, we envision a number of relatively small ontologies, some of which may overlap, and some of which may be in conflict. Our approach to semantic mediation is through the construction of binary mappings amongst all relevant ontologies. Thus, rather than have a single wrapper ontology, any ontology can serve as a wrapper for any other. Dynamic ontology negotiation is thus far more likely (see section C.3.5). Although it is an open question whether this approach will work for the web as a whole (due to the vast number of micro-models that people use to express themselves), there is strong reason to believe it will work in science domains, where there is typically general agreement on an underlying paradigm, coupled with a small number of formalizations of various aspects of that paradigm. Even in the case where there is no dominant paradigm, there are typically a small number of schools of thought in which research is grounded.

A major activity of our first two years will be the construction of the ontologies necessary to support the queries of biodiversity researchers. This effort will be led by the SFSU and UC Davis teams. The first and easiest step will be the conversion of existing formalisms (e.g., the ITIS thesaurus; the Darwin Core, several related NBII/USGS invasive species information catalogs and GIS systems) into OWL ontologies. The next step will be to create mappings amongst the concepts expressed in these ontologies. More difficult knowledge engineering tasks will involve the construction of ontologies for ecological interactions, and geospatial reporting and querying. We will not be starting from scratch, as a number of groups have already begun investigating these areas.

The USGS has committed to funding workshops that seek broad community involvement in this ontology-building effort, and the Santa Fe Institute has offered to host these workshops. (see attached letters of support.) USGS and USDA-sponsored workshops in the last year, two hosted by the UCD group, have made considerable progress in defining consensus semantics and core ontologies for cross-agency invasive species applications.

**C.3.2 Addition of Semantic Mark-up.** The most effective way to mark up data is at the source. Just as tools like Oracle’s WebDB automate the generation of web content from relational databases, so will our tools automate the generation of semantic web content from relational databases, GIS systems, and other data sources. Since a goal of NBII is to house data from all its providers (literally thousands of federal, state, and local agencies) in database or GIS systems, our success in this area will have huge impact on our target domain.

**Markup by Humans.** Our tools need to allow the domain expert, who is not necessarily an IT expert, to create such mark-up. Such tools must allow for the markup of individual web pages, in the same way that HTML editors allow non-experts to create and publish web pages. Figures
4 and 5 show screen shots of the SMORE prototype markup editor we have developed that makes it easy for authors to link to multiple ontologies as they create a web page, document, database, image archive, or any other web-based resource.

These tools eventually must be built in a way that they tie into the “business processes” of the working scientist—that is, rather than learning a whole new set of tools, the basic web tools of the scientist must include mechanisms that make it EASIER for the scientist to produce web content (appropriately marked up) while authoring papers, performing experiments, creating and logging data, and the other day to day activities of the working researcher.

The UMCP group has been a major developer of such tools, and has produced to date a tool kit providing the first prototype tools of these types. The tools make it possible for the scientist to create RDF data tied to ontologies (written in DAML+OIL or OWL) and to harness and collect that data to make other processing easier. Thus, if the user defines a pointer to a particular concept, say “Asian Longhorn Beetle” – an instance of the class “invasive species” in some ontology -- then the ontological information is used to create a “form” that the user can fill in (either by drag and drop, browsing, or directly typing values). Fields such as “native ecosystems”, “known effects”, “know remediations” and the like are provided, and these may be either filled in with the values found in various data resources, or can allow the scientist to add or change values. Their local changes are stored locally, but linked to the more global information (much as a personal web page may be linked to a University site, but not change that site without further authorization). The information created can be used by search engines, markup tools, web page creators, and other web tools to provide further access for other users.

Despite the strong first steps we have taken in providing tools for scientists, they are only the start. The work funded by this grant will be used to develop new versions of these tools specifically tailored to the scientist, and tied to primary scientific data sources (for example, we are already working with both Nature Publishing Group and BioMed Central to build versions of our tools tailored to their content and processes) -- in this case, to the sources of NBII data. In addition, we will be working to extend these tools to do more in the area of database access and integration, to integrate with and manage Grid
Services and to develop new versions tied to primary scientific data analysis tools (including statistics packages and visualization tools). A key feature of our work will be a tight spiral development of our specialized tools with the scientists we are supporting, to make sure the tools fit into their processes and enhance and reinforce scientific research, rather than derailing it with new tools not tailored to scientists needs.

**Markup Through Machine Learning**

One of the potential impediments to the widespread implementation of the semantic web is the cost of adding semantic information to documents. Billions of pages already exist without such markup, and there is currently little incentive for authors of new web pages to add machine-readable semantic content. To help overcome this hurdle, we are developing a system that takes as input a small corpus of documents containing DAML tags and learns how to add such tags to similar documents lacking them.

