Task-oriented digital library interactions: An end-user perspective on interoperability

Tamara Sumner
University of Colorado at Boulder and DLESE Program Center
sumner@colorado.edu

Mary Marlino
DLESE Program Center, University Corporation for Atmospheric Research (UCAR)
marlino@ucar.edu

Terence Smith
Alexandria Digital Earth Prototype, University of California at Santa Barbara
smithtr@cs.ucsb.edu

Extended Abstract

Over the past two years, the Earth system education community has organized and begun development of the Digital Library for Earth System Education (DLESE). This community effort has it roots in a number of educational, social, and scientific priorities, including:

- the public demand for improved science, mathematics, engineering, and technology education at all levels;
- the need for specific training for instructors and students to effectively use scientific information, methods, and tools;
- the value added to the research enterprise by translating new discoveries, information, data sets, and images about the Earth into effective instructional activities;
- the public's need for access to reliable information about the Earth to inform personal and societal decisions.

In August of 1999, under the sponsorship of NSF and NASA, representatives of the Earth System education community came together to institute a governance system and an operational arm (DLESE Program Center) to support the design and development of a community-sponsored and community-owned digital library, dedicated to offering high-quality electronic resources that foster active, inquiry-based learning about the Earth at all educational levels. When fully operational, DLESE will offer peer-reviewed teaching and learning resources (e.g., simulations, models, maps, images, text, data sets, etc.), interfaces and tools to allow student exploration of Earth data, services to help users effectively create and use educational resources, and a community center to facilitate sharing, collaboration, and excellence in Earth systems education. DLESE users include learners and instructors in all venues, many of whom are also resource contributors, developing educational materials, providing scientific knowledge, and evaluating DLESE holdings. To date, significant progress has been made on many aspects of the library building process: the community has been organized, a governance structure has been established, a usable collection is available, a community center is emerging, and a working version of the library is now in use (http://www.dlese.org).

A critical interoperability challenge facing DLESE is bridging the gap between how end-users informally express their spatial and temporal information needs and how geo-referenced collections are formally indexed. Geo-referenced collections are indexed by geospatial footprints (e.g., a point, line, box, or polygon region on the surface of the earth specified by latitude and longitude) and temporal footprints (e.g., a point or period in time). For data holdings, geo-references may be the only metadata available. As illustrated by the scenarios in Figure 1, end-users typically express their information needs in terms of place names, earth system events, and informal spatial and temporal qualifiers (i.e., before, after, next to, around). Most educators and students are unwilling or unable to spend the time to learn how to translate their informal information needs into the necessary latitudes and longitudes. As such, this is a significant barrier for digital libraries trying to promote active, inquiry learning in the earth sciences, which often depends on students being able to locate and analyze earth data.

1) Two community college science teachers are putting together an introductory module on global climate change. They want their students to use the past history of Lake Mendota (climate and CO2 levels) to think about and predict what might happen in the future. They already have maps and images of Lake Mendota, and some textual resources that they put together. What they need are climate records from Lake Mendota going back in geologic time, including warm periods when the Earth's poles were not topped by ice caps, and a student-friendly modeling environment.

2) Some students from a Seattle high school are studying the effects of Mt. St. Helens eruptions on local geography. They are looking for data: about the regrowth of vegetation around Mt St. Helens, on the pre-explosion snowpack, and on the extent of the post-explosion mudslides.

3) A professor is searching for "before and after" data, maps, and images showing the effects of major hurricanes for use in a college level environmental sciences course.

4) Gaining access to specimens presents a huge problem for paleontology teachers and researchers. Two university teachers want to develop a digital library of significant fossils - a virtual paleontology museum – that will assemble information on key specimens, and can be tied to other information such as the geographic and stratigraphic distribution of other paleontological resources.

Figure 1. Library usage and collections development scenarios submitted by DLESE community members that explicitly or implicitly depend on spatial and temporal information (underlined).

To address this challenge, we are embarking on an ambitious research program in partnership with the Alexandria Digital Earth Prototype project to investigate an end-user perspective on digital library interoperability based on the notion of 'task-centered interactions.' The goal is to enable geoscience students, teachers, and researchers to focus on their geoscience objectives, rather than the intricacies of searching across distributed, geo-referenced digital libraries. The basic idea is to develop client tools and middleware services that support a tight integration between working, retrieving relevant information, and producing new digital content.

One thread of the research program will examine how digital library clients can enable users to easily create 'enriched' representations containing embedded resources; e.g., maps annotated with related images and data sets, or simulations annotated with maps, text, and data sets. Planned client tools include a simulation construction tool, a concept mapping tool, and a geospatial analysis tool. These clients will analyze the user's context, i.e. the work product, and pass the results of this analysis to the middleware services as an implicit geo-referenced query. For a geoscience researcher, this could mean using a prototype microclimate model, as the query itself, to locate relevant data sets to test the model's generality. For a group of fifth graders preparing a presentation on hurricanes in North Carolina, this could mean using the current presentation draft to locate pictures and videos of hurricanes that have swept through that state. The clients will be designed to enable end-users to easily and incrementally embed into their work product (e.g., their presentation slides, concept map, or simulation) related digital resources (or links to resources) to create an enriched document containing embedded geo-referenced metadata. Built into the clients will be enough intelligence to automatically analyze the enriched document context to help end-users construct powerful geo-referenced queries.

Another thread of the research program, already underway, is developing middleware services based on explicit knowledge models and machine learning algorithms. These services will perform useful semantic mappings between different resource representations (simulations, concept maps, maps, images, text, data sets) using geoscience domain concepts with spatial and temporal dimensions. Specifically we are creating four middleware services:

- An placename gazetteer that maps geographic places (e.g., Colorado, Darmstadt) into geospatial footprints;
- An event gazetteer that maps significant earth system events (e.g., Hurricane Hugo, the May 1994 eruption of Mt. Pinatubo, or the 1979 mudslides in Los Angeles) into geospatial and temporal footprints;
- A free text spatio-temporal parser based on machine learning algorithms that bridges between the informal language people use to express their geospatial and temporal information needs ("forest fires near Los Angeles") and the formal query language required by gazetteers;
- A collections registry service that directs queries to appropriate collections based on the recorded geospatial and temporal extents of registered collections.

Should this extended abstract be selected, we would like to develop a full paper for the DELOS workshop on interoperability that presents:

- A detailed example of a task-based digital library interaction; namely, a scenario where a geography faculty creates an independent lab activity whose goal is having undergraduate students understand the concept of the "hydraulic geometry" of river channels.
- Use this scenario to discuss the technical issues required for a vertical slice of intelligent clients, middleware services, and distributed collections to work together to support the desired end-user usage experience.

We would also be very interested to present or discuss this research program formally at the workshop.