1-7-2003

Biodiversity and Ecosystem Informatics: Event and Process Tagging for Information Integration for the International Gulf of Main Watershed

Mary-Kate Beard-Tisdale
Co-Principal Investigator; University of Maine, Orono, beard@spatial.maine.edu

Neal Pettigrew
Co-Principal Investigator; University of Maine, Orono, nealp@maine.edu

Jeffrey Kahl
Co-Principal Investigator; University of Maine, Orono

Malcolm L. Hunter Jr.
Co-Principal Investigator; University of Maine, Orono, mhunter@maine.edu

Marilyn Lutz
Co-Principal Investigator; University of Maine, Orono

Final Report for Period: 10/2001 - 12/2002
Submitted on: 01/07/2003
Principal Investigator: Beard-Tisdale, Mary-Kate .
Award ID: 0131912
Organization: University of Maine
Title:
Biodiversity and Ecosystem Informatics: Event and Process Tagging for Information Integration for the International Gulf of Main Watershed

Project Participants

Senior Personnel

Name: Beard-Tisdale, Mary-Kate
Worked for more than 160 Hours: Yes
Contribution to Project:

Name: Pettigrew, Neal
Worked for more than 160 Hours: Yes
Contribution to Project:

Name: Kahl, Jeffrey
Worked for more than 160 Hours: Yes
Contribution to Project:

Name: Hunter, Malcolm
Worked for more than 160 Hours: Yes
Contribution to Project:

Name: Lutz, Marilyn
Worked for more than 160 Hours: Yes
Contribution to Project:

Post-doc

Graduate Student

Name: Watson, Stephanie
Worked for more than 160 Hours: Yes
Contribution to Project:
Stephanie Watson helped coordinate and organize the workshop, developed the project webpage, helped develop the on-line catalog, and developed the test ontology. She was supported by this project.

Name: Woo, Seung-Ki
Worked for more than 160 Hours: Yes
Contribution to Project:
Seung-ki Woo developed an on-line prototype event viewer to view events once they have been identified. He was funded by this project.

Name: Harvey, Christine
Worked for more than 160 Hours: Yes
Contribution to Project:
Christine Harvey helped with workshop organization, on the development of the digital catalog and on development of the lobster ontology. She was funded by a university fellowship.

Undergraduate Student

Technician, Programmer

Other Participant

Research Experience for Undergraduates

Organizational Partners

Other Collaborators or Contacts

The following list includes the list of participants for the workshop

Name Affiliation Expertise
Seth Barker Maine Department of Marine Resources Data Management
Kate Beard Dept. of Spatial Information Science and Engineering, University of Maine Digital Libraries, Spatial Databases
Judy Blake Jackson Laboratory Information Management, Molecular Biology
Robert Branton Department of Fisheries and Oceans, Environment Canada Biological assessment
Ken Foote Woods Hole Oceanographic Institution Physical Oceanography, Engineering
Merry Gallagher Maine Department of Inland Fish and Wildlife Fish habitat, fish biology
Kathleen Hornsby Dept. of Spatial Information Science and Engineering, University of Maine Spatio-temporal databases
Malcolm Hunter Department of Wildlife Ecology, University of Maine Biodiversity
Steve Kahl Senator George Mitchell Center for Watershed Research Freshwater Geochemistry
Marilyn Lutz Fogler Library, University of Maine Cataloging Scientific Collections, Digital Libraries
William Mackaness Dept. of Spatial Information Science and Engineering, University of Maine Visualization of spatial information, mobile devices
Deirdre Mageean Margaret Chase Smith Center for Public Policy Demography, Environmental Policy
Nathan Michaud Island Institute Community Planning
Robert Morris Computer Science, University of Massachusetts at Boston Web-based distributed databases
Judith Pederson Sea Grant, Massachusetts Institute of Technology Marine science and policy
Neal Pettigrew University of Maine, School of Marine Sciences Physical Oceanography
Robert Stevenson Department of Biology, University of Massachusetts at Boston Environmental informatics/visualization
Thomas Trott Suffolk University Benthic Ecology
Lou Van Guelpen Huntsman Marine Science Center, Department of Fisheries and Oceans, Environment Canada Biodiversity
Peter Vaux Maine Aquatic Biodiversity Program Limnology, aquatic ecology
Will Wollheim Water Systems Analysis Group, University of New Hampshire Hydrology and water quality modeling
Greg Zielinski Maine State Climatologist and Research Associate Professor of Quaternary and Climate Studies, University of Maine Climatology