For tasks that involve text understanding, like adding DAML tags to web pages, computer programs will not perform as well as humans. However, we believe that learning systems can play two key roles. First, because brain cycles are expensive and CPU cycles are cheap, learning systems can magnify the impact of a small hand-annotated corpus by annotating similar documents. As the number of DAML-enabled web pages increases, applications will proliferate, and the incentive for humans to annotate additional pages will increase, leading to more fodder for the learning system. Second, given a collection of documents with DAML tags, as the content of the documents changes, the learning system can automatically update the tags to reflect the new content. This will make maintenance of the semantic web, which will be as dynamic as the current web, easier and less prone to errors of omission.

Our current work in this area has been tested in the domain of talk announcements, such as are found on web sites of computer science departments. Such announcements give information about the speaker, topic, location, date, time, and so on. We have implemented a system, called Stalker [Muslea01] that learns to add hierarchically organized tags, such as HTML or DAML, to documents by extracting rules from a small corpus of previously annotated documents. We developed a number of extensions to the basic algorithm that yield significant improvements in precision and recall for most DAML tags in talk announcements from a variety of sources (i.e. university web sites) [Krueger03].

Over the next five years, we will extend this work in a variety of ways. The issues that we will address include the following: As the pages to be annotated decrease in similarity to the pages from which rules were acquired, the quality of the resulting annotations decreases. An important research goal is to develop methods that will automatically determine when the learner-agent needs to ask for additional training documents from a human so as to maximize quality and minimize expense. This will also make it possible for the system to seek out documents on the web and make judgments about whether it is capable of adding meaningful tags. Other sources of information can be incorporated profitably into the system. For example, we will explore the use of expert modules that have information about instances of a particular ontology to increase the accuracy of the system. Ultimately, the semantic web itself will be able to serve as an expert module in some domains.

**C.3.3 Markup, Discovery, and Composition of Scientific Services.** Consider a network environment which includes two types of services; data-producing services and data processing services. An example of a data-producing service might be one associated with a sensor which is producing data, a processing one can be one which runs a FIR (far infrared) filter on that sensor data. At a more complex level, this can include data produced by a biological sampling or sensing device which would be processed by some sequencing or visualization software, or other such complex program. The concept of a “semantic grid,” linking databases to processors on the computational grid (represented in the Open Grid Service Architecture [OGSA] by use of semantic web information management techniques is a special case of this, and one we will explore in this project. We will also use other services including device services, written in Microsoft’s Universal Plug and Play [UPnP] language and in the web standard Web Service Description Language [WSDL]. One of the unique features of our research is the ability to combine all
these together by use of an extended version of the DAML-S\(^1\) [DAML-S] language. We have already done full groundings for WSDL and UPnP and have a partial grounding for OGSA. This funded research will enable us to extend our WSDL tool set to OGSA, and to work on the further automation of service composition.

We have already built a semi-automated composer [Sirin03] which creates a workflow of services that can solve the user’s need in a goal-driven way. The user starts the composition process by selecting one of the services registered to the composer, and specifying some input to that process. For example, the user can choose “FIR filter” as a service and provide the input “a sensor service” (meaning the system would be free to choose one) or some specific sensor. Similarly, the user can specify a particular visualizer or analysis device and a specific dataset, or can specify (using an ontology) any data set meeting certain characteristics. The system uses a filtering technique based on the "non-functional

\(^1\) Co-PI Hendler was the creator of the DAML-S project, and serves as a member of the steering committee for both the US DAML-S effort and a new international project, funded jointly by DARPA and the EU/IST program, to develop a new version of the language as a prequel to industrial standardization. PI Finin and co-PI Hendler both serve on Working Groups within the World Wide Web Consortium working on standardizing service descriptions on the web, and providing a liaison between DAML-S and these industrial standardization efforts.
attributes” of the service—that is, the ontological properties that are not directly inputs and outputs to the service. In the case of a sensor these are features such as sensor location, type, deployment date, sensitivity, etc. The system is also extensible, which we believe will be an extremely important functionality for use by scientists. Any composition generated by the user and the system can be automatically realized as a DAML-S CompositeProcess, thus allowing it to be reused at a later time, or used by the system for composition with other services.

Our primary research will be on extending the composer in a number of ways. First, the current implementation of the system executes the composition by invoking each individual service and passing the data between the services according to the flow constructed by the user. This method is primarily dictated by the DAML-S and WSDL specification which both describe the web services as an interaction of either a request/response or as a notification messaging between two parties. However, this centralized coordination suffers from scalability and availability problems. Rather, we need a special framework where each node abides by a set of system rules to conduct the execution process by directly passing its result to the next service. (This is sometimes referred to as service choreography, an active area of development for business services on the web, but as yet largely unexplored for scientific data and Grid services—see [Hendler03] for a discussion of this problem and related work in the area.)

