Gauging Agency Involvement in Environmental Management Using Text Analysis of Laws and Regulations

Julia A. Ekstrom, Gloria T. Lau, Jack C.P. Cheng, Daniel J. Spiteri, and Kincho H. Law

jekstrom@lbl.gov, {glau, cpcheng, law}@stanford.edu, danspiteri@gmail.com

Stanford University, Stanford, California, 94305

Abstract

This paper presents an open source application that uses the text analysis of laws and regulations to gauge government agency involvement in any given topic related to coastal and ocean management. It is well-established that management of the coasts and oceans is transitioning to integrate ecosystem concepts and considerations into management decisions. To implement such a transition, baseline knowledge of ecological systems and management systems is needed. Much work has focused on the compiling and synthesizing of ecosystem understanding, but relatively little effort has provided comparable information about management from a comprehensive perspective. In this paper, we describe our exploration and development of an accurate metric to gauge government agency involvement, which represents an important aspect of management. The results of three text analysis-based metrics (frequencies of statutes and regulations, legal sections, and terms) are tested against survey results completed by domain experts. Results showed that the frequency of sections and terms were similarly accurate when compared to survey results. Further, we report an open source tool we have developed that allows users to perform the agency involvement analysis. A variety of applications and potential uses are described. This highlights an avenue for digital government approaches to progress natural resource management in dealing with emerging problems of today and of the future.

Keywords

Environmental management, terminological taxonomies, legal inventory, regulation, government agency, ecosystem-based management.

1. Introduction

Scientific studies of fishery collapses, harmful algal blooms, hypoxic zones, invasive species and other threats indicate that the ocean health is in decline.[1] Emerging threats of climate change, ocean acidification, sea level rise, continued coastal development, and others plague the projected future of marine ecosystems.[2] The impacts of these threats are often as consequences of the culmination of multiple source activities.[3] One key to restoring, mitigating, and preventing further destruction is in strategically altering management institutions that guide the current

---

[1] Ransom A. Myers and Boris Worm, "Rapid worldwide depletion of predatory fish communities" *Nature* 423 (2003): 280-283.

Daniel Pauly, Villy Christensen, Johanne Dalsgaard, Rainer Froese and Francisco Torres, "Fishing Down Marine Food Webs" *Science* 279, no. 5352 (1998): 860-863.

PM Glibert, S Seitzinger, CA Heil, JM Burkholder, MW Parrow, LA Codispoti and V Kelly, "The Role of Eutrophication in the Global Proliferation of Harmful Algal Blooms: New perspectives and new approaches " *Oceanography* 18, no. 2 (2005).

Ransom A. Myers and Boris Worm, "Rapid worldwide depletion of predatory fish communities" *Nature* 423 (2003): 280-283.

Daniel Pauly, Villy Christensen, Johanne Dalsgaard, Rainer Froese and Francisco Torres, "Fishing Down Marine Food Webs" *Science* 279, no. 5352 (1998): 860-863.

PM Glibert, S Seitzinger, CA Heil, JM Burkholder, MW Parrow, LA Codispoti and V Kelly, "The Role of Eutrophication in the Global Proliferation of Harmful Algal Blooms: New perspectives and new approaches " *Oceanography* 18, no. 2 (2005).

[2] Millennium Ecosystem Assessment, *Ecosystems and Human Well-being: Synthesis*, (Washington, D.C: Island Press, 2005).

J.C. Orr, V.J Fabry, O Aumont, L Bopp, S.C Doney, R.A Feely, A Gnanadesikan, N Gruber, A Ishida, F Joos, R.M Key, K Lindsay, E Maier-Reimer, R Matear, P Monfray, A Mouchet, G. Raymond, R.G Najjar, G-K Plattner, K.B Rodgers, C.L Sabine, J.L Sarmiento, R Schlitzer, R.D Slater, I.J Totterdell, M-FY Weirig, Y Yamanaka and A. Yool, "Anthropogenic ocean acidification over the twenty-first century and its impact on calcifying organisms" *Nature* 437 (2005).

Joint Ocean Commission Initiative, "One Coast, One Future: Securing the Health of West Coast Ecosystems and Economies, (Accessed Jan. 15, 2009 at http://www.jointoceancommission.com/resource-center/1-Reports/2009-01-15_One_Coast_One_Future.pdf)," (Washington, D.C. 2009).

[3] Benjamin S. Halpern, Karen L. McLeod, Andrew A. Rosenberg and Larry B. Crowder, "Managing for cumulative impacts in ecosystem-based management through ocean zoning" *Ocean & Coastal Management* 51, no. 3 (2008): 203-211.

Benjamin S. Halpern, Shaun Walbridge, Kimberly A. Selkoe, Carrie V. Kappel, Fiorenza Micheli, Caterina D'Agrosa, John F. Bruno, Kenneth S. Casey, Colin Ebert, Helen E. Fox, Rod Fujita, Dennis Heinemann, Hunter S. Lenihan, Elizabeth M. P. Madin, Matthew T. Perry, Elizabeth R. Selig, Mark Spalding, Robert Steneck and Reg Watson, "A Global Map of Human Impact on Marine Ecosystems" *Science* 319, no. 5865 (2008): 948-952.

practices so that decisions are made accounting for the multitude of environmental impacts.[4] Government agencies, non-governmental organizations, policy makers and other ocean policy constituents now seek to alter management so that it is guided by ecosystem principles and considerations.[5]

Historically, the oceans have been managed within isolated sectors of state and federal government. Currently, there is a growing momentum to transition out of the sector-based approach into an ecosystem-based management (EBM) system.[6] A major roadblock to implementation of EBM is that it requires coordination and communication among sectors within and between levels of government.[7]

Fundamentally, coordination in any domain of management requires knowledge of baseline information about what agencies need to collaborate and in what capacity; such foundational information is not always easily accessible, depending on the complexity of the issue. For instance, when Hurricane Katrina hit New Orleans, the appropriate government agencies did not respond rapidly. The delay in response was largely due to the lack of knowledge

---

[4] H. Ken Cordell and John C. Bergstrom, *Integrating Social Sciences with Ecosystem Management: Human Dimensions in Assessment, Policy, and Management*, (Champaign, IL: Sagamore Publishing, 1999).

H. J. Cortner, M. G. Wallace, S. Burke and M. A. Moote, "Institutions matter: the need to address the institutional challenges of ecosystem management" *Landscape And Urban Planning* (1998): 159-166.

Joint Ocean Commission Initiative, "One Coast, One Future: Securing the Health of West Coast Ecosystems and Economies, (Accessed Jan. 15, 2009 at http://www.jointoceancommission.com/resource-center/1-Reports/2009-01-15_One_Coast_One_Future.pdf)," (Washington, D.C. 2009).

[5] Cassandra Barnes and Katherine W McFadden, "Marine ecosystem approaches to management: challenges and lessons in the United States" *Marine Policy* 32, no. 3 (2008): 387-392.

[6] Karen L McLeod, Jane Lubchenco, Steve R Palumbi and Andrew A Rosenberg, "Scientific Consensus Statement on Marine Ecosystem-Based Management " *Communication Partnership for Science and the Sea*, (2005).

Andrew A Rosenberg and Karen McLeod, "Implementing ecosystem-based approaches to management for the conservation of ecosystem services" *Marine Ecology Progress Series* 300 (2005): 270-274.

Ken Sherman, M Sissenwine, Villy Christensen, A Duda, G EHempel, C Ibe, S Levin, D Lluch-Belda, G Matishov, J McGlade, M O'Toole, S Seitzinger, R Serra, HR Skjoldal, Q Tang, J Thulin, V Vandeweerd and K Zwanenburg, "A global movement toward an ecosystem approach to management of marine resources" *Marine Ecology Progress Series* 300 (2005): 275-279.

[7] H. J. Cortner, M. G. Wallace, S. Burke and M. A. Moote, "Institutions matter: the need to address the institutional challenges of ecosystem management" *Landscape And Urban Planning* (1998): 159-166.

