

Semantically Enabling the SEMAT Project: Extending Marine Sensor Networks for Decision Support and Hypothesis Testing.

Trina Myers, Ian Atkinson
e-Research Centre, James Cook University
Queensland, Australia
[Trina.Myers, Ian.Atkinson] @jcu.edu.au

Ron Johnstone
Centre for Marine Studies, University of Queensland
Queensland, Australia
rnje@uq.edu.au

Abstract—The SEMAT project is a multi-institution/multi-discipline program developing advanced wireless sensor networks to collect, store, process and interpret data in coastal systems. The marine environment, specifically coral reefs within the Great Barrier Reef, is one of the initial deployments for a prototype SEMAT network. Wireless sensor networks are being deployed to extract environmental data for research into environmental issues such as climate change, water quality and ecosystem health. Remote monitoring networks in remote marine locations are logistically challenging. However, the interpretation of the complex multidimensional data generated is a problem of at least equal complexity. Application of the semantic tools and methods developed in the Semantic Reef project are being mapped onto the SEMAT use-cases with the goal to develop a data model capable of complex inference, as well as conventional data storage and analysis.

Keywords-ontology; sensor networks;semantic web

I. INTRODUCTION

The SEMAT (Smart Environmental Monitoring and Analysis Technologies) project is developing advanced remote wireless sensor networks designed to collect data in marine environment locations. Selected coral reefs within the Great Barrier Reef (Heron Island) and estuaries along the Queensland coast (Morton Bay) are the target locations for pilot SEMAT sensor network implementations [7]. These prototype networks are being deployed to extract physical data associated with the water column to support research and monitoring into environmental issues such as climate change and water quality. Marine algal blooms are another target of interest. Monitoring remote, marine locations demands the physical sensors and communications be robust, adaptable and smart to maximize the quality and timeliness of data retrieval. It is also essential to minimize cost. All of these factors are being actively investigated as part of the SEMAT program of works.

The SEMAT sensor architecture is developed as an end-to-end sensor network infrastructure that spans the sensor nodes, communications networks (both above and below water), deployment infrastructure, through to the data management, presentation, and analysis software systems. A very wide range of aquatic environments are being targeted in SEMAT. Sensor networks will gather data across the full

spectrum of estuarine, coastal and oceanic environments; including fresh water storages.

As well as a comprehensive approach to data gathering, a particular focus of SEMAT is that it incorporates an integrated systems approach to data storage, analysis and management tools needed to provide synthesised data to ecosystem managers. At present SEMAT is relying on conventional approaches to data management (repositories and relational databases) and data processing. Scientific workflows are being developed to orchestrate the processing and quality assurance of raw data

Independently from SEMAT we have also been developing the Semantic Reef project to create a hypothesis testing tool for use in coral reef ecosystems [11]. This system can be regarded as a Knowledge Base (KB) consisting of a modular ontology hierarchy. The elements of the hierarchy range from informal taxonomies through to formal logic and inference-based ontologies. The architecture is designed to maximise both reusability and usability of ontology's inside the Semantic Reef. The overall goal is to enable marine researchers to pose arbitrary hypotheses about environmental data gathered from *in situ* observations, and to explore phenomena, such as climate change affects, as they relate to the entire ecosystem rather than one component at a time. The system has been substantiated as a predictive tool by being configured to act as an alert of events such as coral bleaching by mimicking common indices with well defined inference rules.

The purpose of this paper is to describe the reconfiguration of the SEMAT integrated environmental monitoring approach to include the semantic and ontological logic from the Semantic Reef. Mapping the ontology's from semantic reef to the SEMAT use cases is a component of this work. However, the primary step is to ensure the workflows that deal with the raw SEMAT sensor flows will tag the data with Semantic information (i.e., URI's) and map them to a Jena triplestore. Data will then be stored in the SEMAT databases but accessible to the triple store via the URI tagging.

Once sensor data is in triple form it can be easily accessed for Semantic queries or mapping to expert systems, such as the Semantic Reef system.