Activities and Findings

Research and Education Activities: (See PDF version submitted by PI at the end of the report)

Findings: (See PDF version submitted by PI at the end of the report)

Training and Development:

Two of the graduate students participating on the project attended the workshop and participated in the discussions. The discussion of proposal preparation at the workshop and subsequent to it provided them with exposure and first hand experience in strategies for research proposal preparation. One of the gradaute students futher assisted with research project reporting.

Outreach Activities:

Journal Publications

Books or Other One-time Publications

Web/Internet Site

URL(s):

www.spatial.maine.edu/~bdei

Description:

The Web site was created as a form of communication between participants in the project as well as a public information piece. The web site includes:
À A description of the planning grant and workshop
À Workshop agenda (PDF format)
À White paper on events and processes (PDF format)
À White paper on ontologies (PDF format)
À Presentations from the workshop (PDF format): Dr. BeardÆs Workshop Introduction, Dr. Judith BlakeÆs Gene Ontology Presentation, and Dr. Neal PettigrewÆs Presentation on links between atmospheric and oceanographic events and processes over multiple spatial and temporal scales
À Workshop Report (PDF format)
À Links to Related Web sites
À An XML digital catalog of scientific resources for the Gulf of Maine
À An on-line questionnaire to identify issues related to data access and integration

Other Specific Products

Product Type: Data or databases

Product Description:

An XML document/digital catalog of scientific resources for the Gulf of Maine was created. The XML syntax can be queried based on the content-based tags used, such as <Spatial_coverage>Maine</Spatial_coverage>. The digital catalog includes website descriptions (general description, URL, spatial coverage, temporal coverage, monitoring frequency, political area; the latter for human impacts data only) for websites categorized as ôenvironmentalö or ôhuman impactsö. The environmental category includes websites related to biological, chemical, physical or other information. It is further categorized by whether the website is a searchable database, has raw data, or is a list of links to other resources. The human impacts category includes websites related to legislation, regulation, population or other information. It is also further categorized by whether the website is a searchable database, has raw data, or is a list of links to other resources.

Sharing Information:

The digital catalog is published on line (accessible via link from the www.spatial.maine.edu/~bdei website). The catalog contains active links to the websites listed within. Programming is currently underway to provide a query interface to obtain websites matching certain criteria (i.e., by the tags in the XML document), such as ôSpatial Coverage = Gulf of Maineö and ôTemporal Coverage=1997-2002ö so that any researchers can easily query the catalog

Product Type: Ontology

Product Description:

An ontology was developed for environmental events and processes related to the American lobster (Homarus americanus). The ontology is the first step in the development of a larger event and process ontology. The ontology was developed in the DAML+OIL language, with OILED 3.4 and FACT open source software.

The ontology builds on three primary concepts: object, event, and process. The initial role of the ontology is to formalize the definitions of an event, process and object in a machine-readable way. The object class includes taxonomic information about the American lobster as well as information about geographic features (terrestrial and hydrographic). The event class includes the subclasses, biological event, chemical event, physical event, and other event. The process class includes a similar set of subclasses. The ontology does, for example, define, ôwhat is an event?ö and ôwhat is a process?ö

An additional role of the ontology is to define and formalize the relationships between the primary concepts û object, event, and process. This is a key role of the ontology in that it, in information space, provides the formal connections between concepts that are actually linked in the real world.