Second, research in Grid computing, a critical infrastructure for scientific service composition, is exploring the development of OGSA, which is an extended version of WSDL, which defines specific “service description elements” and “ports” making the services available on the grid. We are developing a version of DAML-S that can directly reason about the properties of OGSA services and extend the OGSA capabilities into new choreographies, managed by the composer. This would allow the service composer to generate a plan of how to achieve a goal and produce a workflow. This workflow would then be handed to an analysis system to examine it for cost, quality of service, efficiency, resource use, etc.

Third, service composition management and execution monitoring is an as yet largely unexplored area. When the workflow developed above is run against a set of scientific data, many problems can arise - these can be computational problems (a needed resource is unavailable, perhaps a crashed server or attempt to access a device without proper authority) or problems relating to the scientific process itself (lack of provenance on the data, datarange issues, unexpected dataset noise, etc.) These latter problems also include issues of later reuse of this data—for an extreme example, consider those papers based on the Lucent Laboratory materials reports that are now considered to be fraudulent. Many results based on that data will now need to be re-explored. The service composition and monitoring system could track the derivation and sourcing of data to help with these problems, as well as with more mundane scientific problems such as date of publication, authorship, etc.

Researchers at the NBII’s California Information Node and the National Institute for Invasive Species Science have spent considerable time building composite services, out of S-Plus and ArcInfo functions, to build distributions from point data on the basis of several assumption sets (see Figure 7).

<table>
<thead>
<tr>
<th>Current Predictive Modeling Capabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. ArcGIS: Input satellite data, veg., soils, topography, etc.</td>
</tr>
<tr>
<td>3. S-Plus: test residuals for auto-correlation and cross-correlation (Morans-I) and find the best model (ordinary least squares, gaussian, etc. using AICC criteria).</td>
</tr>
<tr>
<td>4. S-Plus/Fortran: If spatially autocorrelated, run kriging or co-kriging models.</td>
</tr>
<tr>
<td>5. ArcInfo GIS: develop map of model uncertainty from S-Plus output, Monte-Carlo simulations, observed-expected values.</td>
</tr>
<tr>
<td>6. ArcView: produce maps of current distributions, potential distributions, and vulnerable habitats, with known levels of uncertainty.</td>
</tr>
</tbody>
</table>

Figure 7. Algorithm pipeline for predicting potential invasive species geographic distributions.
This is a laborious process, and must be repeated for each new class of statistical query. Using the DAML-S markup component of our service composer to generate service description for the functions of common software packages will be one of our initial undertakings. The extent to which we can automate the construction and execution of this and similar pipelines is one of the research questions of this project.

C.3.4 Information retrieval and the semantic web. We will develop and prototype techniques for information discovery and retrieval over documents and queries that contain a combination of free text and semi-structured information in the form of semantic web markup and annotations. We will extend these to handle images, data sets, services and other non-textual objects containing metadata encoded in semantic web languages. Finally, we will investigate, analyze, and evaluate trade-offs in when, where and how much inferencing over the semantic web annotations should be done -- at document indexing time, during query preprocessing, post retrieval or some combination of all three. We have begun to explore [Shah02] techniques to support information retrieval over documents and queries that contain a combination of free text and semi-structured information in the form of semantic web markup and annotations and have implemented a prototype system. The ability to retrieve over semi-structured annotations and metadata will provide a means to extend retrieval to other content not normally suitable for IR processing, including images, audio objects, forms, and embedded data, scripts and programs.

While research has been done on hybrid information systems [Finin93, Abitebou97, Martin99] and information retrieval over documents that contain XML and other semi-structured data (e.g., [Fuhr00, Egnor, Bar-Yossef99, Chinenyanga01]) there has been no work to date on systems that combine free and semi-structured text with inferencing capabilities on the backend.

At the end of the first year we will have completed and released a basic system for indexing and retrieving documents with text and semantic web markup. The system will perform a focused crawl of the relevant portions of the web to discover documents, indexing them using both their text as well as associated semantic web annotations. A web-based query system will be developed to allow environmental scientists as well as students to perform searches over this collection using a combination of free text and semi-structured information. At the end of the second year, we will release an enhanced version of the system that will be able to handle environmental images and embedded datasets with their associated metadata. This will be supported by the development of appropriate ontologies for describing images and datasets in the environmental sciences domain and tools to add and annotate images and datasets in RDF and OWL. At the end of the fourth year, we will release a version which is able to extend the annotations of documents and queries with inferences licensed by the ontologies as well as those added by optional heuristic and expert systems. This will enable a system which goes beyond simple retrieval to do question answering using information retrieved from text, annotations, images and data. At the end of year five, we will have completed extensive evaluations of the complete retrieval system, studying and analyzing the contributions and costs of the different components.
C.3.5 Ontology Maintenance, Evolution and Mapping. The semantic web is not and never will be centrally controlled, with universally agreed-upon ontologies under the control of a single authority. Even within a domain, not all data providers will use the same ontology. Moreover, ontologies change over time as their users’ knowledge about and conceptualization of the underlying domain evolves. This is particularly true in an active scientific disciplines such as biodiversity and biocomplexity. For all these reasons we will need tools to track ontology evolution (versioning) and to map ontologies (or fragments thereof) to one another, in order to reconcile the differing underlying conceptualizations of ontologies and their inevitable evolution over time. Closely related is the “data provenance” problem [Buneman00,-01,-02] – relating derived information to its ultimate sources in a distributed environment.