The data available from the SEMAT sensor network will be in the most flexible and accessible form after it is prepared with semantic annotations. Additionally, data will be open and available to researchers worldwide as linked data for semantic queries and mapping to the growing number of ontologies, such as the coral reef ecosystem ontologies of the Semantic Reef model.

The following sections will introduce the SEMAT and Semantic Reef projects. An outline of the benefits of coupling live data to a semantic KB is also presented before the paper concludes with a description of the methods to be employed.

II. SEMAT

A. Background

The SEMAT project was awarded funding under the Queensland Government’s National and International Research Alliances Program, commencing in 2009. The multidisciplinary project incorporates Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP) researchers from the University of Queensland, James Cook University, Queensland Cyber Infrastructure Foundation and the University of Melbourne along with researchers from the Torino Foundation, Polytechnic di Milan and the Danish Hydraulics Group Australia.

The project aims to progress a working proof-of-concept next generation wireless marine sensor network to a commercial prototype that will enable real time monitoring of the marine and other aquatic environments. The system is utilising an integrated, sensor network solution using underwater radio-frequency communication [3, 7]. The project has already generated new links between ISSNIP and other groups associated with the international project partners, as well as with key end-users of such a network system. For example, the SEMAT technology is integral to a new project delivering information into the Asian Knowledge Platform for Climate Adaptation being conducted under the auspices of the United Nations Environment Program (UNEP), Swedish International Development Agency (SIDA) and Stockholm Environment Institute (SEI).

The central aim of the SEMAT project is to build a “smart” wireless sensor architecture that will allow users to apply both their existing hardware and emerging technologies within a multi-scale monitoring system that allows them to interrogate and task aspects of the sensor network according to need and monitoring outcomes. This includes the ability to feed monitoring data into intelligent analytical tools that streamline the synthesis of results and the subsequent management decision-making. Critically, the SEMAT project is prioritising the adaptation of existing technologies where possible, and the use of highly innovative invention so as to deliver the cheapest but most robust and adaptable solution possible. It is intended that the system fill both the need in developed countries, as well as

the more profound knowledge and data collection gap that exists in developing countries globally.

III. SEMANTIC REEF

A. Background

The Semantic Reef project employs Semantic Web (SW) and scientific workflows to resolve data integration and synthesis problems and assist with knowledge discovery for coral reef ecosystems [11, 12]. This project aims to streamline the data analysis of remotely sensed data from the Great Barrier Reef, by balancing the capacity of individual researchers to benefit from the growing volume of data and near-real time data feeds available to them. The Semantic Reef model is a tool for hypothesis-driven research and problem-solving that accesses and infers information, from unconnected datasets and hence the more efficient investigation and synthesis of disparate data streams. An example application is the provision of alerts for unusual domain-specific events or uncovering phenomena in data through coupling disparate data to ecological and environmental ontologies for querying and deriving inferences.

The Semantic Reef Project is a platform for researchers to combine disjoint data into a single KB and to pose questions of the data. Remote sensor data and data available via the Web are retrieved by scientific workflows that integrate the data into a KB for further synthesis and analysis. Any necessary calculations and reformatting are automatically completed by the workflow, which then routes the data to the KB.

The semantic KB consists of a hierarchy of ontologies, created in Protégé [14], that generically describe coral reef ecosystems and the complex relationships within them. The strategy applied in the development of the Semantic Reef ontological architecture was to maximise re-usability and usability, simultaneously, and was accomplished through the separation of concept definitions, instance data and rules (Fig. 1).