Sharing Information:

The ontology will continue to be developed. The goal is to make the ontology queryable and to use it for environmental event and process detection across the Web. At present, the World Wide Web Consortium (W3C) has yet to recommend a standard query language for ontologies developed in DAML+OIL. Once that recommendation is made, the query functions will be implemented.

Contributions

Contributions within Discipline:

Currently events and processes are only implicitly represented in scientific repositories and there are few explicit links between different scienctific information repositories.
This project raised the awareness of the importance of connecting environmental events and processes and the advantages of being able to explicitly identify events and processes within scientific repositories such that they are searchable entities.

Contributions to Other Disciplines:

The library and information science fields have not dealt substantially with environmental databases. Information retrieval from environmental databases and particularly retrival and presentation of spatio-temporal based information requires some different strategies from traditional text based retrieval and information display. By working with representatives from the libary community we have started to build a foundation for improving access mechanisms to diverse enrionmental data repositories.

Contributions to Human Resource Development:

Contributions to Resources for Research and Education:

Contributions Beyond Science and Engineering:

Categories for which nothing is reported:

Organizational Partners

Activities and Findings: Any Outreach Activities

Any Journal

Any Book

Contributions: To Any Human Resource Development

Contributions: To Any Resources for Research and Education

Contributions: To Any Beyond Science and Engineering

Findings

Workshop Outcomes

We had hoped that the workshop might result in some consensus on definitions for process and event. However this was not the case. We received comments on the straw definitions (Event:A change in the status quo over a space-time interval; Process:A temporal sequence of changes that transform something). When working groups reported, it was clear that the definitions would continue to evolve but that the distinction between events and processes was useful. The groups' discussions resulted in an initial set of important processes and events in the Gulf of Maine watershed.

Comments ranged from general agreement to:

- How do we define the status quo in dynamic systems?
- The term "sequence" may excessively limit what we can categorize as a process.

Further comments on events and processes and their explicit representations included:

- An event is defined in space and time.
- An event is a signal. A process responds to an event.
- An event might be the result of a process that has yet to be identified.
- An event has time associated with it, regardless of whether the event is discrete or not.
- An event has a short time scale relative to the time scale of the process of which it is part.
- An event is not ongoing, nor is it periodic.
- An event has a beginning and an end. Processes can be described as independent of a time element.
- Time scale and cultural (i.e., value sets) perspectives influence whether you define something as a process or an event.
- Events could be turning points, extremes etc. within a process. Processes are punctuated by events -e.g., maximum and a minimum
- An event has an associated magnitude or intensity

The working groups also defined events and processes by example. The following list includes example events and processes on which there was general agreement that the phenomena was clearly an event or a process.

Events:

Storms
Spawning of an eddy from a current
Phytoplankton bloom (red tide)
Larval settlement
Spawning
Hatching
Oil spill
Site development
Storm runoff/spring melt
Spawning event

Species invasion

Processes:

Seasonal heating and cooling
Tidal mixing
Ocean circulation
Grazing
Life history of an organism
Sprawl
Sedimentation
Riverine input
Erosion
Predation
Herbivory
Eutrophication

A number of phenomena were not to be easily distinguishable as events or processes. Examples given included:

Migration
Stratification
Extinction

This blurring in some cases between events and processes could be an issue of scale or a matter of perspective.

With respect to the scale issue, working groups identified large, meso, and small scale process and events primarily by example. Examples of large scale processes and their connections to other scale processes and events included:

Climatology/Meterology
Insolation (solar irradiance) - seasonal to annual to decadal influence, affected by volcanic eruption, greenhouse gases, dust from the continents
North Atlantic Oscillation - annual to decadal influence
Jet Steam - variations from seasons to years
Weather - atmospheric conditions (some are spatially extensive but temporally brief) They affect sunlight amount.