We envision a set of base ontologies, each of which provides the conceptualization (concepts and relationships between concepts) for a scientific discipline (e.g., bioscience, physical science, environmental science, etc.) related to biodiversity and biocomplexity. These base ontologies will have variants that reflect different views or perspectives of individual researchers or groups [Weinstein99], or reflect the evolution of a given ontology -- its versions [Heflin00, Klein01]. As in real life, the evolution is gradual, with occasional large changes due to Kuhnian paradigm shifts. It is impractical and inappropriate to construct a huge, universally shared, ontology that horizontally encompasses base ontologies of all involved scientific disciplines and vertically their variants [Bailin01, Peng02], as suggested by some earlier work on shared ontology [Gruber93].

In the environment we envision, problems occur when an agent encounters a concept defined in a foreign ontology. Our objective is to develop a suite of enabling technologies that supports semantic resolution in such situations. Our concerns here are more practical than theoretical, and we will pay close attention to issues such as if a technical solution can be scaled up to a large number of ontologies, if a new ontological variant can be easily handled in a timely fashion, and if semantic mapping service can adequately support required reasoning tasks, etc. We will work closely with the W3C Web Ontology Working Group so that the solution we develop in this project meets the requirement they are developing for ontology evolution. Our basic approach is outlined next and involves three aspects – static translation, dynamic resolution, and versioning.

Static Translation. In general, there are only a few base ontologies in the system, and they are shared by many and are relatively stable. Therefore it is both desirable and feasible to construct translations between them before system deployment. These translations will be static (subject to versioning), providing a stable semantic basis for the entire system. Several approaches will be pursued for establishing such translations.

- A logic based approach Specifies a set of rules that logically relate the source concepts with target concepts (e.g., equivalent, similar etc.). This is largely manual, involving domain experts, and is similar to what is presently pursued by NBII and the ontology merging approach in PROMPT [Noy00] and OntoMerger [Dou02]
- A machine learning approach learns from exemplars to automatically classify source concepts into the target ontology. Text-based auto classifiers developed in IR communities (e.g., Rainbow [Prasad02]) are particularly attractive because concepts in scientific ontologies are often associated with text descriptions and may have links to related articles.
- Approaches based on ground instances relate concepts in different ontologies based on shared ground instances. This requires identifying a set of ground instances whose semantics is agreed upon by both source and target ontologies, this can be helped by techniques from natural language processing [Stumme01, Bailin01] and information retrieval [Prasad02].

These approaches are more complementary to than competing with each other. Moreover, some approximate mapping methods and metrics will be developed to address the issue of partial semantic matching between concepts from different ontologies [Stucken00b]. Architecturally, for the sake of flexibility and scalability, translations will be better not to reside within individual ontology servers but managed by special agents that publish the service and perform translation on demand.
Dynamic Semantic Resolution. The system will have many ontology variants reflecting the views of agents representing different participating parties. For such ontologies, it is more appropriate to resolve semantic difference dynamically (on-the-fly) when they occur at run time. Dynamic semantic resolution can be seen as an abductive or evidential reasoning process, consisting of a cycle of evidence-collection and hypothesis-generation phases [Peng02].

- **Evidence-collection.** Since variants are extensions of base ontologies, most local concepts are defined in terms of concepts of base ontology concepts. The definition of a source concept thus forms an important type of evidence, which can be collected through a sequence of queries (“What do you mean by estuary?” “What do you mean by outer limit in your definition of estuary;” etc.) via agent communication, much as people try to clarify meanings during a conversation [Peng02, Balin01].

- **Hypothesis-generation.** This phase identifies the target concepts as candidate mappings of the given source concept based on accumulated evidence. It is widely accepted that ontology mapping is seldom exact [Stucken_children00b]. Operations based on crisp logic often do not provide a satisfactory solution to this issue [Koller97]. Instead, we will develop similarity or confidence metrics to rank alternative candidate hypotheses based on some approximate reasoning frameworks (e.g., Bayesian [Pearl88, Koller97], fuzzy logic [Zadeh65, Straccia98, JoKri00], or rough sets [Stucken_children00b, Pawlak82]).

Managing Ontology Versioning. A sequence of versions of an ontology represents the evolution or incremental changes of that ontology. Different changes have different ramifications [Heflin00, Klein01]. Some are backward compatible (the inferences based on previous versions still hold in the new version), others are not (the inference becomes non-monotonic). When a new version is created, it needs to be made known to other components of the system, especially those that use rules extracted from the ontology for their computation. Our research in this direction will focus on version representation and control of version creation, dissemination and removal. We will also address related data provenance and data pedigree issues.

