

# Similarity Search for Harvesting Information to Sustain our Forests

Lois Delcambre,<sup>1</sup> Mathew Weaver<sup>1</sup>, Timothy Tolle,<sup>4</sup> David Maier,<sup>1</sup> Eric Landis,<sup>5</sup>  
Shawn Bowers<sup>1</sup>, Patty Toccalino,<sup>2</sup> Fred Phillips,<sup>3</sup> Nicole Steckler,<sup>3</sup> Craig Palmer,<sup>6</sup>  
Julia Norman<sup>2</sup>, Rupini Tummala<sup>1</sup>, Sonali Varde<sup>1</sup>

<sup>1</sup>Computer Science Dept./<sup>2</sup>Environmental Science & Eng. Dept./<sup>3</sup>Mgt. of Science & Tech. Dept.  
Oregon Graduate Institute  
20000 NW Walker Road  
Beaverton, OR 97225

<sup>4</sup>Region 6  
Strategic Planning Office  
Pacific Northwest Region  
USDA Forest Service  
Portland, Oregon 97204-3440

<sup>5</sup> Natural Resource  
Information Management  
10911 69th Street, North  
Stillwater, MN 55082

<sup>6</sup> Harry Reid Center for  
Environmental Studies  
University of Nevada,  
Las Vegas  
4505 Maryland Pkwy  
Las Vegas, NV 89154-4009

lmd@cse.ogi.edu, mweaver@cse.ogi.edu, [ttolle@fs.fed.us](mailto:ttolle@fs.fed.us), maier@cse.ogi.edu,  
elandis@ix.netcom.net, shawn@cse.ogi.edu, [toccalino@ece.ogi.edu](mailto:toccalino@ece.ogi.edu), fphillips@admin.ogi.edu,  
steckler@admin.ogi.edu, palmerc@nevada.edu, norman@ese.ogi.edu, rtummala@cse.ogi.edu,  
svarde@cse.ogi.edu

[www.cse.ogi.edu/forest](http://www.cse.ogi.edu/forest)

## 1. Introduction

The USDA Forest Service is partnered with researchers at the Oregon Graduate Institute to develop a web-based portal for forest information. The project is focusing on the information needs of the Adaptive Management Areas of the Pacific Northwest. Forest Service (Northwest Region). The Adaptive Management Area Program involves a number of federal agencies including: the USDA Forest Service (including the Region 6 Office and the Pacific Northwest Research Station), the USDI Bureau of Land Management, and the USDI Fish and Wildlife Service

The Adaptive Management Area program encourages the development and testing of technical and social approaches to achieve desired ecological, economic, and social objectives. There are ten Adaptive Management Areas in the program, as shown in Figure 1. Each of the Adaptive Management Areas ranges from about 92,000 to nearly 500,000 acres of federal lands area. And most are associated with communities that have been impacted socially and economically by reduced timber harvest from federal lands. One goal of the Adaptive Management Program is to help others to learn from the collective experience thus learning to manage & managing to learn.<sup>1</sup>

---

<sup>1</sup> Learning to manage & managing to learn is the slogan of the Adaptive Management Area Program.

## 2. Requirements for our Project

In the first six months of our project, we focused on interviewing a broad range of partners in the four target Adaptive Management Areas shown in Figure 1. We also visited Mt. Rainier National Park, a possible partner in our project. We wanted to know:

- internal and external users of information
- internal and external information providers
- national, organizational, program mandates
- human, financial and physical resources
- content, quality, updating, security, purpose, accessibility of information
- metadata and controlled vocabularies in use
- desires and visions