The ontology hierarchy contains both re-usable and usable levels. Built using an intra-ontology and inter-ontology methodology [12], where firstly the concepts of a

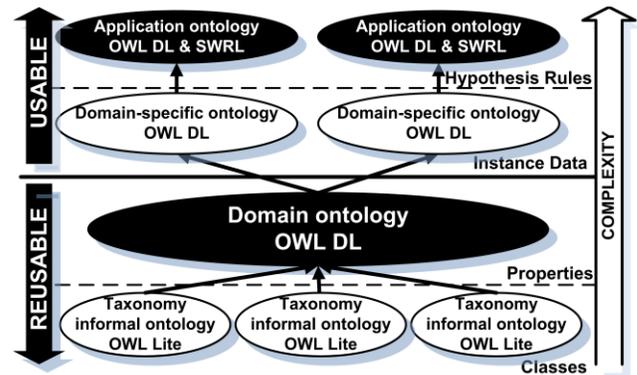


Figure 1. The Semantic Reef knowledge base hierarchical ontology design to maximise re-usability and usability.

coral reef system are defined in a series of separate concepts (“intra” methodology) and then as one holistic concept (“inter” methodology). A formal reef domain ontology, which applies descriptions logics to define the complex relationships between the imported informal taxonomies and ontologies below, was created to generically model any reef system. This level of the KB is the “re-usable” component and independent of the location, type or community mix of a coral reef. At a higher “usable” level, ontologies are developed for explicit applications. First, domain-specific ontologies are created, dependant on the reef of interest, and populated with data specific to that region. Second, at the final application level of the KB, ontologies are created and inference rules designed that pose hypotheses, in syllogistic form, or infer predictive events. Because the instance data is separate from the inference rules the same KB can be re-used for many types of hypotheses. Additionally, as the domain-specific usable ontology imports the coral reef generic ontologies, it can be easily repopulated with newly streamed data.

Scientific workflow technologies are employed to access and process the datasets for mapping to the domain-specific ontologies of the KB. The Semantic Reef architecture uses the Kepler scientific workflow system [1] which is an adaptive software program to capture complex analyses in a flow where data is taken through a series of analytical steps. Within this architecture the workflow collects both raw data from remote sensors as well as existing archives and repositories (e.g. sea surface temperature maps, physical survey data and related data products) by enlisting both web and grid services. Kepler automatically processes the raw data via a series of workflow steps, and passes the results to the KB.

Once the ontologies have been populated by the workflow, the data can be reasoned over and inferences can be made. For example, a domain expert, either a marine scientist or reef manager, can query the KB to extract information of interest, pose observational hypotheses or as an alert system by inferring events.

B. Mapping Live Sensor Data to the Knowledge Base

The design goal here is to use the remotely collected data streams of sensor data from SEMAT to populate the Semantic Reef KB that can be used, as an exemplar, to produce climate change projections for the GBR region.

The capabilities of the Semantic Reef system are presently being tested to infer a coral bleach warning with remotely sensed sea surface temperature data. The alerts relate sea temperature anomalies to the known susceptibility of coral communities to temperature stress. The data used includes spatial data from NOAA [13] and near real time SST data from Davies Reef, Cleveland Bay and Myrmidon Reef monitoring sites of the Australian Institute of Marine Science (AIMS) weather observing system [8]. The data is available via the data access portal of the AIMS data centre and quality assurance processes are in place. If there are gaps

found in the data feed representational data is substituted. A workflow, created to stream the SST data from AIMS, tag it with a URI, then map the data to the Semantic Reef KB. The workflow actors are then directed to populate the domain-specific ontology with the temporal reef instances and assert the designated date/time and temperature data-type property values.

Once the KB is populated with the relevant data Semantic Web Rules Language (SWRL) rules [5], which mimic common bleaching metrics, are applied to predict coral bleaching events. Inference rules are created to infer any instances of a particular reef to belong to a categorised bleach watch class automatically, dependant on the rise in temperature above average and the susceptibility of the coral community to bleaching.

Modern researchers require assistance in streamlining the data processing and analysis phase of research in light of the growing quantities of data. Ideally, scientists are best left to concentrate on the research studies, not on the IT tools they employ to process the data, or the need to develop their own fundamental data management tools to address the large data streams. By corollary, an end-to-end hypothesis system with the capacity to automatically pose questions and derive knowledge from live streaming data has positive implications for many research paradigms.

