ECOLENS: INTEGRATION AND INTERACTIVE VISUALIZATION OF ECOLOGICAL DATASETS

Cynthia Sims Parr[1], Bongshin Lee[1,2*] and Benjamin B. Bederson[1,2]

[1]Human-Computer Interaction Lab, UMIACS, University of Maryland, College Park, USA and [2]Department of Computer Science, University of Maryland, College Park, USA

Direct correspondence to: Cynthia Sims Parr, csparr@umd.edu

ABSTRACT

Complex multi-dimensional datasets are now pervasive in science and elsewhere in society. Better interactive tools are needed for visual data exploration so that patterns in such data may be easily discovered, data can be proofread, and subsets of data can be chosen for algorithmic analysis. In particular, synthetic research such as ecological inter-action research demands effective ways to examine multiple datasets. This paper describes our integration of hundreds of food web datasets into a common platform, and the visualization software, EcoLens, we developed for exploring this information. This publicly-available application and integrated dataset have been useful for our research predicting large complex food webs, and EcoLens is favorably reviewed by other researchers. Many habitats are not well represented in our large database. We confirm earlier results about the small size and lack of taxonomic resolution in early food webs but find that they and a non-food-web source provide trophic information about a large number of taxa absent from more modern studies. Corroboration of Tuesday Lake trophic links across studies is usually possible, but lack of links among congeners may have several explanations. While EcoLens does not provide all kinds of analytical support, its label and item-based approach is effective at addressing concerns about the comparability and taxonomic resolution of food web data.

Keywords: food webs, visualization, data integration, taxonomy

INTRODUCTION

Ecologists are performing more meta-analyses and many researchers are integrating their datasets to achieve analyses that are unparalleled in geographic, historic, and topical scope (e.g., review by Storch and Gaston, 2004, publications from NCEAS http://nceas.ucsb.edu). Integrating large numbers of datasets is fraught with pitfalls, and problems are difficult to catch: have like elements been properly merged, have all inconsistencies among datasets been recognized and handled, are appropriate metadata available for further assessment of dataset quality? Furthermore, exploration of these multiple datasets for trends or testable patterns prior to statistical analysis is tedious, as it relies on complex SQL queries, spreadsheet macros, or specialized applications.

Data sharing itself poses an additional set of problems. A corpus of data is typically chosen by investigators for a particular purpose, and then made available to others, perhaps in a data clearinghouse. However, other investigators may have different criteria for their own purposes and may need only a subset, or they may choose an intersection of two or more corpuses; with some datasets appearing in more than one corpus. These are problems faced by any domain where multiple data sources must be explored and selected for further analysis.

In this paper, we are particularly interested in addressing these problems for ecological interaction analysis. The food web research community has a long history of data sharing and integrated analyses (reviewed in Pimm, 2002). Here the datasets typically involve networks of organisms and their trophic relationships, as well as associated population characteristics, flows, and organismal attributes. While there are clearly trophic relationships, or links, within datasets, many organisms have been studied in multiple places by multiple researchers in the same or different habitats, and so there are relationships across datasets as well.

Though graph visualizations are often used in network analyses (Lima, 2006), these have typically focused on visualizing one network at a time. They emphasize the nature of the linkages among nodes in a particular network. Most food web visualizations use a node-link diagram, laid out in 2-D or 3-D space (e.g., Christian and Luczkovich,

---

1999; Dunne et al., 2006). Primary tasks they support are identifying clusters and the distribution of node or link attributes across these clusters. Where multiple webs are available for visualizing, they must be viewed one at a time with no support for choosing which web to view.

PaperLens (Lee et al., 2005), winner of the 2004 InfoVis contest (Fekete et al., 2004), and its successor NetLens (Kang et al., in press) illustrate an alternative approach. PaperLens was designed to allow analysis of trends in research publications and exploration of topics, authors and other publication metadata. It provides easily sorted and scrolled tables, whose items are coupled to relevant items in related tables. Linkages are revealed primarily by interaction with these tables, but also by a "Degrees of Separation Links" diagram. Information is summarized into bar charts, and also linked to the items in tables that generate the bars. PaperLens was designed for exploration of a digital library, not for visualizing scientific data.