Lawerence Juda and Timothy Hennessey, "Governance profiles and the management of the uses of large marine ecosystems" *Ocean Development and International Law* 32, no. 1 (2001): 43-69.

Joint Ocean Commission Initiative, "One Coast, One Future: Securing the Health of West Coast Ecosystems and Economies, (Accessed Jan. 15, 2009 at http://www.jointoceancommission.com/resource-center/1-Reports/2009-01-15_One_Coast_One_Future.pdf)," (Washington, D.C. 2009).

for what agencies were responsible, who needed to coordinate with whom, and what the chain of authority was supposed to respond.[8]

Given the sector-based nature of ocean and coastal management, compounded with the overlapping nature of the activities and natural resources,[9] there is a strong need for digital government tools to systematically generate and access baseline data about government agency involvement. Arming government agencies with tools to retrieve basic information about what agencies should be involved in the issue at hand without weeks or months of analysis could facilitate inter- and intra-agency coordination as well as strategic policy-making.[10] This type of information retrieval tool could also be useful as a starting point to direct longer and more in-depth analyses. Such further analyses would include those currently conducted for policy and legal analysis, which would include court cases and other non-statutory materials (including legislative histories), international treaties, business decisions by the government, budget and implementation information, and other pertinent information.

Traditionally, retrieving information about what agencies should coordinate for a given issue and in what capacity is performed by personnel involved coupled with legal analysis. Such a process can be lengthy, but often a situation requires rapid response, as with Hurricane Katrina. Ekstrom and Lau (2008) presented a preliminary algorithm that maps out what agencies are involved in any variety of topics by virtue of laws and regulations using relative term frequencies of topics.[11] Such an approach allows a user to identify objectively what agencies are involved in management of a topic across sectors and levels of government. While this *involvement* measure does not necessarily translate directly into management action, it does provide an objective and quantitative measure of *assumed* involvement that can be harnessed from the laws and regulations (and eventually other types of management-relevant documents). Knowing relative agency

---

[8] John R. Harrald, "Agility and Discipline: Critical Success Factors for Disaster Response" *The ANNALS of the American Academy of Political and Social Science* 604 (2006): 256-273.

GAO, "Hurricanes Katrina and Rita: Coordination between FEMA and the Red Cross should be improved for the 2006 hurricane season," In United States Government Accountability Office: Repot to Congressional Committees (ed.), (Washington, DC 2006).

[9] Larry Crowder and Elliott Norse, "Essential ecological insights for marine ecosystem-based management and marine spatial planning" *Marine Policy* 32, no. 5 (2008): 772-778.

[10] Julia A Ekstrom, Oran R Young, Steve Gaines, Maria Gordon and Bonnie J McCay, "A tool to navigate overlaps in fragmented ocean governance" *Marine Policy* 33, no. 2 (2009).

[11] Julia A Ekstrom and Gloria Lau, "Exploratory text mining of ocean law to measure overlapping agency and jurisdictional authority," *Proceedings of the Digital Government Research Conference*, (Montreal, Canada 2008).

involvement does not indicate whether or not necessary coordination is occurring, but rather provides a first step for users needing baseline management information to identify emerging issues (e.g. tidal energy, offshore aquaculture, wind farm development, etc.). The next steps forward in developing a retrieval system to make baseline management data easily accessible is to develop and explore parameters that quantitatively reveal relative degree of agency involvement in any user-defined topic.

1.1 Invasive Species Background

Invasive species management challenges provide an example of overlapping jurisdictions and needs for coordination in coastal and ocean management.[12] Aquatic invasive species management in the United States costs an estimated $9 billion each year.[13] One species alone, the zebra mussel, cost the nation over five billion dollars for the damaging water intake pipes in the Great Lakes region.[14] The State of California has identified 607 aquatic invasive species in the State's estuarine waters. There are over twenty pathways (commonly referred to as "vectors") through which non-native aquatic species are introduced into the state waters. These include (but are not limited to) ballast water exchange, commercial fishing gear, recreational boating, aquarium trade, live bait and live seafood imports, and aquaculture of non-native species.[15]

One of the main goals set by the California Aquatic Invasive Species Management Plan was for the State to conduct an analysis of existing management, identifying what laws and

---

[12] Pew Oceans Commission, "America's Living Oceans: Charting a Course for Sea Change," (2003).

DC Schmitz and D Simberloff, "Need: A national center for biological invasions" *Issues in Science and Technology* (2001): 57-62.

Clare Shine, Nattley Williams and Lothar Gündling, *A Guide to Designing Legal and Institutional Frameworks on Alien Invasive Species*, (IUCN, 2000).

US Commission on Ocean Policy, "An Ocean Blueprint for the 21st Century Final Report of the U.S. Commission on Ocean Policy," (2004).

[13] California Department of Fish and Game, "California Aquatic Invasive Species Management Plan" Available at http://www.dfg.ca.gov/invasives/plan/ (accessed December 2008).

David Pimentel, "Economic and Ecological Costs Associated with Aquatic Invasive Species," In K Wakefield and A Faulds (eds.), *Proceedings of the Aquatic Invaders of the Delaware Estuary Symposium, Malvern, Pennsylvania, May 20, 2003. (http://www.sgnis.org/publicat/proceed/aide/pime2003.htm)*, (2003).

[14] David Pimentel, "Aquatic Nuisance Species in the New York State Canal and Hudson River Systems and the Great Lakes Basin: An Economic and Environmental Assessment" *Environmental Management* 35, no. 5 (2005): 692-702.

[15] California Department of Fish and Game, "California Aquatic Invasive Species Management Plan" Available at http://www.dfg.ca.gov/invasives/plan/ (accessed December 2008).

regulations the State already has that pertain to each specific pathway and invasive species. Additionally, one of the plan's primary tasks is to identify which agencies are and should be involved in management of invasive species. Given the complexity and long list of pathways through which non-native species are introduced into the state waters, this can be a time consuming project. We used this existing management challenge in California as a scenario to explore the utility of the agency involvement metric.

1.2 Objective

We began this project seeking to determine the most accurate parameter for gauging agency involvement. This investigation continues the work from Ekstrom and Lau (2008) which presents a preliminary technique that displays term frequencies in laws organized visually around their relevant agencies (Figure 1). Lines are drawn from each law and regulation to the authoritative government agency (represented by an acronym). Each document is represented by a node (pink = regulation, red = statute) which is sized by the value of term frequency for the term queried. Thus, in Figure 1 the nodes have been eliminated with only a line remaining for those laws and regulations in which the term "fishing" does not occur. To progress this technique a step further, we sought to quantify agency involvement again using the laws and regulations, with the help of a domain-specific taxonomy. In this work, we incorporated domain expert survey response in order to verify the accuracy of various parameters used in the analysis. We also used the survey data to determine if one parameter was more accurate than another, and to determine whether and how using lower level taxonomic terms would increase the accuracy of the metric.

After conducting a domain expert survey and running the analyses, we found our six measures yielded similar results, all of which were quite accurate in identifying the top most agencies involved. As such, using the most accurate set of parameters, we developed an application that provides this agency involvement metric for public use. Thus, this paper is divided into two parts. First, we present the techniques of the analysis and accuracy tests, including a description of the data and survey implemented. Second, we present the prototype application that provides users with access to the agency involvement data.