IV. THE POTENTIAL FOR OBSERVATIONAL HYPOTHESIS TESTING

The sustainable management of coastal and offshore ecosystems requires the collection of accurate data across various temporal and spatial scales. This includes specific measurement and assessment, as well as management related monitoring of processes and influences. Accordingly, remote data collection and monitoring systems are seen as central tools for ecosystem-based environmental management; they provide new information and knowledge on ecosystem function and performance as well as an evidence-based platform for the development and implementation of management strategies and policies. Recent developments in wireless network technology, energy harvesting techniques, and miniaturization now make it possible to realistically assess and monitor the aquatic environment across temporal and spatial scales unachievable historically due to cost, logistics and available technologies. This emergent capability particularly benefits research in remote or dangerous locations, where many fundamental processes have rarely been studied due to physical and economic constraints. In this context, remote underwater sensor networks represent a key instrument able to gather large continuous data sets about many different biological and biophysical parameters (shown in Table I) in key defined environments as well as across ecosystems and operational domains that are yet to be fully explored. This significantly improves our ability to understand ecological phenomena across a much wider spectrum such that issues including

climate change can be more accurately interpreted and understood.

TABLE I. TYPICAL PARAMETERS MEASURED IN THE MARINE ENVIRONMENT

Physical	Chemical	Biological
Temperature	Total Phosphorous	Fluorescence Chl
Salinity	Phosphate	Fluorescence
Conductivity	Nitrogen	Turbidity
Currents	Nitrate/ Nitrite	Yellow substances
Radioactivity	CO ₂	Phytoplankton
Pressure	Ammonia	Zooplankton
Water level	Silicate	Secchi depth
Wave direction	PH	Nutirents
Wave height	Heavy metals	Photosynthetic light
Solar radiation	Hydrocarbons	PAR

Notably, many of the most influential papers in coral reef science of the past few years have been “synthesis” papers, aggregating long-term observations into new hypotheses and conclusions. The range of problems currently facing coral reefs, such as climate change and disease, and the subsequent search for solutions, is necessitating the use of large multi-disciplinary data sets and the synthesis of acquired knowledge. One well known example of this new class of science would be the recent Inter-governmental Panel on Climate Change reports (IPCC) [6]. The IPCC reports are the product of many scientists working together, in an international, intergovernmental, scientific federation, to foster an efficient research paradigm on climate change. Through activities such as IPCC, future research will be conducted by finding correlations in data and through observational hypothetical studies in the ecological domain.

Projects such as the Semantic Reef aim to balance the inevitable growth in data, in particular remotely sensed data, with the capacity of the individual researcher to create new knowledge. The Semantic Reef model assimilates e-Research technologies that each offer solutions to help resolve the data processing and data management bottlenecks. Automating the data analysis processes should free the scientist’s time to work on the research and not be taxed needlessly with gaining access to and/or reformatting the data. Specifically, the automatic integration of disconnected data sources, with different formats (e.g., spreadsheet data, annotation data, live streaming binary data, etc), for inclusion in a hypothesis, would enrich the hypotheses being tested and also free the researcher’s time to continue with other aspects of the study.

Currently in marine research, there has been an explosive increase in the number of questions being posed about the effects of climate change. For example, on coral reef ecosystems there has been an intense interest in elucidating problems such as coral bleaching. In this case, as with many others, the hypotheses include finding the cumulative combination of ecological factors and stressors that contribute to the tipping point from a healthy state to a

degraded or high mortality state; that is, healthy coral to coral mortality due to coral bleaching. This reflects the need in the marine biology domain for hypothesis-driven research tools and problem-solving methods that allow for the efficient investigation of the disparate data streams and data sources that may be available [4].

The Semantic Reef system has the capacity to pose hypotheses and automate inferences of the available data. The system’s design supports flexibility in theoretic hypothesis design (Fig. 2). More specifically, the researcher is not required to predetermine the exact hypothesis prior to data collection and the population of the KB. Rather, the questions can be as flexible as the researcher requires, they may evolve as new data becomes available or as ideas grow and/or epiphanies emerge. Then, once phenomena in the data are disclosed, in situ observations can be performed to confirm or negate the theory. Because of the flexibility offered to the researcher, the Semantic Reef tool is an appropriate method to theorise about a range of scientific conundrums such as the cumulative causal factors that contribute to coral bleaching.