In this paper, we describe the selection and integration of datasets. Then, we describe EcoLens, an enhanced version of PaperLens that provides effective filtering, querying, and visualizing of multiple ecological interaction datasets. We give examples of results gained by using it on the integrated database and the results of a qualitative evaluation of EcoLens. Finally, we summarize lessons learned and propose new directions for tool development.

METHODS

1 DATABASE DEVELOPMENT

1.1 Data Requirements

The goal of our theoretical ecology research is to develop effective algorithms for predicting trophic links in a system where they are unknown, or where conditions and therefore the known trophic linkages will change. At present we focus on presence or absence of links, i.e. the basic network topology critical for more sophisticated network analysis and modeling projects such as Christian and Luczkovich (1999). Our approach (Parr, in prep) is to take advantage of large numbers of known trophic links and use similarity in attributes or evolutionary relationships to predict whether links exist among organisms whose trophic links have not been studied. Thus, the datasets to integrate include studies of food webs throughout the world, including the names of organisms, their links, and metadata about each of these studies such as when, where, and in what habitat it was conducted. Furthermore, our algorithms require information about the evolutionary relationships among organisms, and the attributes of those organisms. The database should be maintained online in order to better integrate with SPIRE forecasting tools (http://spire.umbc.edu, e.g., Parr et al., in press), and potentially for integration with our other project datasets and tools.

1.2 Data Sources and Integration

Below we describe the original data sources and our process of modifying them to integrate into our database. A current MySQL database schema is available at http://www.cs.umd.edu/hcil/biodiversity.

Taxonomy and evolutionary information
We followed the integrated classifications as in Parr et al. (2004) for animals and ITIS (2006) for plants and other organisms. These compilations of multiple sources provide an internally consistent source of names and allow us to use phylogenetic or taxonomic relationships among food web nodes in our other analyses (in prep). Information from these sources did not need to be modified in order to be integrated. We made some effort to identify and replace synonyms in the food web data with current names using these sources. Typographical errors in food web node names were also fixed wherever they were identified.

Ecological interaction data
We obtained ecological interaction data from online repositories in the most machine-readable format (we could obtain usually ASCII files, but occasionally MS Excel spreadsheets or PDFs). We focused initially on integrating large multi-dataset sources which have already been subject to multiple analyses, such as Cohen's EcoWEB (Cohen, 1989), the trophic webs at the NCEAS Interaction Web Database (Vazquez, 2005), and the Webs on the Webs corpus (Dunne et al., 2006). We also included two webs specifically to see how EcoLens can handle taxon list comparisons (Jonsson et al., 2005). We emphasize that these are merely a starting point and not intended to be wholly representative of the data available. Integration of webs was achieved primarily on two dimensions, habitat and organism names. Habitat categorization was determined by manual inspection of original data files or published studies. It followed the moderately rich biome categories used by the Animal Diversity Web ontology (Parr et al., 2005).

Mapping food web entities (nodes) to the scientific names in our evolutionary data involved 1) moving modifiers such as age and size classes and other descriptors to other database fields; 2) searching taxonomic databases for exact or approximate matches to scientific names; and 3) determining the most appropriate scientific name or names

for a common name. Step 1 was accomplished largely by scripts written in Java (available upon request), steps 2 and 3 were handled first by scripts accessing our own database information, then remaining typos and synonyms were handled by manual inspection and searching of external sources such as FishBase.org or Google.com.

In some cases, a node name needed to be mapped to multiple taxonomic names. For example, "birds of prey" becomes Falconiformes and Strigiformes; "foxes" becomes the individual fox species or genera known to occur in this geographic location. This process is in effect the opposite of constructing trophospecies – where a name is found for a group of taxa that all share the same trophic link. Trophospecies can be sufficient for understanding trophic relationships and a way to avoid problems with taxonomic resolution (reviewed in Dunne, 2005), but de-aggregating trophospecies into taxa is critical for integrating data across multiple food web datasets. We mapped to the narrowest scientific name possible to include all the likely instances in the food web.