Oceanography
Differences in input waters to gulf of Maine through Northeast channel - seasons to years
Gulf Stream meanders (generate eddies) - seasonal to annual influence
Freshwater inputs - snowpack and riverine input; sea ice and snowpack from other basins – seasons to years

Economic Activity
Marine transport - introduced species (decades to centuries), oil spills, collisions
Fisheries
Land-based activities - agriculture and forestry (seasons to decades), pollution, coastal settlements (seasonal to decadal and longer)

Policy, regulation and governance
e.g., fisheries regulations, environmental controls – operate on many spatial and temporal scales

A few processes were identified as meso-scale, such as:

Current location
Grazing
Population dynamics
Human sprawl
Riverine input
Fishing
Perdition
Transportation
Contamination

Some of these such as riverine inputs overlapped with large scale processes, so some blurring of scale ranges was apparent.

Fine-scale events and processes were defined in terms of their spatial and temporal constraints. In most cases it was observed that events and processes that occur at fine spatial scales also occur at fine temporal scales or vice versa. One counter example given for the Gulf of Maine was Minas Basin. In this case the fine temporal scale of the tidal fluctuation actually exposes a very large spatial area at low tide. Another observation was that one spatial dimension of an event or process can be quite constrained while others are not. For example some processes or events may have a wide horizontal extent but little depth. Another example offered was bird migrations where the migration is constrained to a narrow path while the migration distance is quite long. The group also discussed that in some case the onset of an event can be instantaneous but that the effect is persistent. Examples given were pollution spills and incidents of introduced species. The following table provides example of fine scale events and processes along with processes or events that constrain them spatially or temporally and whether the constraint is predominantly spatial or temporal.

Events/processes	**Duration**	**Constraints**	**Temporal/Spatial?**
Tidal fluctuation	Daily		T
Vertical migration	Daily	Swimming, currents, light	T/s
Pollution	Variable		T/s
Larval dispersal/retention	Weeks	Hydrography	S
Introduced species	Instantly-persistent		S
Floods	Hours/weeks	Floodplain	T/s
Storms	Hours/days		T/s
Nutrient mixing	Variable	Upwelling/frontal mixing	S
Fishing activity	Variable	Where the fish are	T/s
Hypolimnic	Seasonal	Temperature/rainfall, wind,	T/s

stratification		oxygen demand	
Feeding concentrations		Zooplankton patching, stratifications	T/s
Spawning/nesting	Hours/weeks	Suitable habitat, food, temperature, light	T/s
Migration	Seasonal	Restricted spatially, swimming speed, currents	S
Zonation	Seasonal	Light, competition, predation, substrate	S
Blooms	Seasonal	Day length, mixing, temperature	T/s

Datasets for which event and process ontologies would be useful were also identified. Mechanisms for monitoring and assessing human impacts (including the synergistic effects of the impacts and the appropriate spatial and temporal scale for monitoring) were also discussed as important to include in ontologies of events and processes.

One conclusion was that definition of events will be subject to multiple interpretations by different researchers and explicit detection and recording of events within datasets will be based on different criteria. An example discussed was drought events. Multiple definitions of droughts exist thus any record of a drought stored in a database of drought events should be associated with the defining criteria.

Conclusions regarding the test ontology development

1. ontologies provide an effective means of formalizing environmental concepts and their inter-relationships in a machine-readable way
2. the ontological formalization, and implementation, of important environmental events and processes will likely improve access to, and integration of, data.
3. ontology development tools founded in Description Logics (which provides support for reasoning among concepts) are still in their infancy.
4. The need exists to develop ontology language extensions and tools for defining the relationships and rules necessary to represent spatial-temporal data (e.g., environmental events and processes). For example, the DAML+OIL language could be extended to include the eight topological relationships between spatial objects (Egenhofer and Franzosa, 1991) and the interval relations defined by Allen (1983).

Allen, J. 1983. Maintaining Knowledge about Temporal Intervals. Communications of the ACM. 26(11) 832-843.

Egenhofer, M. and Franzosa, R. 1991. Point-set Topological Spatial Relations, *International Journal of Geographical Information Systems,* 5 (2): 161-174.