The delayed retrieval of data in current methods of remote data collection (using data loggers) hampers effective observational hypothesis research. *In situ* data collection in marine environments presently entails lengthy delays. This is a common problem given the collection methods being employed (i.e., data loggers). Commonly, the delay intervals can be several months before the data is retrieved for

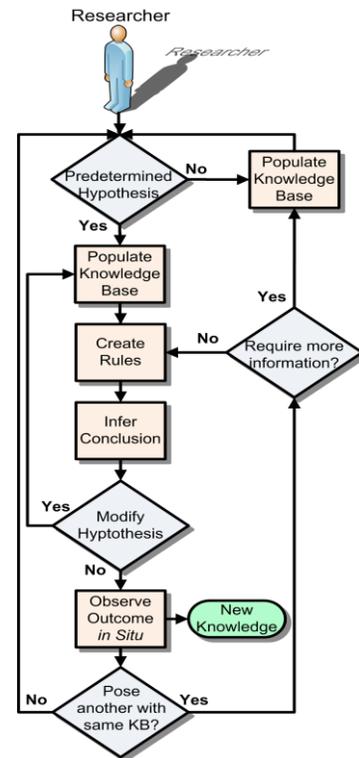


Figure 2. The hypothesis flowchart for posing questions of the Semantic Reef KB. The propositions are fully flexible in light of new ideas or additional interesting data.

processing and analysis, which is not conducive to timely observational hypothesis testing. The need for access to real-time data is leading to the widespread development of real-time marine observation systems. It can be anticipated that a vast amount of sensed environmental data will be available in coming years. A key requirement for effective environmental management will be processing and reducing this data into meaningful information. The value of coupling cutting edge sensor technologies with semantic inference is evident in this context.

The addition of an “adaptation layer” that links the SEMAT remote sensing infrastructure with automated data analysis tools such as the Semantic Reef has great potential for problem solving and the creation of new knowledge.

V. DATA EXTRACTION METHODOLOGY – THE ADAPTATION LAYER

The SEMAT architecture consists of sensor nodes that gather data autonomously and transmit it to one or more network hubs located on the sea surface, which in turn forward the data to a network server (Fig 3). The network hub nodes are positioned on a buoy and equipped with high-speed radio communications to the Internet and hence the network server. The network hub nodes communicate with the sensor nodes directly, via an underwater RF channel. The network hubs (represented by buoys) allow much richer network connectivity, creating multiple data collection points for the underwater RF network [7].

The adaptation layer of the SEMAT architecture aims to couple the live sensor data to linked data platforms and storage repositories while maximising efficient accessibility, analysis and processing functions.

A. The Adaptation Layer Workflow

On arrival to the sensor network server, semantic annotations are applied to the data. The procedures and standards for the addition of semantic annotations, metadata and ontologies to sensor applications and deployments are being developed by an open information space called the Semantic Sensor Web (SSW) [9]. The SSW project proposes to annotate sensor data with semantic metadata to provide contextual information for situational awareness. The aim is to incorporate W3C and Open Geospatial Consortium (OGC) standardisation and extend them with SW technologies to create a SSW, providing an environment for improved query and reasoning in a sensor domain [15]. Computer-readable meaning is enabled by adding semantic annotations to existing standard sensor languages of the Semantic Web Enablement (SWE) specification, such as Sensor Model Language (SML) and the Sensor Observation Service (SOS).

Workflows are employed to automate the annotation process by adding URI’s to the data. Then, represented in RDF format, the data can be ported to the Jena triplestore [2]. Jena is one of the most widely used Java APIs for RDF and OWL and still implemented as a basis for the Protégé-OWL framework. Jena provides services for model representation, parsing, database persistence, querying and some visualization tools [10].