For those nodes where a taxonomic name was not possible to assign, we mapped to a controlled vocabulary. For example, Dissolved Organic Matter in one food web and DOM in another were both mapped to the same name, but DOM and POM (Particulate Organic Matter) were mapped to different names. We will refer to these also as "taxon" names though of course these are not evolutionary units or groupings.

With de-aggregated taxa, it is necessary to de-aggregate links. When a trophic link was reported between two nodes, and one or both of the nodes maps to more than one taxon name, we assumed that there was a link among all the resulting taxa. Webs from Jonsson et al. (2005) were already provided in both aggregated and de-aggregated forms.

Non-web trophic information
Most food web research involves trophic links reported in the context of source or sink or population webs. Source webs include a basal organism and all the organisms that eat it and the relationships among them. Sink webs include a top predator and all the organisms it eats. Population webs include a community of organisms and all the links among them. We used categorizations from the original sources to indicate each web's type. In addition, our schema allows evidence of trophic links that do not come from web studies at all, but from sources that report only lists of prey for a given predator or lists of predators for a given prey or general food habits. This type of information is readily available in online encyclopedias and greatly increases the scale of available knowledge in terms of taxa, habitat, and geographic coverage. It augments the more comprehensive food web studies.

Food web attributes
Food web researchers often compare overall web characteristics, such as number of species or nodes (S), number of links (L), and connectance (L/S^2). We included the published or original values for the webs, and others we calculated based on our new de-aggregated taxa and links. Though other quantitative attributes of webs are possible to calculate (Pimm, 2002), we did not attempt to do so because of known lack of comparability of these measures across datasets collected under diverse conditions. Quantitative link strength or flow measures were not possible to compute because we currently use only presence/absence data which is more widely available. Given our newly mapped taxon names, we were able to compute the percentage of each web's taxa which were species (or subspecies), above species, or unknown (either the entity is not truly taxonomic or its level could not be determined).

Taxon attributes
The common name and rank of each organism were obtained from the taxonomic sources described above. To demonstrate how natural history characteristics could be integrated, we downloaded maximum mass from Animal Diversity Web (Myers et al., 2006) using their advanced inquiry search. Finally, we determined from our data tables the number of food web studies in which each organism appears.

2 ECOLENS SOFTWARE DEVELOPMENT

### 2.1 Visualization Requirements

The goal of EcoLens is to allow biologists to explore a collection of food webs, find webs of interest, and then visualize an individual food web. As described above, our data consists of several elements such as food web study details, taxa, and habitats. Inspired by our successful experience with PaperLens, we tightly coupled multiple views to show relationships among these data elements. Within each food web, trophic relationships among taxa are important as well. Therefore, we wanted our design to combine the PaperLens overview technique with a guiding metaphor proposed in TreePlus (Lee et al., in press): "Plant a seed and watch it grow." Using this philosophy, users start with a specific node and incrementally explore the network, avoiding complexity until it is necessary. Through the overview, users can easily find not only interesting trends and patterns in the dataset but also particular webs of interest. Once they find desired webs to look at, they can investigate each food web. We consider labels essential for both overviews and details, while the need to see every single item in a single overview is not important. While

EcoLens is implemented with a food web dataset, we aimed to support a variety of general tasks having to do with understanding multiple datasets and their integration.

2.2 Evaluation Methods

For internal evaluation, we constructed a list of questions that EcoLens might help a biologist answer. We then tried using EcoLens to answer these questions and reported the results in the Dataset characteristics section below.