Activities

The objectives of this project were to:

- hold a workshop to establish partnerships among future collaborators (a diverse group of researchers, agency representatives, and non-governmental organizations)
- explore the feasibility for integrating scientific information through ontologies of events and processes
- develop a proof of concept prototype for event and process tagged information.

Workshop

A day-and-a-half workshop was held June 13-14 2002 with representatives from multiple disciplines and multiple institutions from among the states and provinces around the Gulf of Maine The workshop involved a series of discussions on the integration of scientific data sets collected for the Gulf of Maine watershed. Presentations were made to introduce participants to focal issues. After the presentations, participants broke into working groups to address specific questions related to events and processes.

Dr. Kate Beard opened with a discussion of the workshop goals and objectives, which were to:

- discuss methods for integrating diverse datasets (varying in completeness and geographical extent),
- begin development of an ontology of events and processes (both environmental and human induced),
- explore the interaction of processes at various spatial and temporal scales, and
- identify mechanisms for ongoing collaboration.

As background and context for the workshop, Dr. Judy Blake from the Jackson Laboratory presented work on the development of a molecular biology ontology as a potential model for the development of an environmental event and process ontology for the Gulf of Maine watershed.

Dr. Steve Kahl introduced the first working group assignment which was to define *events* and *processes* and consider whether a distinction between events and processes was useful to the development of an ontology for Gulf of Maine watershed data. The discussion was seeded with straw definitions for *event* and *process*, as follows: a) *event* - a change in the status quo over a space-time interval, b) *process*- a temporal sequence of changes that transform something.

Dr. Neal Pettigrew opened the second round of working groups by introducing the concept of processes and events operating over, and interacting between, different spatial and temporal scales. Dr. Pettigrew discussed this idea in the context of the North Atlantic Oscillation and how its position was affected by climatic conditions. He demonstrated how the oscillations impacted smaller scale processes – eventually linking the large-scale oscillations to the fine-scale process of lobster shedding in an estuary. The second working group assignment was to consider similar large, meso, and fine-scale events and processes operating within the Gulf of Maine watershed and to examine the interdependencies. The working groups identified important events and processes operating over different scales and began to describe their interactions.

The third round of working groups focused on the interactions between human (anthropogenic) events and processes and environmental events and processes in the Gulf of Maine watershed. Dr. Malcolm Hunter justified the separation between human and environmental processes by observing that human processes were premeditated, often with partial knowledge of their outcome, and that, as humans, we are responsible for our actions. The groups' discussions focused on characterizing anthropogenic events and processes and examining their typical scales of impact.

The workshop closed with a discussion on a future collaborative research proposal. It was suggested that any proposal should include a two-tier structure of participation: 1) a broad group with representation from relevant organizations, agencies and research institutions, and 2) a smaller technical group focusing on issues of data integration through ontologies, as well as methods of integrating distributed databases. The larger group could include representation from State and Provincial Government (e.g., Departments of Marine Resources, Environmental Protection, Conservation, Wildlife, and Planning), the Gulf of Maine Council (especially the Data and Information Management Committee), GOMINFOEX, Regional Association for Research in the Gulf of Maine (RARGOM), Canadian Researchers, Northeastern Governors and Eastern Canadian Premiers, Gulf of Maine Ocean Observing System (GoMOOS), Census of Marine Life, Gulf of Maine Aquarium, Gulf of Maine Educators, Center of Marine Biodiversity, industry representation, Conservation Law Foundation, Conservation Council of New Brunswick, CEC (NAFTA), and more inland representation.

The smaller technical group will be defined by the Principal Investigators. The small group would develop an example ontology to be included in the collaborative proposal. The participants identified 1) lobster landings; and 2) Atlantic salmon population and distribution as two compelling issues in the Gulf of Maine watershed. Plentiful data exist for both issues, but virtually no data integration has being accomplished, thus making them ideal candidates.

Ontology Development

As a follow on from the workshop, an ontology was developed for environmental events and processes related to the American lobster (*Homarus americanus*). The ontology is the first step in the development of a larger event and process ontology. The ontology was developed in the DAML+OIL language, with OILED 3.4 and FACT open source software.