Data presentation and storage is initially in both SQL relational database and triplestore platforms.

Ontologies, employed via the workflows, will reside at a higher layer for data integration. Here, there are a number of appropriate predefined ontologies available for implementation including: a common sensor ontology, a geospatial ontology and a time ontology. These ontologies are being developed for the Sensor Standards Harmonization endeavour, which involves organisations such as the US

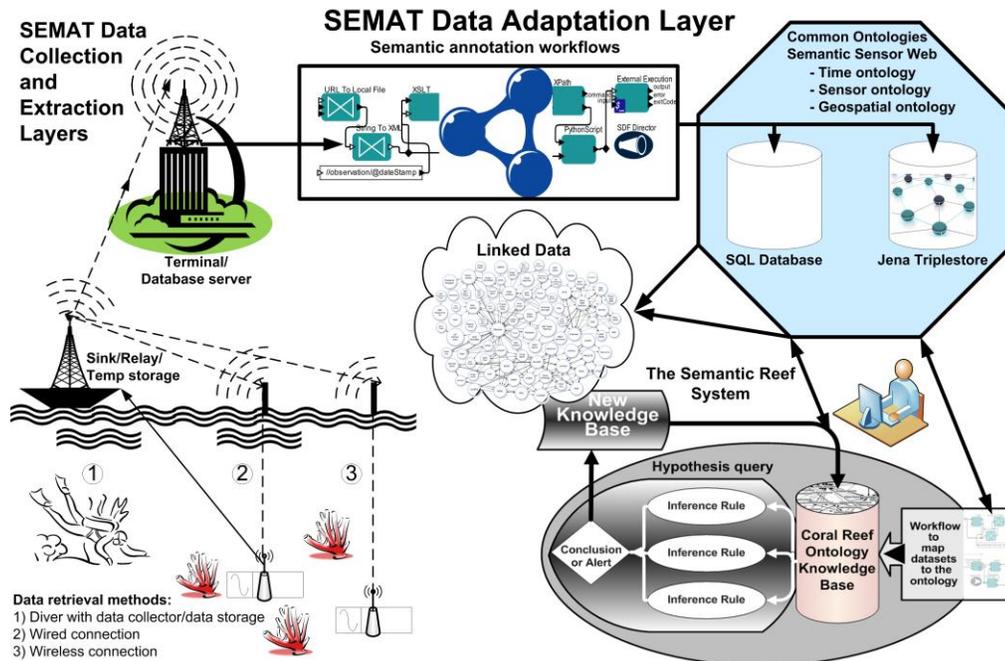


Figure 3. The SEMAT architecture with the adaptation layer.

National Institute of Standards and Technology , W3C and OGC [15].

Once in triple form, the data is accessible as linked data for query and inference by many applications. The triples can be mapped directly to the Semantic Reef KB via the scientific workflows, in near real time for analysis and inference.

VI. CONCLUSION

The Semantic Reef project is a model for hypothesis testing on coral reef ecosystems and SEMAT is an end-to-end data collection and presentation infrastructure. The coupling of these two developing initiatives presents a platform where marine researchers will be able to pose observational hypotheses over a rich range of environmental and anthropogenic factors in real-time.

The SEMAT sensor infrastructure is developing new technologies to handle the data collection and fusion in marine environments. The streamed real-time data, which is the product of the above- and underwater RF sensor networks, include not only the commonly sensed data types such as sea temperatures, but also many other important parameters used in marine research, from the microbiological level through to the reef geomorphology.

The data adaptation layer the SEMAT hierarchy creates an abridgeable semantic portal for linked data access, retrieval, synthesis and integration. At this layer the data is semantically enabled through automated annotation via the SSW emerging standards. The outcome of this process will be a resource available for mapping to the Semantic Reef KB for real-time observational hypotheses and predictions for coral reef research.