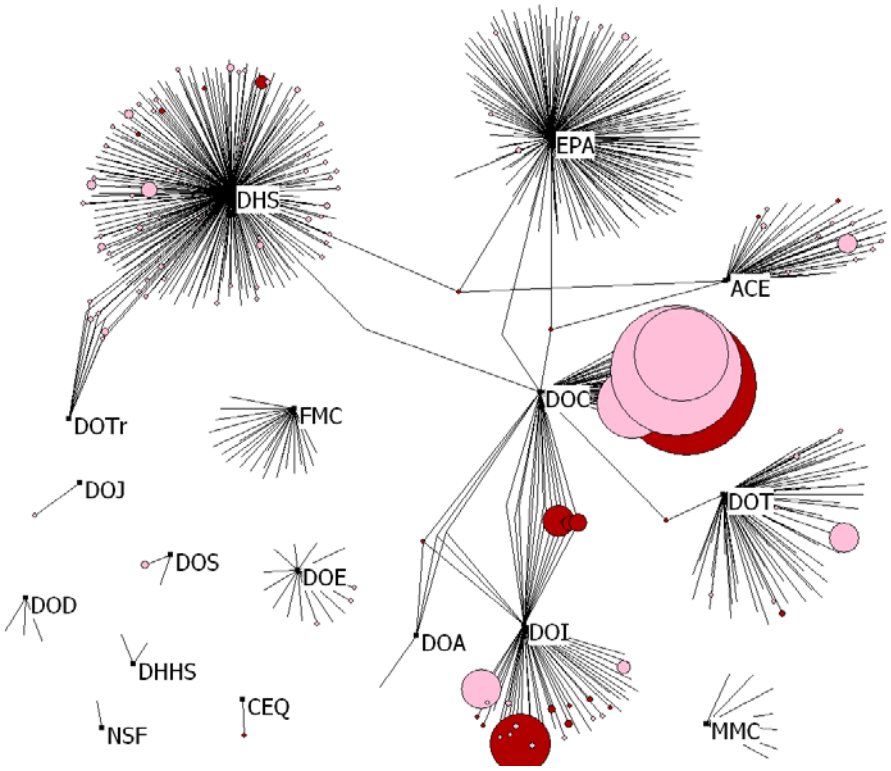


Figure 1: Network diagram depicting relative federal agency involvement in topic of fishing, by virtue of the term frequency in the laws and regulations.[16] *Lines are drawn from laws (red nodes) and regulations (pink nodes) to authoritative agencies, and node size varies with topic frequency.*

2. DATASET

2.1 Document Collection

To develop and test the six measures of the agency involvement metric, two types of data were used for this exploratory analysis: (a) document collection of marine and coastal laws and regulations; (b) a record of agency responsibility for each document. For the latter dataset, each regulation was tagged for what agency wrote it and each statute was tagged for what agency or agencies Congress granted authority to implement it.

[16] Developed in the followed works:
Julia A Ekstrom and Gloria Lau, "Exploratory text mining of ocean law to measure overlapping agency and jurisdictional authority," *Proceedings of the Digital Government Research Conference*, (Montreal, Canada 2008).

Julia A Ekstrom, Oran R Young, Steve Gaines, Maria Gordon and Bonnie J McCay, "A Tool to Navigate Overlaps in Fragmented Ocean Governance" *Marine Policy* 33, no. 2 (2009).

2.1.1 Scope of Document Collection

The document collection used is composed of a comprehensive set of statutes and regulations related to the marine and coastal region of the California coast of the United States. The documents are codified federal United States and the State of California laws and regulations from the Year 2006.[17]

2.1.1.1 Record of Agencies to Documents

An important metadata for the law collection is the agency authority for each document. In the form of an agency-by- document table (

Table 1), the agency or agencies with responsibility to implement each statute or regulation was recorded.[18]

Table 1. Sample of record of agency to documents. Row headers are the documents (sample of federal U.S. statutes used in the analysis). Column headers are a sample federal agencies (ACE: Army Corps of Engineers; EPA: Environmental Protection Agency; DOC: Department of Commerce; DHS: Department of Homeland Security; DOT: Department of Transportation). Cells with a one (1) indicate the agency has the assumed responsibility to implement the law. Cells with a zero (0) marks the laws over which the agency does not have direct responsibility.

	ACE	EPA	DOC	DHS	DOT
Clean Water Act	1	1	0	1	0
Fishery Conservation and Management Act	0	0	1	0	0
Deepwater Port Act	0	0	1	0	1

[17] Julia A Ekstrom, "California Current Large Marine Ecosystem: Publicly available dataset of state and federal laws and regulations" *Marine Policy* 33, no. 2 (2009): 532-535.

Julia A Ekstrom and Gloria T Lau, "Exploratory text mining of ocean law to measure overlapping agency and jurisdictional authority," *Proceedings of the Digital Government Research Conference*, (Montreal, Canada 2008).

[18] Julia A Ekstrom and Gloria T Lau, "Exploratory text mining of ocean law to measure overlapping agency and jurisdictional authority," *Proceedings of the Digital Government Research Conference*, (Montreal, Canada 2008).

2.2 Terminological Taxonomies

Using a single word and phrase in a query to represent a concept often is not sufficient in information retrieval systems.[19] Several approaches for querying document collections are typically used to assist in information retrieval, including structure-based queries, Boolean searches, context queries, and natural language queries.[20] Terminological taxonomies, a more advanced approach to query-building, are a hierarchical organization of terms and phrases to define a single topic or concept. This approach has shown to increase the accuracy of information retrieval when they are created for domain specific inquiries and constructed using domain specific vocabularies.[21]

2.2.1 Constructing Terminological Taxonomy

We sought to determine the benefit, if any, of considering a lower (more detailed) taxonomic level to retrieve agency involvement information from natural resource management law and regulation. It is important for system users to understand the benefits and limitations of using only general terms, as opposed to also incorporating specific terms, to define their topic. To explore such benefits and limitations, Ekstrom, in consultation with domain experts, constructed a domain specific terminological taxonomy using the California Aquatic Invasive Species Management Plan.[22] This document contains an extensive description of the individual pathways of aquatic invasive species in the State of California and a full species list (with vernacular and scientific names). Given that pre-defined terminological taxonomies do not necessarily exist for every domain, a user could use such a management plan document to construct topic queries from either the general-only or general and specific levels of terms. The taxonomy was created using

- pathway industries (human activities or industries through which invasive species enter California)

---

[19] Ronen Feldman and James Sanger, *The Text Mining Handbook: Advanced Approaches to Analyzing Unstructured Data*, (Cambridge: Cambridge University Press, 2007).

[20] Ricardo Baeza-Yates and Berthier Ribeiro-Neto, *Modern Information Retrieval*, (New York: ACM Press, 1999).

[21] Jack C. P. Cheng, Gloria T. Lau, Kincho H. Law, Jiayi Pan and Albert Jones, "Regulation Retrieval Using Industry Specific Taxonomies" *Artificial Intelligence and Law* 16, no. 3 (2008): 277-303.

Jack C. P. Cheng, Gloria T. Lau, Kincho H. Law, Jiayi Pan and Albert Jones, "Improving Access to and Understanding of Regulations through Taxonomies" *Government Information Quarterly* 26, no. 2 (2009).

[22] California Department of Fish and Game, "California Aquatic Invasive Species Management Plan" Available at http://www.dfg.ca.gov/invasives/plan/ (accessed December 2008).

- categories of invasive species

The pathway industries and categories of species were composed of a general level (L1, Table 2) and a more detailed level of terms (L2, Table 2). We divided the Management Plan's aquatic invasive species into four general categories: fish, plant, invertebrate, and amphibian. For the higher level of taxonomy, we investigated eleven pathway industries: commercial fishing, recreational boating, recreational equipment, aquarium and aquascaping trade, live bait, live seafood import, aquaculture, shipping and navigation, and drilling platforms, and amphibious and sea planes. Each of these industries is defined by the State of California to facilitate the entrance of non-native aquatic species into the state waters.[23] Each general category contained a variety of more specific terms to define each concept.

Table 2. Sample of terminological taxonomy applied.