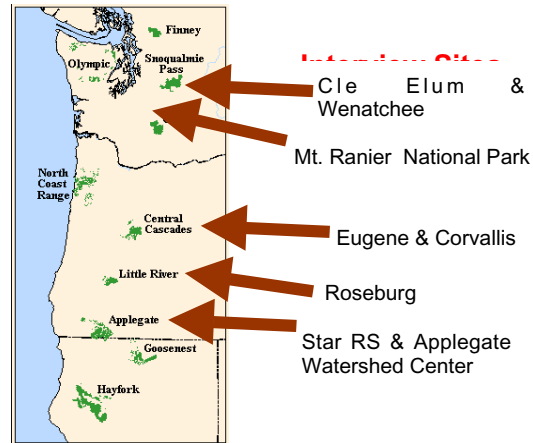


Figure 1: Ten Adaptive Management Areas

Prospective system users want easy access to information for users with various levels of expertise. They very often search for documents based on place names or geographic features. Users are generally more interested in summary information than raw data. That is, there is more interest in reading the conclusions, findings, recommendations, decisions, etc. rather than having access to the datasets gathered by scientists and subject matter experts. This is particularly significant given that scientists are often reluctant to share datasets, particularly before results have been published. We observe that information of interest is not necessarily available electronically; some information is electronic but not online. We have decided to focus our initial system on meeting the needs of land and resource manager. Common questions that a resource manager needs to answer include:

What do we know about this place? Can you find information about similar places, e.g., similar watersheds, similar forest types, similar topography, similar rural communities, and similar natural resources, such as: recreation, wilderness, minerals, water, timber, and wildlife?

Can you find information on similar activities or results, e.g., thinning, planting, burning, community assistance, and public participation?

Can you find a particular document type, e.g., Environmental Impact Statement, inventory, monitoring result, research study, map, photograph, charts, decision document, and so forth?

Can you find information about people such as What else has this person published? Reviewed? What is this person working on these days?

We learned that information providers are concerned about making contributions simple and complying with the mandate for FGDC metadata. There is a shortage of time and money for information management, beyond data collection. The primary information products are assessments, studies, surveys, and environmental impact statements. Finally, we learned that multiple agencies are involved in most projects.

### 3. Computer Science Research

We are constructing a sandbox to demonstrate the technology that we are developing. For the sandbox, we are working on the problem of providing similarity search where similarity can include any of the various domains mentioned above, such as: place, forest type, vegetation, topography, activity type, and document type. Our approach is to develop a sophisticated metadata facility that describes each domain of interest.

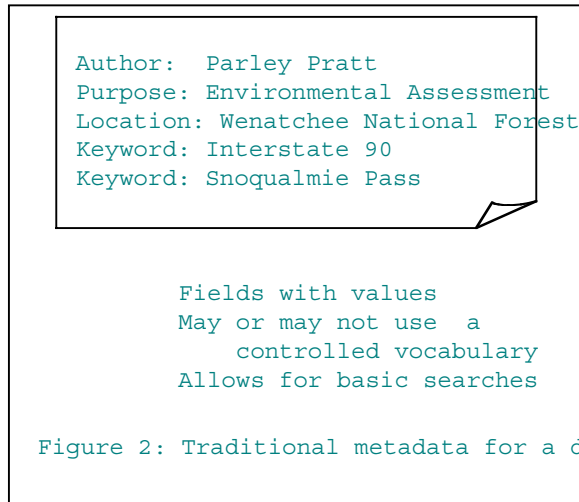


Figure 2 shows some typical metadata fields that might be associated with a document, with basic search capability.

In our work, we are representing both the properties (e.g., field names) and the terms (e.g., the values) as explicit objects. In Figure 3, we see that Document 1 has the same metadata as the document in Figure 2. But Parley Pratt, Environmental Assessment, Wenatchee National Forest, Interstate 90, and Snoqualmie Pass are all represented directly in the metadata (shown in rectangles). Also, the properties, Author, Purpose, location, and Keyword are represented directly (shown

in ovals). Using this approach we are able to represent the explicit hierarchy of properties and terms and use this for enhanced searching. For example, we know that the Wenatchee National Forest is part of the Cle Elum Ranger District, as shown by the red arrow from Cle Elum to Wenatchee. This means that if the user wants documents about Cle Elum Ranger District, the system could automatically return documents about the Wenatchee National Forest and vice versa.