The ontology builds on three primary concepts: *object, event*, and *process*. The initial role of the ontology is to formalize the definitions of an event, process and object in a machine-readable way. The object class includes taxonomic information about the American lobster as well as information about geographic features (terrestrial and hydrographic). The event class includes the subclasses, biological event, chemical event, physical event, and other event. The process class includes a similar set of subclasses. The ontology does, for example, define, “what is an event?” and “what is a process?”

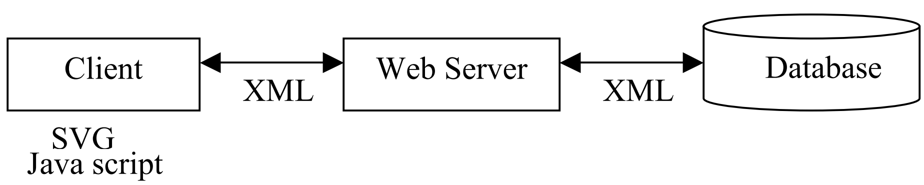
The second role of the ontology is to define (also in a machine-readable way) and formalize the relationships between the primary concepts – object, event, and process. This is a key role for the ontology in that it provides the formal connections between concepts that are actually linked in the real world.

The third role of the ontology is to provide explicit criteria for detection of events or processes within raw datasets based on the definitions of concepts and their relationships. Such criteria could be used for designing data mining algorithms for event or process detection.

Prototype Event Viewer
A web-based prototype was developed to demonstrate presentation of identified events. The operation of the viewer assumes that events identified from various scientific data sources have been explicitly represented in a database. An XML model of weather events was developed along with a small test database and an associated SVG based viewer accessible at (http://acsm.spatial.maine.edu/kai/casco2.asp)

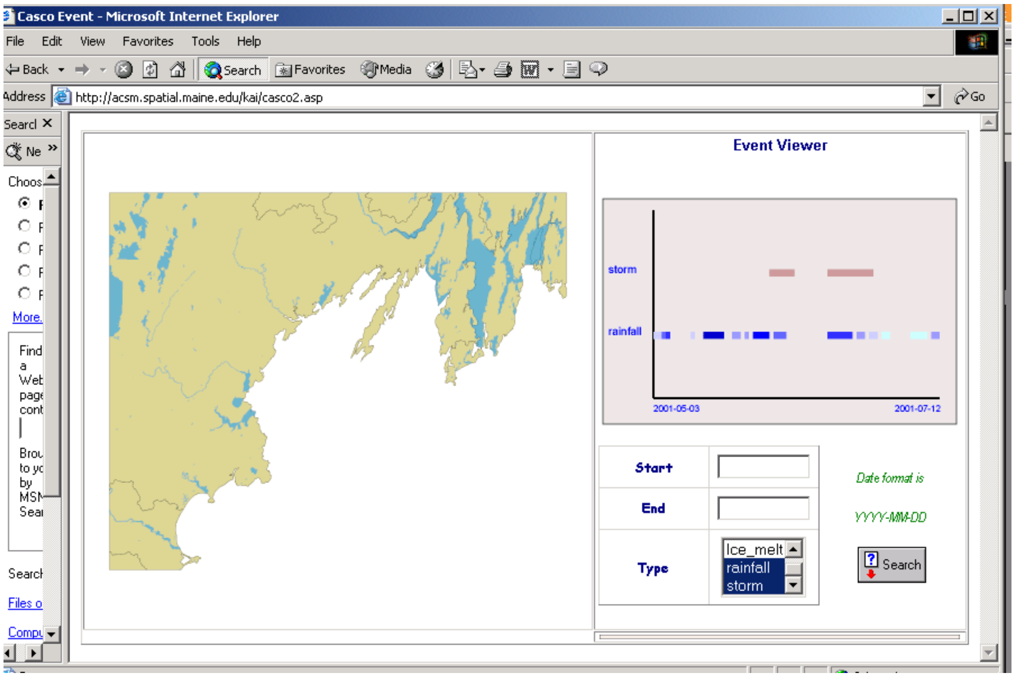
The model represents two types of events (static and dynamic) and the viewer was designed to display different spatial temporal behaviors of events. Under the model events have a unique identifier, type, a start and end time, or duration, a spatial location, and magnitude or intensity. Dynamic events can have multiple time referenced locations and varying intensity levels.