Concept	Taxonomic Level	Term(s)
Commercial fishing	L1	Commercial fishing, commercial fisheries, commercial fishery
	L2	Gear
	L2	Fishing net
	L2	Fishing line
	L2	Trawl
	L2	Trap
Aquaculture	L1	Aquaculture, mariculture, fish farming, tuna pen, sea ranching
	L2	Trade species
	L2	Hitchhiker species
	L2	Parasite
	L2	Stock enhancement
Invasive invertebrate	L1	+(invasive exotic introduced nonindigenous imported nonnative "non-native" "biological pollutant" alien cryptogenic established) +("invertebrate" "invertebrates")
	L2	Asian overbite clam, Corbula amurensis
	L2	Channeled apple snail, Pomacea canaliculata
	L2	Chinese mitten crab, Eriocheir sinensis
	L2	European green crab, Carcinus maenas
	L2	Golden mussel, Limnoperna fortune
	L2	New Zealand mudsnail, Potamopyrgus antipodarum

[23] California Department of Fish and Game, "California Aquatic Invasive Species Management Plan" Available at http://www.dfg.ca.gov/invasives/plan/ (accessed December 2008).

	L2	Northern Pacific seastar, Asterias amurensis
	L2	Quagga mussel, Dreissena bugensis
	L2	Sabellid polychaete, Terebrasabella heterouncinata
	L2	Shipworm, Teredo navalis
	L2	Zebra mussel, Dreissena polymorpha

A use case of our system is given here. A user interested in the topic *commercial fishing* can query our system using the phrase *commercial fishing* to gather relevant laws. Perusing the management plan, the user might want to expand into lower level taxonomy terms to define what constitutes the topic *commercial fishing*, such as *gear*, and specific gears including *traps*, *fishing line*, *fishing nets*, and *trawl* which are the specific avenues through which invasive species enter California waters through the commercial fishing industry (Table 2). However, the value of defining topics by using the lower level (L2) list of terms and phrases from the taxonomy is unknown to the user. As such, we sought to test whether it is necessary to include the more detailed terms, lower in the terminological taxonomy, in a search query to accurately retrieve all the relevant agencies involved in management of the topic.

2.3 Exploring parameters

In order to gauge agency involvement using our collection of laws and regulations, we compute the occurrence *frequency* of the *topic* in our collection, where each document is tagged with its enforcing agency. As this is a pilot study, it has never been established what *frequency* we should be recording, and thus we will be testing three frequency parameters here:

- Parent document (codified chapter or division) count per topic
- Legal section count per topic
- Term frequency per topic per agency

Apart from the frequency parameter, we also need to establish the definition of a *topic*. Here, we will investigate the value of lower level taxonomy terms. Two techniques were applied to define the topic parameter to test the added benefit of terminological taxonomies:

- Without taxonomy (single concept defined by general level, L1)

- With taxonomy (single concept defined by combined general and specific levels, L1 and L2)

The three frequency parameters and two topic parameters combine to six measures of interest. Using these six measures to devise the agency involvement algorithm, we aimed to identify whether any measure generates the full array of agencies involved in a topic. In addition, we sought to determine whether any (or all of) the measures accurately reveals the top most involved agencies.

2.3.1 Frequency of Document Unit

The first parameter used to measure agency involvement was the frequency of documents containing the topic query under the responsibility of each agency (Table 3).

Table 3. Sample of recorded document frequency by agencies for one concept without (L1) and with (L1, L2) inclusion of the terminological taxonomy.

Topic	Level	Agency			
		ACE	**EPA**	**DOC**	**DHS**
Commercial fishing	L1	3	3	22	20
	L1, L2	3	18	27	32

2.3.2 Frequency of Legal Section

As the second parameter to measure agency involvement, we calculated the number of legal sections containing the topic query under the assumed responsibility of each agency (

Table 4). Text analysis is often performed on elements derived from larger documents. Increasing the granularity of a set of documents enables a higher resolution of analysis. Documents are typically divided based on structure or size in more digestible elements. For example, a corpus of text from a book may be divided into chapters, paragraphs, or sentences for more detailed analysis.[24] Similarly, laws and regulations are organized in sections, which is the smallest consistent composition in which these documents are structured. As such, we use the frequency of individual sections as one of the exploratory parameters for agency involvement.

[24] Klaus Krippendorff, *Content Analysis*, (Thousand Oaks: Sage Publications, 2004).

Table 4. Sample of section frequency by agencies for one concept without (L1) and with (L1, L2) inclusion of the terminological taxonomy.

Topic	Level	Agency			
		ACE	**EPA**	**DOC**	**DHS**
Commercial fishing	L1	7	4	115	60
	L1, L2	25	33	212	110

2.3.3 Frequency of Concept Term

The third parameter was number of occurrence of the topic query in the entire document corpus (Table 5). Regular expression is leveraged for pattern matching and computing the occurrence frequency of concept terms.

Table 5. Sample of term frequency by agencies for one concept without (L1) and with (L1, L2) inclusion of the terminological taxonomy.

Topic	Level	Agency			
		ACE	**EPA**	**DOC**	**DHS**
Commercial fishing	L1	8	4	343	84
	L1, L2	19	257	1362	125

2.4 Survey

The survey was designed to gather agency involvement information from domain experts from whom we could verify the accuracy of the parameters. We targeted individuals with experience in management of pathways related to aquatic invasive species in California. The survey was performed by asking the six experts to rank the top agency or agencies, up to five, that they know are responsible for management of a given topic. A list of the fifteen topics was included in the survey (eleven of which were pathways and four of which were species groupings). Respondents were asked to only fill in information for those they were confident.

Each of the three frequency parameters was run for the fifteen topics in our taxonomy at the broad level (L1) and at the more detailed level (L2). Thus, for each topic there were six measures of results for comparison. The parameter results were compared against survey data filled out by six agency domain experts in ocean and coastal management to evaluate the accuracy of the parameters.

3. EVALUATION AND ASSESSMENT

3.1 Results

Preliminary results showed that the term frequency parameter (Table 5) and the section frequency parameter (Table 4) generate accurate results of agency involvement, based on the top ranked agency involved, more often than the document frequency parameter. Error tests (Root Mean Square Error) based on top four ranked agencies showed no notable difference in error among the three parameters, nor between the application of the parameters with and without the taxonomy.

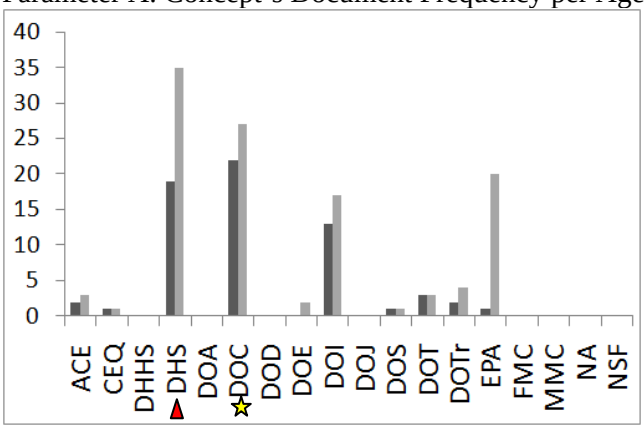
In evaluating the accuracy of the parameters, we found that for some topics all of the parameters were accurate. For analysis of federal laws and agencies, all three frequency parameters performed on two of the ten topics (*recreational boating* and *aquaculture*), with and without the taxonomy, accurately revealed the highest ranked agency. For analysis on the State laws and agencies, all three frequency parameters performed on eight of the twelve topics, with and without the taxonomy, accurately revealed the highest ranked agency.