For external evaluation, we asked ten ecologists to use EcoLens several times and fill out a survey. Four responded. These ecologists had contributed data and were asked to evaluate the mapping of the food web nodes to taxonomic names. We also asked them specific questions about the interface. The survey included both Likert-scale questions and open-ended questions. This kind of formative qualitative evaluation is not expected to demonstrate clear advantages over existing systems but provide insight into advantages worth quantitative study.
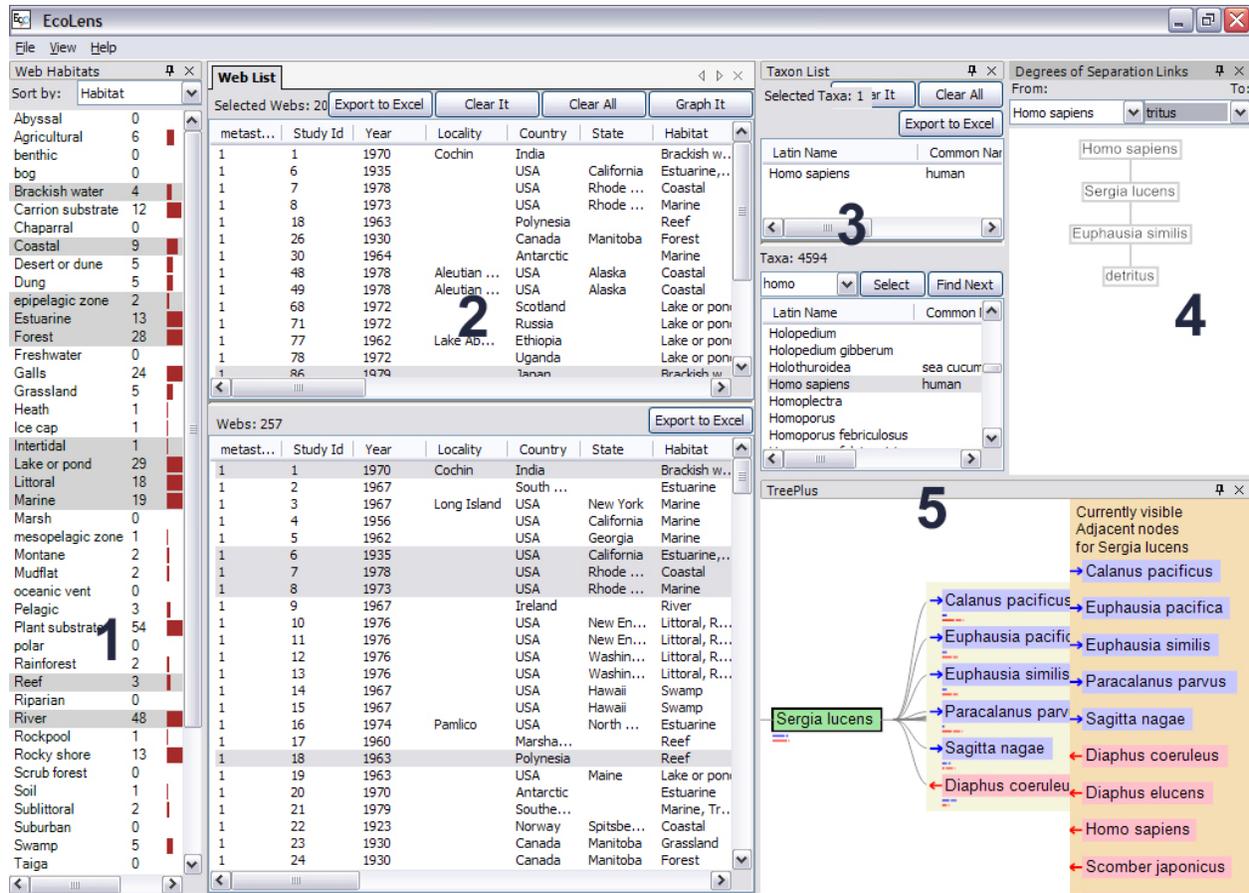

Figure 1. EcoLens provides easy exploration of relational data tables by sorting and selecting in tabular form from complete lists to selected lists (for both web list and taxon list views, the list at the bottom is complete list and the one at the top is slected list), coupled with graphical representations in a bar chart (1), degrees of separation view (4), and network visualization (5).