The architecture of the prototype is as shown below.

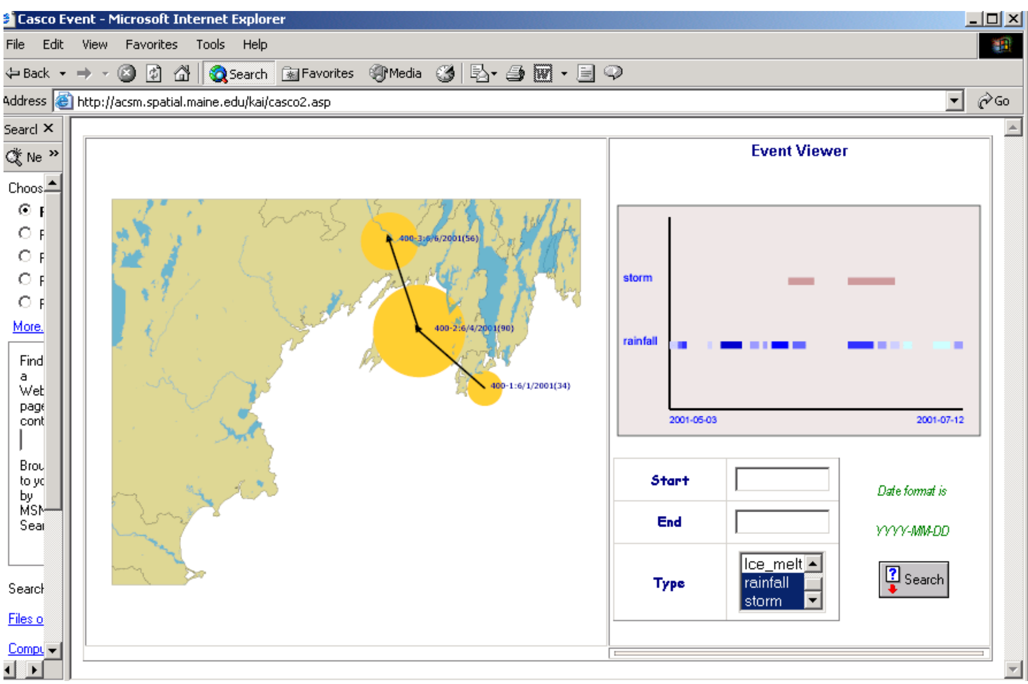


In the webserver, XML documents are parsed and queries are processed by ASP. Data retrieved from the database in response to a query are wrapped in XML and sent to client application. The returned XML documents are parsed by the client and transformed into SVG by javascript. SVG is displayed by an SVG viewer plugin. SVG, an XML grammar describes two-dimensional graphics and supports complex scalable graphics in a highly compressed format. As an XML embedded language it provides hyper links on vectors and line groups through scripting or programming languages.

The viewer allows users to query with start date, end date or event type. Query results are displayed in a time chart initially. A user can select multiple events types and each type is displayed by symbolized lines in the time chart. After query results are displayed, detailed information about a specific event is displayed when a user moves the cursor over a line representing an event in the time chart. Individual events can then be displayed on a map to represent their spatial locations. Screenshots of events displayed in the viewer are shown below.



Selection and representation of multiple events in the time chart.



Representation of the progression and intensity of a storm in the event viewer.

Presentations:
Event And Process Tagging For Information Integration for the International Gulf of Maine Watershed, Digital Government Conference 2002. Redondo Beach, CA. May 2002.

Environmental Process and Event Relations, ACTOR Workshop, Lucerne, Maine, November 3, 2002