Of the three frequency parameters, document frequency is shown to be the least accurate measure of agency involvement. In several cases, the document frequency parameter generated results inconsistent with the survey. For example, for the topic of *commercial fishing* in federal law (Figure 2), two notable inconsistencies between use of the taxonomy and without the taxonomy are in the document frequency count for the Environmental Protection Agency (EPA) and the Department of Homeland Security (DHS). Both inaccuracies are generated using the document frequency parameter (Parameter A, Figure 2) while the other two parameters yielded results consistent with the surveys (Parameters B and C, Figure 2). Without the lower level taxonomy terms, the EPA shows minimal involvement, whereas incorporation of the taxonomy shows the EPA is highly involved, surpassing the rank of the third and fourth agencies (Parameter A, Figure 2). The dramatic increase is primarily due to the inclusion of the term *trap*. The laws and regulations under the EPA containing this term relate to waste control standards and

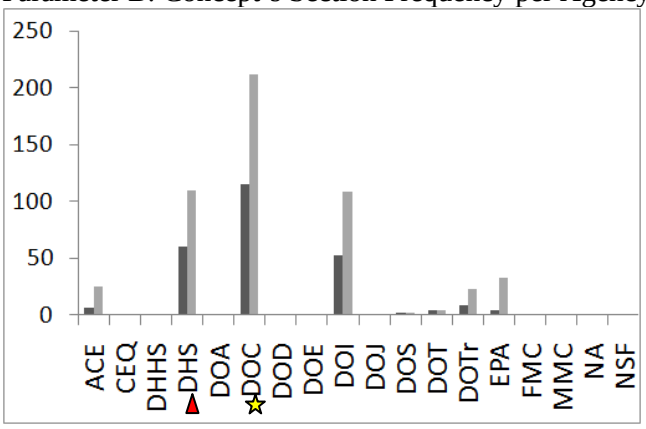
monitoring procedures, which is a different context than the intended fishing traps. For example, one section of the EPA's regulations of Emission Standards refers to *trap* in the following manner (emphasis added): "Particulate *trap* means a filtering device that is designed to physically *trap* all particulate matter above a certain size."[25] Therefore, this Level 2 term creates misleading results. Another error generated by the document frequency parameter is that when the taxonomy is applied, the DHS ranks as the top most involved agency, above the Department of Commerce (DOC), generating result inconsistent with the survey responses. The inconsistency is likely because the DHS documents in which the topic *commercial fishing* occurs tend to contain one or few references to the topic. In addition, the documents under DOC in which the topic *commercial fishing* occurs tend to contain many references to the topic.

[25] "40 C.F.R. 1039" (2006).

Parameter A. Concept's Document Frequency per Agency



Parameter B. Concept's Section Frequency per Agency



Parameter C. Concept's Term Frequency per Agency.

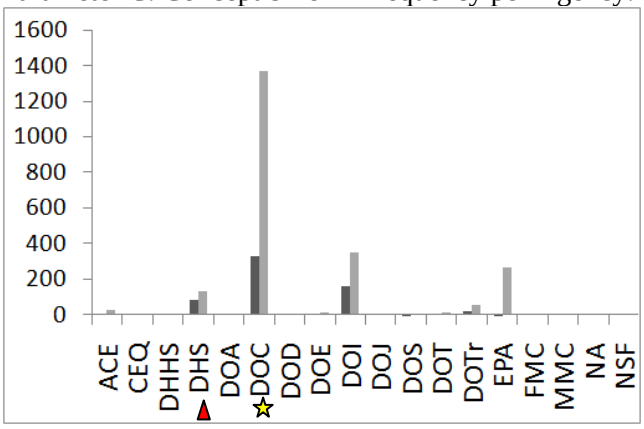
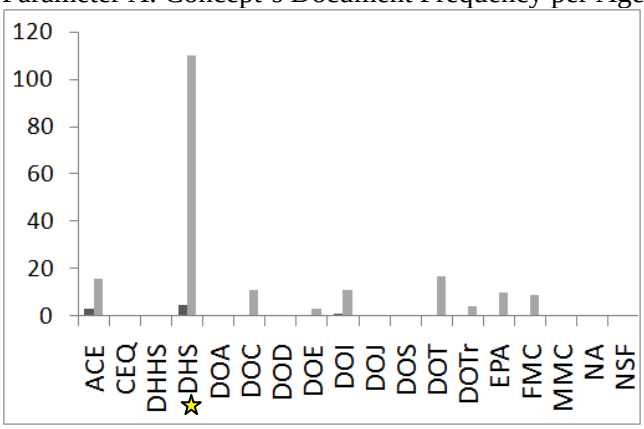
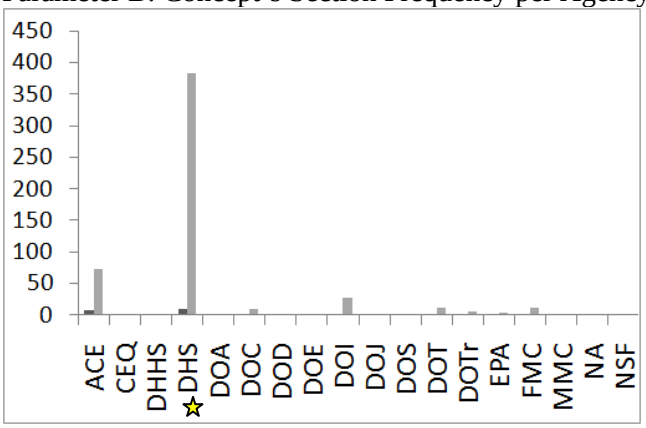


Figure 2. Bar charts of exploratory agency involvement metrics for the topic of commercial fishing. Parameters run without terminological taxonomy in dark gray and those run with taxonomy in light gray. Primary and secondary agencies according to surveys are indicated with a star ( ) or triangle ( ) next to the agency acronym.

Parameter A. Concept's Document Frequency per Agency



Parameter B. Concept's Section Frequency per Agency



Parameter C. Concept's Term Frequency per Agency.

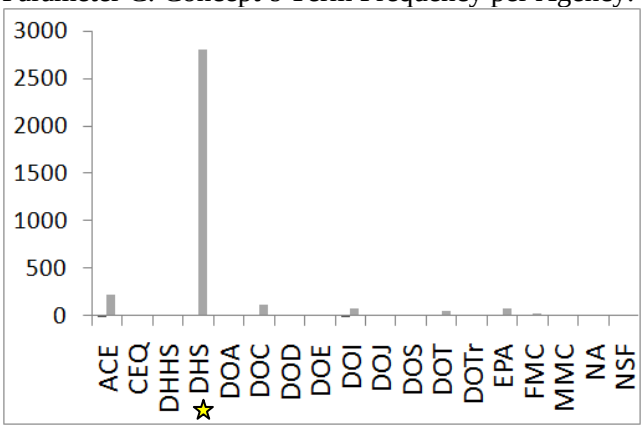


Figure 3. Bar charts of exploratory agency involvement metrics for the topic of recreational boating. Parameters run without terminological taxonomy in dark gray and those run with taxonomy in light gray. Primary and secondary agencies according to surveys are indicated with a star (☆) next to the agency acronym.

Contrary to the topic of *commercial fishing*, some cases demonstrated an added benefit of applying the taxonomy. For example, *recreational equipment* without the use of the terminological taxonomy did not occur in any federal law or regulation. But the use of the lower level terms did reveal agencies involved. In other situations, as with *recreational boating*, the rank of agency involvement remained consistent with and without the taxonomy (Figure 3). However, results for the two topic parameters, namely with and without the consideration of the taxonomy, were identical for the top two agencies and generated very similar overall results.

3.2 Interpretation of Preliminary Results

3.2.1 Comparing Parameters

The most notable finding from our preliminary investigation of developing an agency involvement metric is that the parameters of the term frequency and section frequency were more consistently accurate than the document frequency. In examining the highest ranked agency involved, we found the taxonomy did not increase the accuracy of the parameters, and in the case of *commercial fishing*, decreased the accuracy of the parameters.

Those topics with inconsistent parameter results (federal *shipping*; State of California: *recreational boating* and *shipping and navigation)* tended also to have inconsistent survey results. Thus, the variability in the survey data may indicate that no clear lead agency exists for managing these issues. Instead these issues may be managed by a handful of agencies and survey respondents could be familiar with different aspects of the management.