C.4 Educational Portal; on-line Student Journal; and National Competition

The educational gateway we build for NBII will comprise several components, which we describe below.

Early Warning Portals. The invasive species domain naturally integrates research and education. At UC Davis, Co-I Quinn routinely has students in his undergraduate classes make observations of both invasive and threatened species, and report them to the state Heritage program and NBII national invasive species datasets. Quinn is not alone. Throughout the country, teachers and volunteer groups at all levels have organized class outings to do species inventories of parks, ranches, back yards, etc. Students have been very excited to see their field reports show up on the web as part of the national ecological record. We plan on formalizing Quinn's approach by providing a portal that will allow students and teachers (and any interested amateurs) to learn what invasives their region is at risk for, how to identify those species (e.g., by providing polyclave keys), and how to report on their observations. This activity will not only capture the interest and enthusiasm of learners, but has the potential to mobilize a vast student and volunteer army in the battle against invasive species.

Formulating Science Questions. We plan to leverage ongoing efforts in “Virtual Telescopes in Education”\(^2\), an NSF/NSDL-funded activity at UMBC [Hoban02a], to guide students in conducting food web research (ultimately ending in a publication - see below). VTIE is developing a tool (SISTER, [Hoban02b]), to guide learners in the formulation of a scientific question leading to the design of their scientific investigation. In the astronomy application, the possible questions are numerous, but the choices are constrained by the types of measurements that can be made with the instrumentation available at the telescopes participating in the program. Similar constraints would bound the set of questions appropriate

\(^2\) [http://vtie.gsfc.nasa.gov](http://vtie.gsfc.nasa.gov)
for learners studying the food web. For example, the students in the 2000 Environmental and Ecological Modeling class taught at SFSU generated food webs with the Niche Model (Williams and Martinez 2000) and compared them to empirical food webs using visualizations. In this investigation, the parameters of the model would limit the questions the students can address. We will broaden SISTER beyond astronomy to address the scientific process generally so that any scientific discipline may use it, and in particular, so that we may use it as an educational tool for teaching biodiversity.

**Reporting Results: Student Journal.** Also leveraging work ongoing in VTIE, we will provide an online student journal, including the tools to contribute to and manage the journal. Once learners have made their proposed measurement, they analyze their data, draw conclusions and report their results. While studies indicate that students readily engage in computer-based educational activities, this final step, reporting results, has been found to be the most difficult to consistently achieve (see e.g. Hoban, Blum and Farrell 2002). We will assist learners in composing a paper based on their experimental findings. Learners will be presented with a dynamically-generated HTML form which contains all of the elements of a typical paper. The general layout for the paper will be provided, focusing the authors on writing rather than presentation. The system provides for the process of reviewing (usually, in this case, review by the teacher or a combination of the teacher and a group of the learner's peers), and, upon successful review, the papers will be published to the online student journal. The combination of Proposal Generation and Paper Writing provides the learners with essential pieces in the end-to-end process of experimental science. The semantic web technologies developed as part of this ITR effort, such as the annotation capabilities, will add the ability for learners to work together on projects, which follows more closely the practice of contemporary scientists.

**Annual Competition.** While we primarily focus on the development of tools for the semantic web, the ultimate test of our success is whether these tools can be adopted by users to solve problems relating to biodiversity and biocomplexity. To this end, we propose to coordinate a series of national competitions based on the use of our tools to solve a problem relating to invasive species. Such an activity will not only serve as a testbed for use of our tools, but also will broaden student involvement beyond our own universities. The premise of the competition is for student teams to develop a collection of resources using our tools that could be used to determine if a recently discovered exotic species is a potential threat. We envision the following framework for the competition:

**Phase 0 (Preface):** Using our tools, we develop a set of resources describing a particular ecosystem and make these available at the Invasive Species Competition website as an example. For example, the WOW (Webs on the Web) resource will be developed to allow ecosystems to be selected according to habitat types (e.g., tropical reef, temperate rainforest, etc) and/or locations (e.g., Amazon, South Africa, etc.). We post the schedule and the rules for the competition at the website. (Winning entries from previous years will be posted as the competition evolves.)

**Phase I (Qualification):** Student teams choose an ecosystem to study. Using our infrastructure creation tools, they develop a set of resources that could be used in the event an exotic species invades their ecosystem. For example, students could specify NBII resources that suggest several plausible candidate species whose geographic expansion suggests it may invade their chosen system and WOW resources that describe ecological interactions of the invader. The students submit their resource library to the Invasive Species Competition Website. The Panel of Judges (chaired by CO-Pi Quinn, Martinez and Schnase, and to be assembled along the lines of a review panel) reviews the entries and determines the top entries which will move forward to the Final Competition (Investigation).

---

3 This work, including the students as co-authors, was presented at the prestigious 2001 Cary Conference on “The Role of Models in Ecosystem Science.”