We are also able to represent additional relationships of various types between terms. As an example, we may know that Interstate 90 passes through the Wenatchee National Forest or that the Wenatchee National Forest has Western Hemlocks. Then, if a document is known to be about Wenatchee National Forest, we would also know that it is near I-90 and may have Western Hemlock (as shown by the green arrows in Figure 3). Finally, we will support a flexible similarity search by traversing various relationships in the metadata. We will demo our technology for similarity search at the dg.o conference.

### 4. Plans for the System

Ultimately in the system we plan to integrate GIS and map data, a gazetteer to support searching by place name, and document storage with our enhanced metadata. We also intend to use appropriate controlled vocabularies or metadata standards of use for our domains. We are currently evaluating controlled vocabularies for possible inclusion in our project. We are using documents from several of the Adaptive Management Areas and we are building metadata and search capability to support several scenarios specified by the user community within the Forest Service and the Bureau of Land Management.

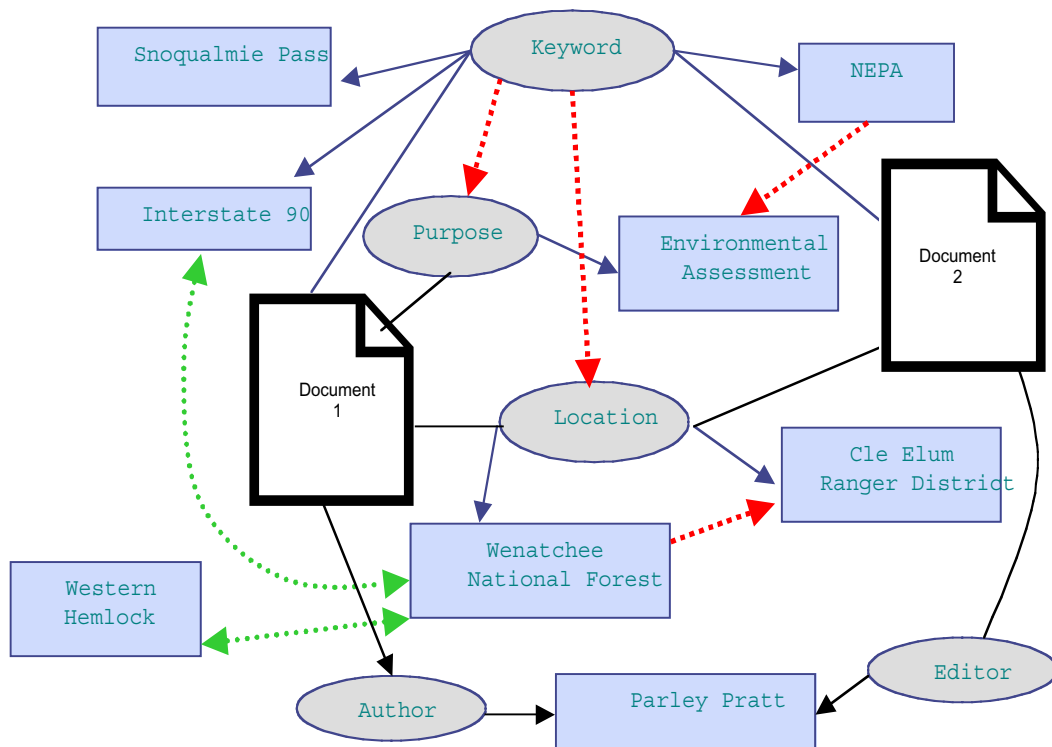


Figure 3: Enhanced metadata — to support similarity search

## 5. Management Science Research

We are studying portals business models to determine their critical success and failure factors. We are also studying the inhibitors and facilitators of information sharing in organizations. The idea is to identify critical factors that have a practical impact on willingness to share information. We are planning to apply a relative variety analysis as an extension of Data Envelopment Analysis (a public program evaluation tool) to measure the efficiency and flexibility of participating agencies. Finally, we are going to use the Adaptive Management Areas as a case study to evaluation the formation process and critical success factors as part of an ongoing study of R&D Consortia.