3.2.2 Utility of Incorporating Taxonomy

Overall there was very little notable added benefit of including terms from the lower level of the taxonomy, with two exceptions. One topic, *recreational equipment*, at the high level terminology did not occur in the federal portion of the document collection (recreational equipment). Incorporation of the detailed terms showed that federal agencies were in fact involved in the management of the issue. Thus, the use of lower level terms is suggested for cases where the higher level terms do not appear. The second added benefit of the taxonomy was that in most cases it appeared to create a bigger margin of difference between the agencies involved, thus resulting in an agency involvement chart with observable differences among agencies.

In analysis of highest ranked agency in the federal level, inclusion of the more detailed lower level terms increased the accuracy of the results in only two topics, *recreational boating* and *shipping and navigation*. In analysis of the highest ranked agency for the state level, inclusion of the more detailed terms did not increase the accuracy of the parameters for any topic. Inclusion of the lower level terms decreased the accuracy of the document frequency parameter for two topics (*invasive plants* and *recreational boating*).

Given that in some cases the inclusion of the taxonomy caused a decrease in accuracy, this serves as a warning for users implementing lower level terms. It is important that such terms are selected and implemented with caution because it can generate erroneous results, especially if used with the document frequency parameter.

4. AGENCY INVOLVEMENT TOOL

As we have determined in the previous section that the term frequency parameter and the section frequency parameter provided similarly accurate results, we can now implement a publicly accessible tool to generate the basic agency involvement metric using either frequency parameter, and we chose to use the section frequency count. The topic parameter will be defined by the user. The functionality of gauging agency involvement is part of a larger application (MINOE) geared to assist those interested in ecosystem-based management to navigate through the morass of law and regulation related to any topic.[26]

4.1 Features and Functionality

Foremost, the application gives easy access to determining what agencies have laws and regulations that discuss a user-defined topic. This information can be generated for single and multiple geopolitical jurisdictions, such as across federal and state, or state to state. A user therefore can identify what agencies at federal and state agencies relating to management of a topic. Take, for example, the case in which state agency personnel are involved in designing a new plan of management for aquatic invasive species. They may need to identify the following:

- The other California state agencies, if any, involved in the topic so that resources can be shared and relevant plans can be coordinated, or gaps in management can be identified and filled strategically.

[26] The beta version of *MINOE: Management Identification for the Needs of Ocean Ecosystems* is available for download at http://minoe.stanford.edu .

- The federal level agencies involved, if any, and through what laws and regulations

To begin, the user defines the concept of interest using terms and phrases. A user also may create more advanced definitions of the concept with Boolean search capabilities to capture synonyms, as shown below:

"live bait" (+import +live) "fishing bait" "live freshwater bait"

The user may enter multiple or single topics of interest in the user interface, as illustrated in Figure 4.

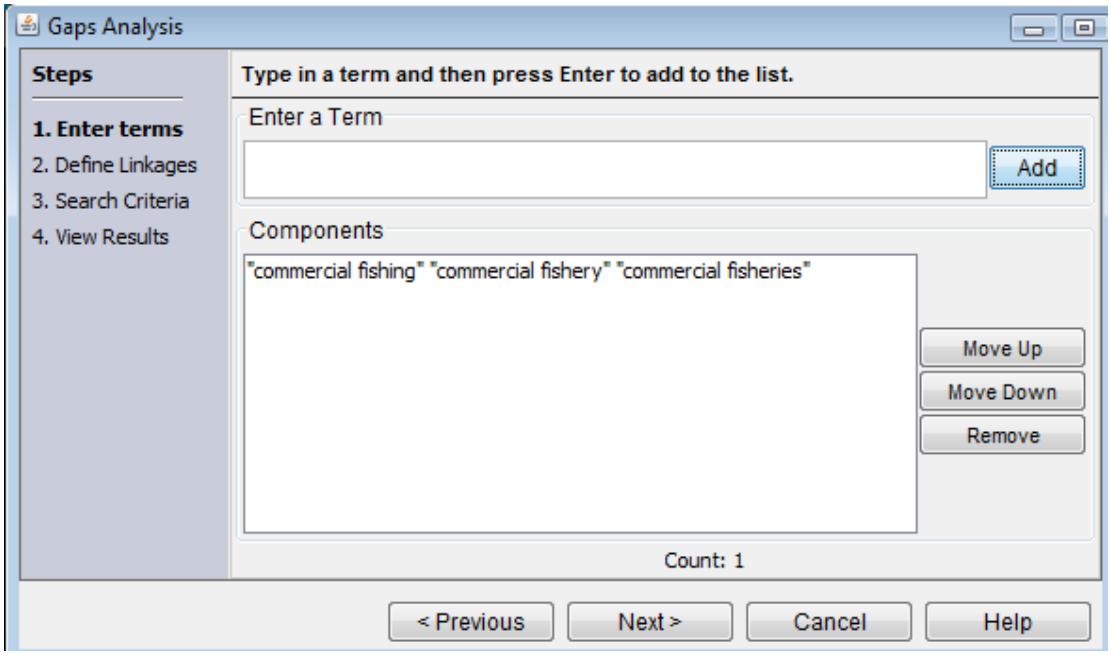


Figure 4. Form for user to input terms, phrases or Boolean search to define each concept of inquiry

Then a window opens to allow the user to select the search criteria (Figure 5). These include the geopolitical jurisdiction and the document scope. Currently the application contains four geopolitical jurisdiction options with three states (Washington, Oregon, and California) and federal United States. The document scope currently includes codified statutes and regulations.[27]

[27] Julia A Ekstrom, "California Current Large Marine Ecosystem: Publicly available dataset of state and federal laws and regulations" *Marine Policy* 33, no. 2 (2009): 532-535.

Julia A Ekstrom and Gloria T Lau, "Exploratory text mining of ocean law to measure overlapping agency and jurisdictional authority," *Proceedings of the Digital Government Research Conference*, (Montreal, Canada 2008).

The user also may select a specific document or group of documents to include in the analysis, as a filtering option.

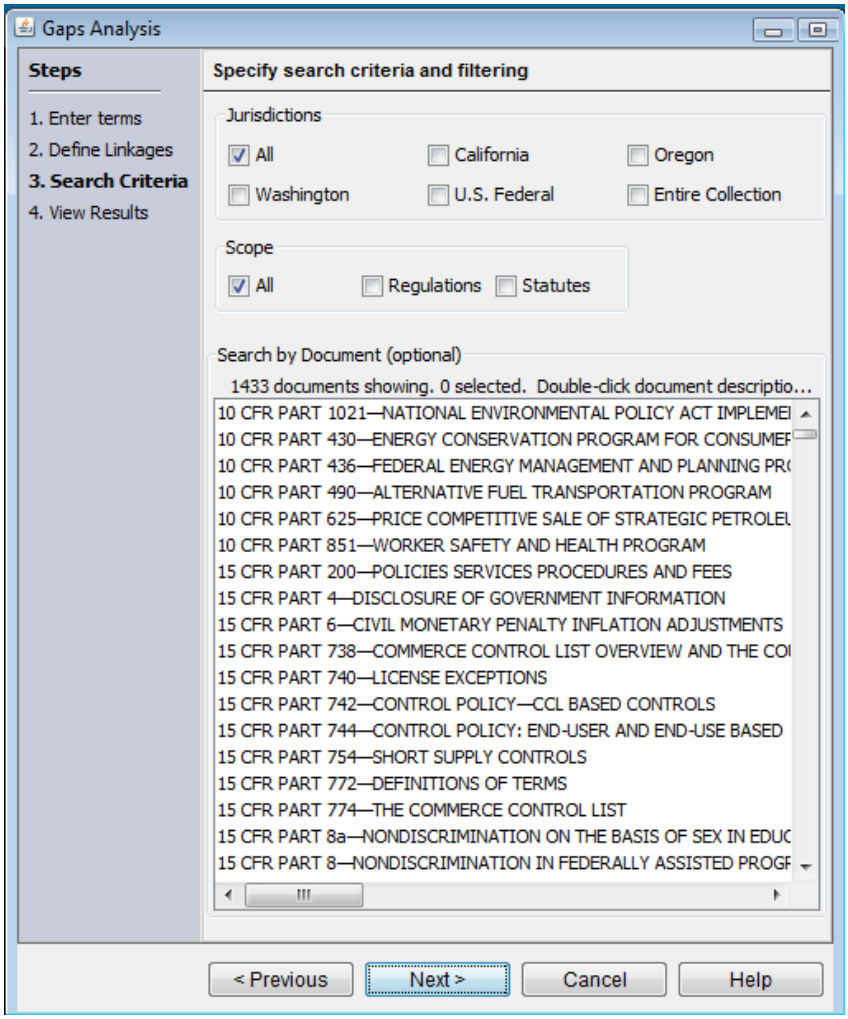


Figure 5. MINOE's filtering options

Once filtering criteria are selected, the next interactive window provides a portal to the agency involvement metric and the text of the corresponding laws.