4 Appropriate measures will be taken to ensure the privacy of the authors since many will be minors.
Phase II (Investigation): Each of the entrants receives a communication which describes a “discovery” of an exotic species in their ecosystem. Using the resources gathered in Phase I, the student teams, using our infrastructure exploitation tools, attempt to determine if the species will pose a threat in their ecosystem, and what can be done about it. They will present their results and their arguments to the panel of judges. The judges will rank the participants.

In year one, we will develop the example resource library and the Invasive Species Competition website. The entire competition will be held over the Web. We will hold the first two competitions (in years two and three) for undergraduates. This will give us time to reach out to the informal education and high school communities. In years four and five we would simultaneously hold competitions for undergraduate level teams, and for high school level teams (through community organizations - e.g. scouts, environmental clubs, etc. will be strongly encouraged to participate). Investigator Hoban, who will coordinate the competition, currently runs a “robot mission to Mars” competition at NASA GSFC, and has had success in getting industry support in the form of cool prizes (palm pilots, printers, etc.). So, we expect there to be incentives for schools to become involved. At a minimum, participants in the Final Competition will receive Invasive Species Competition t-shirts, and members of the winning team will each receive a PDA.

Infusion into Educational Environments. We are well aware that one of the largest obstacles to classroom adoption of new learning technologies is that teachers typically feel bound by their mandated curriculum – there’s simply no room for the new approaches. With this in mind, we have asked
Investigator Sakimoto to oversee the infusion of the education portal capabilities into a classroom setting. Sakimoto (GEST) teaches a graduate course for in-service science teachers at the Johns Hopkins University and will incorporate our proposed framework into her instruction. Creating curricular material out of the portal’s activities will likely be a term project for one or more of her students. All of her students will then serve as a conduit into a larger pool of schools. As well, the project’s developments will be directly incorporated into curricular innovations and classroom activities at both SFSU, and UC Davis. Finally, project members will devise a dissemination plan that includes demonstrations and seminars at national educational conferences, such as the NSTA and NECC.

C.5 Broader impacts on science and education.

We have already discussed the large impact we expect this project to have on education in this country. Through the activities described above, the next generation of science researchers will have early exposure to the next generation of science research tools. What’s more, they will use these tools to construct their own understanding of complex ecological phenomena, such as food webs and invasive species. In addition, our research team is involved in a number of on-campus outreach programs, which we describe in this section.

At the University of Maryland, we will make a special effort to help broaden the participation of under-represented groups by working with the UMBC Meyerhoff Scholarship Program and the Center for Women in Information Technology (CWIT). The Meyerhoff Program, founded in 1988, is a nationally recognized program dedicated to increasing the number of under-represented minorities earning doctorates in the sciences and engineering. Students accepted into the program have exceptional retention rates (95%) and GPAs (3.4) and are broadly distributed in scientific fields. Currently, there are over 200 undergraduates in the program at UMBC. The Center for Women and Information Technology, established at UMBC in 1998, seeks to address and rectify women's under-representation in IT and to enhance our understanding of the relationship between gender and IT. At San Francisco State University, this project will support a Mexican-American co-PI and a woman team member to engage undergraduate students in research activities. SFSU is an inner city non-PhD granting institution; two-thirds of the students are persons of color and almost 60% are women.

C.6 Conclusion

We have assembled a balanced team of computer scientists, biologists and science educators to explore how a new generation of semantic web markup languages will enable powerful tools and application to support biological science. These tools and applications will be integrated into the NBII, allowing both researchers and students to convey meaning through the semantic annotation of web documents; to extract and fuse information from multiple, heterogeneous sources in response to a query; and to mine data distributed across the web. Our project will make significant research contributions to computer science and information systems in developing the concept of the semantic web and exploring its realization through working prototypes of tools and applications. We will make contributions to biological science thought the building of explicit machine understandable models (ontologies) of biological knowledge and the exploration of how these can support science. Finally, we will contribute to education, both formal and informal, by showing how a new generation of web collaboration tools can take advantage of semantic annotations to support science education. We are excited by the opportunities and anxious to begin.

C.7 Results from prior NSF support

The PI and CO-PIs have benefited from many NSF research grants in the past five years which we briefly mention here. Senior investigators hold additional NSF grants which space does not permit us to describe.
PI Finin, along with Joshi and Yesha, has an award from CCR (CCR 0070802) to work on data management and systems issues in mobile/pervasive systems. The project, which is in its second year, has already led to a number of publications and partially supports one MS and two PhD students. Finin and Joshi anticipate an award from IIS for a project that will support one PhD student investigating techniques for trust and security on the Semantic Web.