The initial coupling of a KB with live data from the SEMAT project will create a proof of concept focused on the marine science use case. However, the infrastructure of SEMAT, once trialled will be adaptable to many other research disciplines. A redefinition of the ontologies in the KB is all that would be required for an alternative data representation. For example, an entirely different domain such as terrestrial ecology would require many new domain ontologies to define the system. However, common ontologies would remain the same (time, space, etc). Alternatively, in the marine science domain modular ontologies can simply be ported to the KB at either the domain "reusable" ontology level or at the domain-specific "usable" application level. Accordingly, an ontology to describe the concept of a benthic system could easily be added, through importation, to the Semantic Reef KB to better describe the relationships within data.

ACKNOWLEDGMENT

The authors wish to thank the Queensland Government NIRAP, ISSNIP, QCIF, DHI, University of Queensland and James Cook University.

REFERENCES

- [1] I. M. Atkinson, D. du Boulay, C. Chee, K. Chiu, P. Coddington, A. Gerson, et al., "Developing CIMA-based cyberinfrastructure for remote access to scientific instruments and collaborative e-research," Proc. Proceedings of the fifth Australasian Symposium on ACSW frontiers, Ballarat, Australia, (2007).
- [2] D. Brickley, R. V. Guha, and B. McBride, RDF Vocabulary Description Language 1.0: RDF Schema, <http://www.w3.org/TR/rdf-schema/>.
- [3] U. M. Cella, N. Shuley, and R. Johnstone, "Wireless sensor networks in coastal marine environments: a case study outcome," Presented at the The Fourth ACM International Workshop on UnderWater Networks (WUWNet) California, USA (2009).
- [4] T. Hey and A. E. Trefethen, The Data Deluge: An e-Science Perspective, in: Grid Computing - Making the Global Infrastructure a Reality, Berman F, Fox GC, and Hey AJG, Eds.: Wiley and Sons, 2003, pp. 809-824.
- [5] I. Horrocks, P. F. Patel-Schneider, H. Boley, S. Tabet, B. Grosz, and M. Dean, SWRL: A Semantic Web Rule Language - Combining OWL and RuleML, <http://www.w3.org/Submission/SWRL/>.
- [6] IPCC, Intergovernmental Panel on Climate Change, Climate Change 2007: The physical basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge Univ. Press, Cambridge, U. K. 2007.
- [7] R. Johnstone, D. Caputo, U. Cella, A. Gandelli, C. Alippi, F. Grimaccia, et al., "Smart Environmental Measurement & Analysis Technologies (SEMAT): Wireless sensor networks in the marine environment," Presented at the ICT-Mobile Summit, Stockholm, Sweden, (2008).
- [8] S. Kininmonth, S. Bainbridge, I. Atkinson, E. Gilla, L. Barrald, and R. Vidaude, "Sensor networking the Great Barrier Reef," Spatial Sciences Qld. , (2004). pp. 34-38.
- [9] Kno.e.sis.Centre, Semantic Sensor Web, http://knoesis.org/research/semsci/application_domain/sem_sensor/.
- [10] B. McBride, "Jena: A semantic web toolkit," IEEE Journal of Internet Computing, vol. 6(6), (2002). pp. 55-59.
- [11] T. S. Myers, I. Atkinson, and R. Johnstone, "Supporting coral reef ecosystems research through modelling re-usable ontologies," Proc. Knowledge Representation Ontology Workshop (KROW 2008), Sydney, Australia, (2008).
- [12] T. S. Myers, I. M. Atkinson, and R. Johnstone, "Supporting coral reef ecosystems research through modelling a re-usable ontology framework (in press)," Appl. Artif. Intell., vol. 90(24), (2009).
- [13] NOAA, Coral Reef Watch, <http://coralreefwatch.noaa.gov/satellite/index.html>.
- [14] Protégé, The Ontology Editor and Knowledge Acquisition System, <http://protege.stanford.edu/>.
- [15] A. Sheth, C. Henson, and S. S. Sahoo, "Semantic Sensor Web," IEEE Internet Comput, vol. 12(4), (2008). pp. 78-83.