The first screen, as shown in Figure 6, contains the total number of sections in which each concept occurs. A user may view each by clicking the labeled tab, providing a user-friendly way to compare between jurisdictions.

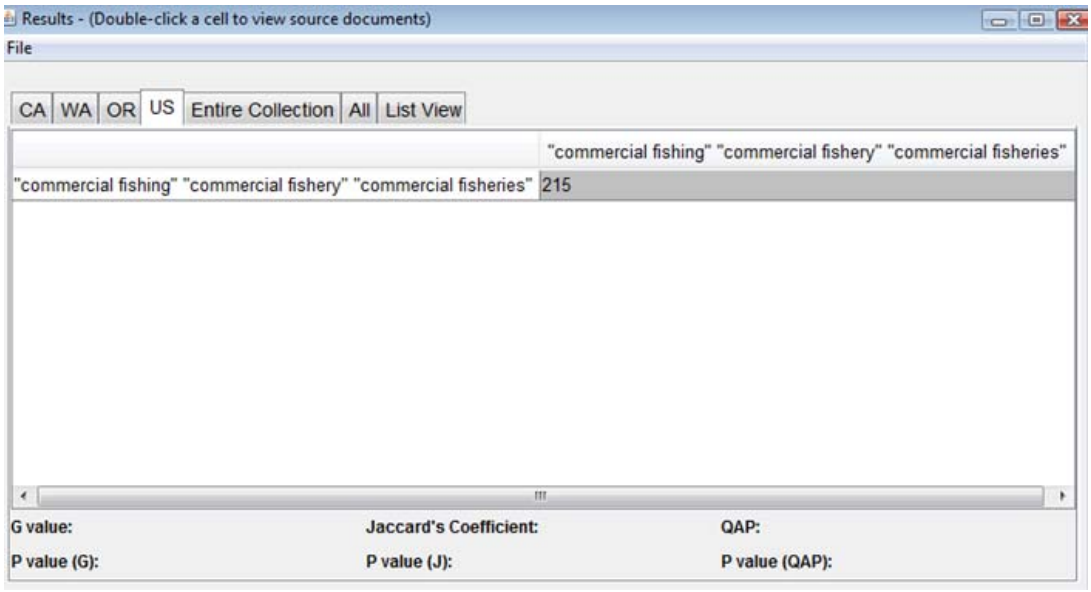


Figure 6. Display of MINOE's initial results screen. Cells contains number of sections of law that contain the single concept.

From this screen (Figure 6), the user can right click on any cell to generate the agency involvement bar graphs, as shown in Figure 7. Agencies are represented along the x-axis and the frequencies of sections of law ("Sections") are represented along the y-axis. To view the full name of an agency, a user hovers the pointer over any of the bars. To view what laws make up each bar, the user may click on the relevant bar, which generates the corresponding list of document names. The number of sections containing the concept is displayed in parentheses following the document name. For example, the box in Figure 8 highlights the Salmon and Steelhead Conservation and Enhancement Act contains eight sections in which the concept of *commercial fishing* occurs (as defined by L1 terms).

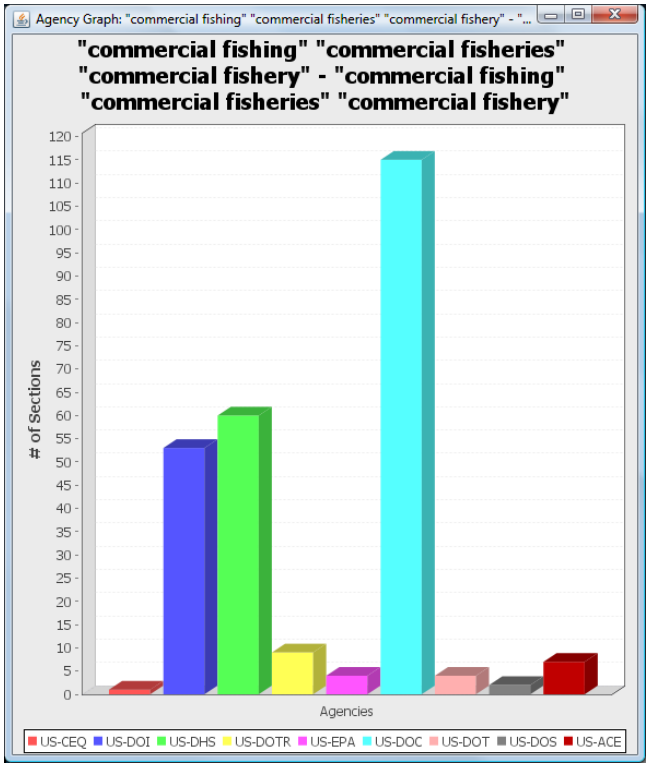


Figure 7. Bar graph gauging agency involvement (based on number of sections containing topic) for topic of commercial fishing without use of taxonomy

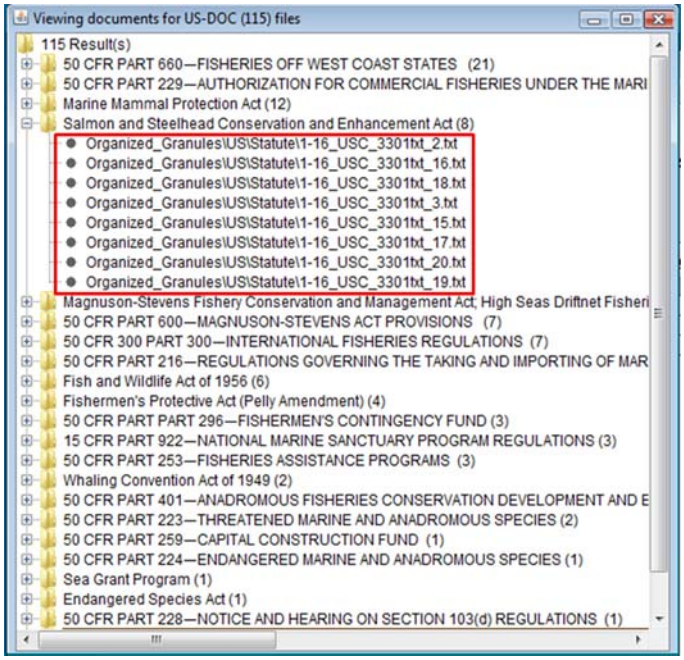


Figure 8. Screenshot of MINOE's law and regulation text viewing and searching feature.

The list of displayed law names may be opened to view the list of individual sections, which then may be selected to view the text. Once the text of the regulation is open, there is a tool bar in the upper left corner that allows the user to search any keyword or phrase (Figure 9). Every occurrence of the search keyword or phrase is highlighted.