Co-PI Neo Martinez was awarded two NSF research grants that also provided team member Rich Williams with post-doctoral support: Null and Natural Food Webs: A Critical Investigation of Biological and Methodological Explanations of Food Web Structure (DEB-9905446, $35,000, 5/1/99 to 12/1/00), and Scaling of Network Complexity with Diversity in Food Webs (DEB-0083929, $100,000, 9/15/00 to 8/31/02). Martinez and Williams sponsor team member Jennifer Dunne’s NSF postdoctoral fellowship in biological informatics: Effects of Biodiversity Loss on Complex Communities: A Web-Based Combinatorial Approach (DEB/DBI-0074521, $100,000, 11/1/00 to 10/31/02). At least 7 articles have been published (Journal of Animal Ecology, Nature) or pre-published (Santa Fe Institute Working Papers), with several in peer-review (Proceedings of the National Academy of Sciences, The American Naturalist, Ecology Letters). NSF also awarded Martinez a grant for an Instructional Environmental Science Computer Lab (CCLI-9950461, 6-15-99 to 5-31-01, $60,000 from NSF with $72,630 match from SFSU; Total: $132,630). The grant has helped equip and staff a research and instructional computer lab, and supported curriculum design (e.g., Environmental and Ecological Modeling taught by Martinez and Williams). Williams received support to write 3D food-web visualization software used to produce scientific visualizations published in Nature (Williams and Martinez 2000, Strogatz 2001) and Science (McMahon et al. 2001). In addition, several conference posters with students and outside researchers as co-authors have resulted. Martinez (PI), Williams (Co-PI), and Dunne (Co-PI) have just been granted $1.5M from the NSF BDI program to develop ecoinformatics tools for ecological networks over 3 yr.

CO-PI Quinn has been a co-PI on a number of NSF grants to investigate watershed processes, most recently (827145 and STAR 9815471 with P. Sabatier), and was one of four UC Davis investigators on the original NPACI award (1999, ACI 96-19020, #375,000/yr). Over the last 4 years, 9 ecology and geography students supported by these awards have received PhDs, and the facilities partly supported by the grants host an additional 8-12 graduate students and postdocs at any one time.

CO-PI Schnase was a Co-PI on the National Science Foundation-sponsored Flora of North America, Flora of China, and Biodiversity and Ecosystem Knowledge and Distributed Intelligence projects.

C.8 Team Responsibilities, Management and Coordination, and Work Plan

Management plan. The senior investigators have considerable experience in managing and working in large coordinated projects sponsored by NSF, NIST, USGS, EPA, NASA, the DoD and industry. Many already work together as part of the DAML, NSDL, and NBII programs, and we expect cross-fertilization to be one of the benefits of this activity. On this project, PI Finin will head the management team, which will have regular telecons and periodic meetings (see coordination plan, below). UMBC will prime the project, and be responsible for over-all administrative support and coordination of subcontracting.

Coordination plan. We will hold a project-wide working meeting twice a year to present progress, share results and coordinate activities, alternating between Maryland and California. In addition, we will hold regular conference calls and ongoing workshops focused on specific tasks. We will establish and maintain an extensive web site for the overall project which will serve not only to disseminate information on the project and research results but will also be a source of information of the broader application of semantic web technology to science and education. Since a major research goal of our project is to explore the use of advanced web technology to support collaborations among scientists, we will use our own group as the earliest of adopters, and our web site will be powered by the tools we are building and techniques we advocate. Students will spend time in residence at other universities for a combination of cross-training and coordination.
Team Responsibilities. Summarizing the details from the technical section, the primary responsibilities of each group involved is detailed below. As stated above, all teams will be involved in management, shared tool building and testing, etc.

- **UMBC eBiquity.** Housed in UMBC’s Computer Science Department, eBiquity explores the interactions between mobile, pervasive computing, multi-agent systems, artificial intelligence, and e-services. The lead on this project, eBiquity’s main responsibility will be to build agent-based tools that allow biology researchers to exploit semantic web infrastructure to further their research aims.

- **UMCP MINDSWAP.** Created after co-PI Hendler returned to academia after serving as Program Manager of the DARPA DAML program, MINDSWAP has become the leader in academic semantic web research and the a key developer of open-source Semantic Web tools. The tools that MINDSWAP will explore for this project fall under two broad categories: those used to create semantic web infrastructure (e.g., markup tools), and those that are part of the infrastructure itself (e.g., service composers). This group will also explore use of Semantic Web tools in concert with the Open Grid Service Architecture.

- **UC Davis Information Center for the Environment.** Led by co-PI Quinn, ICE houses the California Information Node of NBII. Through his leadership role in NBII and related international biodiversity initiatives, Quinn is broadly familiar with the shared metadata standards and shared vocabularies, both existing and emerging, in the biocomplexity domain. His team will lead the effort to construct the ontologies, and mappings amongst the ontologies, that will form the foundation of a semantic web for biocomplexity research. Quinn and his team will also spearhead the dissemination of our tools and techniques throughout the biocomplexity research community.

- **SFSU.** Led by Co-PI Martinez, the SFSU team is a leader in both the data and modeling of food webs, and will closely integrate their newly funded large NSF Biological Databases and Informatics project titled, "Webs on the Web: Internet Database, Analysis and Visualization of Ecological Networks" with the project proposed here. The semantic availability of such data, analyses, and visualizations will greatly enhance our ability to predict the probability of exotic invasions as well as their consequences.

- **NASA GSFC.** Although not receiving funding from this proposal, GSFC will be a major partner in our research. Co-PI Schnase heads the Invasive Species Forecasting System, a joint NASA/USGS project that couples high performance computing with algorithm pipelining to predict spatial distributions of invasive species. We will draw on the considerable achievements of the center in constructing and formalizing pipelines for invasive species research, and will provide them with the means of automating the pipeline construction process. As well, we plan to exploit the massive NASA remote sensing datasets which, to date, have not played a significant enough role in invasive species research.

- **UMBC GEST.** A collaborative agreement between UMBC and NASA GSFC, GEST is a center of excellence in Earth Science, as well as technology, in particular, technologies for educational applications. The GEST team will draw on experience with the NASA Learning Technologies LEARNERS initiative and the NSDL VTIE project to build an educational portal into NBII, and to host an annual competition that will engage students throughout the country in semantic web-enabled invasive species research.

**Additional cooperative support from NBII.** If NSF selects this project for funding, NBII will commit $100K/year to support student researchers, and $50K/year to support workshops on ontology formulation, mapping, etc. This will allow us to support several graduate students as well as additional undergraduate student researchers. Our researchers will have accounts on the NBII development server to (i) provide access to NBII data, and (ii) experiment with our prototypes. NBII will also allow us to make
use of their usability testing facilities. To the extent that training is necessary, we will leverage NBII's existing training programs and expertise.

**Work Plan.** The primary scientific work of this project will occur in a tight spiral of design, implement, alpha/beta test, redesign, deploy, evaluate, and disseminate on a roughly quarterly basis. The spiral will start with the tools described throughout this proposal working with the scientists and scientific web sites as discussed. Scientists will be the testers for the tools designed by the implementers, and implementers will train, evaluate and redesign on an ongoing basis. In addition to this constant ongoing spiral development, proposed highlights of the project will include:

**Year 1**
- Start of spiral development using existing tools, first focus on web site support, crawler development, agent infrastructure
- Conversion of metadata standards and vocabularies to OWL
- Design of ontology for ecology and ecological interactions in OWL; testing of ontology management tools for creating/maintaining this ontology
- Determination of services to be offered in version one web site/tools; markup of these with prototype OWL Service descriptions (based on DAML-S tools)
- Creation of the Education Portal: SISTER and on-line student journal
- Identification/coordination of key IT science challenges (learning markup, evaluation metrics, etc.)

**Year 2**
- Spiral development continues as discussed in work plan
- Deployment of OWL ontologies for main software packages (ESRI;S-Plus; etc.) and for WOW project
- Initial roll out of service infrastructure, testing of semantic composition for this domain
- Initial deployment of crawler with agent-based access to results
- Competition at College level introduced
- Initial Evaluations run at NBII usability testing site

**Year 3**
- Spiral development continues as discussed in work plan
- User evaluations on site and service tools
- First deployment in spiral of automated markup tools for biological cites
- Full-scale College Level competition

**Year 4**
- Spiral development continues as discussed in work plan
- High school competition introduced
- Test and refine semantically embedded combined analyses (geographic/ecology interaction)
- Integrate with analytical pipeline technology developed in SEEK project

**Year 5**
- Spiral development ramps down, final dissemination of tools developed through year 5
- Test and refinement of integrated tools for ecological discovery and analysis of potential invasions and ecological impact
- Evaluation and user testing of all components and the integrated web site
- Full-scale High School competition run
References


[DAML-S] DAML Services, DAML-S 0.7 Draft Release, http://www.daml.org/services/daml-s/0.7/


[Finin01] Finin, Tim, Anupam Joshi, Lalana Kagal, Olga Ratsimore, Vlad Korolev, and Harry Chen, Information Agents for Mobile and Embedded Devices, Fifth International Workshop Cooperative Information Agents, September 6-8, 2001 Modena, Italy.


Science on the Semantic Web


[Joshi02] Joshi, K., Joshi, Anupam, Yesha, Y., On Using a Warehouse to Analyze Web Logs, accepted for publication in Distributed and Parallel Databases, 2002.


Science on the Semantic Web


[NRC, 2002]. Predicting Invasions of Nonindigenous Plants and Plant Pests. Committee on the Scientific Basis for Predicting the Invasive Potential of Nonindigenous Plants and Plant Pests in the United States, Board on Agriculture and Natural Resources, National Research Council (in press)


on Challenges in Open Agent Systems, July 2002, University of Bologna, held in conjunction with the 2002 Conference on Autonomous Agents and Multiagent Systems.