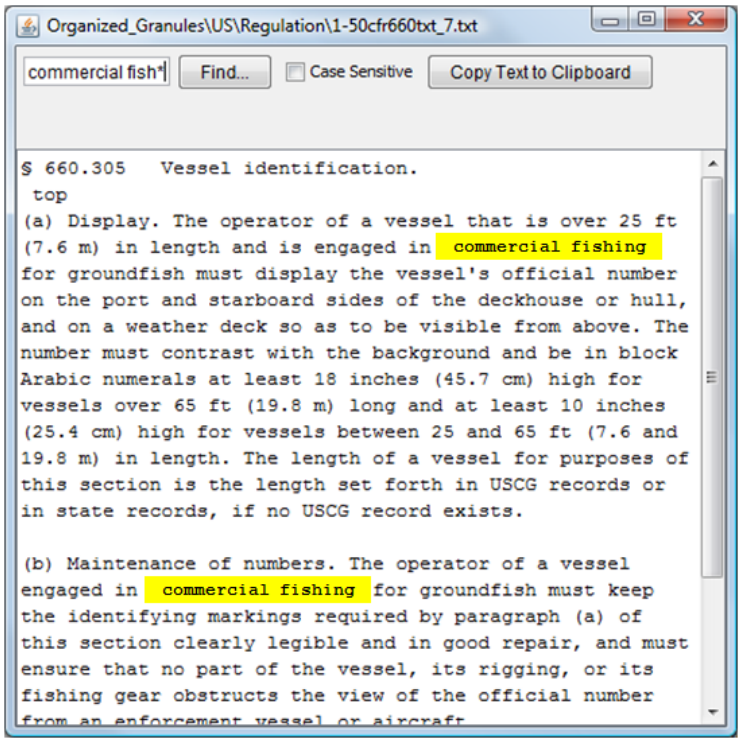


Figure 9. Screenshot of MINOE's regulation text viewing and searching feature.

4.2 Application and Utility

We recognize no application will replace the domain expertise of an experienced practitioner for understanding how an activity is managed. However, given that the application produced highly accurate results of top most involved agencies for state and federal levels of management, our tool does provide a myriad of benefits even in its basic form. Such benefits may assist domain experts as well as non-experts as a first step to further investigation:

- Provides objective estimation of agency involvement using a suite of laws and regulations from comprehensive suite of sectors
- Facilitates inclusion of terminological taxonomies through Boolean search capabilities
- Provides information in a transparent manner, allowing easy access to the source text for determining context

There are several potential users for a system that automatically retrieves agency involvement information, including the government, non-governmental organizations, policy advocates, resource users, concerned citizens, and policy course instructors and students. As a tool for agencies, it could be used to assist in improving collaboration, enhancing strategic resource sharing, and increasing strategic policymaking. Such a metric could assist lawmakers determine what agencies should be involved in a proposed law or policy based on an emerging issue such as ocean acidification or climate change adaptation. The technique could also be useful for individual agencies at a survey level in writing new regulations to determine whether there are resource-sharing opportunities with other agencies in order to fulfill mandates.

In terms of invasive species management, the preliminary technique with further development and verification could be used to help identify what agencies play a role in regulation of entrance pathway-related activities. Beyond government uses, non-governmental organizations and other issue-focused advocacy groups could use such a tool to determine what agencies and through what laws could be targeted for lobbying in order to fulfill the groups' goals. Other types of stakeholders, including resource users or community members may find such a tool useful in that it could promote accessibility to how various issues are managed, and thus helping to provide a pathway for promoting increasing public participation in decision-making.

4.2.1 Tool for building management scenarios

The increasing interest and need for applying the ecosystem concept in management and decision-making have generated a growing number of scenario building tools and research programs focused on valuing ecosystem services.[28] These tools thus far focus on the natural science of ecosystems and are beginning to integrate these data with economic information. However, until now no information retrieval policy or regulation tools have been developed that have capacity to integrate and interoperate with these ecological and economic EBM tools.

Ecopath with Ecosim (EwE), the former of which was originally developed by Jeffrey Polovina and colleagues at the National Marine Fisheries Service,[29] is being built by a team of scientists and software engineers at University of British Columbia to assist in identifying optimal

[28]NatureServe, "Ecosystem-Based Management Tools Network (http://www.ebmtools.org)," (Accessed 2009).

[29]NOAA, "ECOPATH Modeling: Precursor to an Ecosystem Approach to Fisheries Management http://celebrating200years.noaa.gov/breakthroughs/ecopath/welcome.html," ((Accessed 2009)).

management strategies for fisheries management.[30] EwE creates quantitative ecosystem models of direct and indirect linkages between species, habitats, and other ecosystem elements. Using a suite of management scenarios, the EwE can help scientists and managers identify what aspects of the economic supply chain will be affected by various ecosystem changes. Our application presented in this paper could be added to provide users with useful information about the existing legal framework in which ecosystem elements are being regulated, whether the ecosystem linkages are accounted for in any regulations, and what agency or suite of agencies should be involved in the implementing the scenario.

The Natural Capital Project (NatCap), based out of Stanford University in collaboration with The Nature Conservancy and the World Wildlife Fund, has developed a software tool that calculates an economic cost benefit analysis of management decision scenarios based on valuation of ecosystem services.[31] Using the optimal scenarios generated by the NatCap software, MINOE could assist users to retrieve baseline information about what agencies (and through what laws) are involved in the variety of ecosystem elements that pertain to the scenario. This could help users identify most feasible scenarios to follow and, combined with further research, could assist development of policy recommendations to implement the scenario.

4.2.2 Tool for policy makers

The presented tool can also be used for research to identify discrepancies between law and policy. For example, federal environmental regulation may lag behind state environmental policy where states are more environmentally progressive than federal laws. With the new administration, digital government tools, such as MINOE presented in this paper, could be useful to help systematically identify such mismatches. Although further in-depth evaluation would be necessary for policy development, it could assist in the first level of analysis for the federal government to progress its environmental policy.

---

[30]Villy Christensen and Carl Walters, "Ecopath with Ecosim: methods, capabilities and limitations" *Ecological Modeling* 172 (2004): 109-139.

Daniel Pauly, Villy Christensen and Carl Walters, "Ecopath, Ecosim, adn Ecospace as tools for evaluating ecosystem impact of fisheries" *ICES Journal of Marine Science* 57, no. 3 (2000): 697-706.

[31]Gretchen C. Daily and Pamela A. Matson, "Ecosystem services: From theory to implementation" *Proceedings of the National Academy of Sciences* 105, no. 28 (2008): 9455-9456.

4.3 Future Work

There are several prospects for applying the tool presented in this paper in identifying agency involvement. However, the tool is only in its beta version and requires further exploration and expansion of features. In the future, we plan to:

- Explore other parameters that represent management besides law and regulation: E.g. Align with economic data to determine feasibility of tracking implementation through budget allocations

- Apply technique to management plans

- Apply technique to geospatial boundary data of legal jurisdictions. In some cases a law pertains to a specific region or habitat, such as freshwater and not marine. Therefore, incorporating the spatial or other type of zonal tagging would assist in improving the accuracy of the agency involvement metric in development

- Include taxonomy with weighting of terms in lower levels based on each term's relatedness to defining the concept

- Elaborate capacity of user queries with Natural Language Processing algorithms, including investigating meaning implied by the language used in the documents

Through analysis of preliminary surveys and text analysis, we found that of the three frequency parameters used to measure agency involvement (document frequency, section frequency, and term frequency), document frequency was the least accurate metric. In addition, use of lower level terms from our taxonomy was shown useful for some topics and misleading for others. Therefore, we developed a tool that allows a user to input single or multiple levels of terminological taxonomies to retrieve information for a single topic. In addition, we developed the system using the section frequency as the metric to identify agency involvement. Upon further development, there are several opportunities to apply this open source tool, highlighting the need for digital government research to develop tools that can assist improvement of natural resource management.

5. ACKNOWLEDGMENTS

Our thanks to the David and Lucile Packard Foundation (Ecosystem-Based Management Tools Initiative Fund) for supporting this research. The authors would also like to acknowledge partial support by the National Science Foundation grant IIS-0811460. Any opinion expressed in this paper are those of the authors and do not necessarily reflect the opinions of the David and Lucile Packard Foundation or the National Science Foundation.