2.3 System Description

As shown in Fig. 1, EcoLens consists of five main views: 1) web habitats; 2) web list; 3) taxon list; 4) degrees of separation links (DOSL); and 5) TreePlus. The web habitats view shows the list of habitats with the number of food webs in each one. Users can sort the view either by habitat name or by the number of webs. The bottom of the web list view shows all the webs in the database. When some webs are selected either by users or by the system, they are shown in the Selected Webs list at the top of the view. Similarly, the bottom of the taxon list view shows all the taxa in the database and the currently selected taxa are shown in the Selected Taxa list. When users double click on a taxon in the Selected Taxa list, EcoLens opens a dialog box to show the list of studies that contain the selected taxon

(Fig. 2). The TreePlus view visualizes the food web as a node-link, tree-like diagram and the DOSL view shows one of the food chains from one taxon to the other in the web currently being visualized by the TreePlus view.
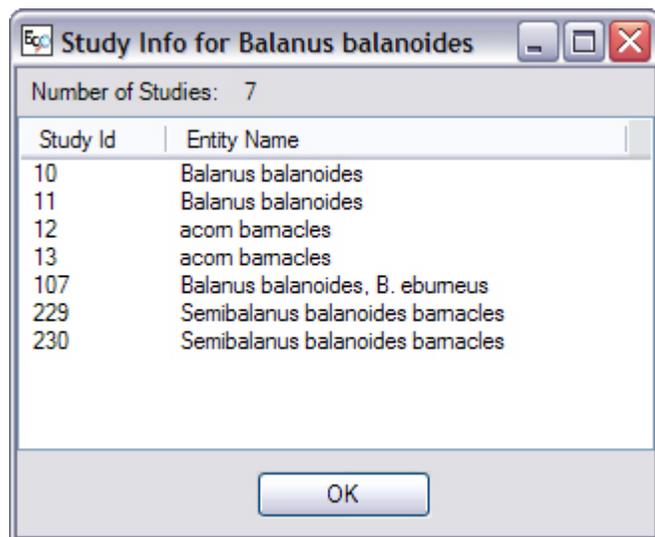
Figure 2. When users double click on a taxon, <u>Balanus balanoides</u>, in the Selected Taxa list, EcoLens opens a dialog box to show the list of studies that contain the selected taxon.

These views are tightly coupled. When users select a habitat in the web habitats view, all the webs from the selected habitat are highlighted in the Webs list. Furthermore, they are displayed in the Selected Webs list for easy access. In addition, all the taxa in these webs are highlighted in the Taxa list and displayed in the Selected Taxa list. For these three views – web habitats, web list, and taxon list – user interactions are symmetric. For example, users can select webs from the Webs list to see habitats for particular food webs or get lists of taxa. Habitats of the selected webs are highlighted in the web habitats view and taxa in the selected webs are shown in the taxon list view. Users can also copy reference information of the selected food webs to the clipboard by selecting the Copy Reference menu option after right clicking on the selected food webs.

Users may visualize an individual food web in the TreePlus view to see trophic links among taxa by double clicking on a web in the Selected Webs list or in the Webs list. They can also press the "Graph It" button after clicking on a web in the Selected Webs list. EcoLens then builds a food web from the database and visualizes it using TreePlus. Since it uses the default root selection mechanism in TreePlus, the taxon with the most connections to others is chosen to be the root. EcoLens also adds all taxa in the web to the "From" combo box in the degrees of separation links view. Once a taxon is selected from the "From" combo box, EcoLens displays all the taxa reachable from the selected taxon in the "To" combo box with the corresponding degrees of separation. When a taxon is selected from the "To" combo box, EcoLens displays one of the shortest food chains between two taxa. When users click on a node in the degrees of separation links view, EcoLens opens the selected taxon within TreePlus. Similarly, when users click on a node in the TreePlus view, EcoLens highlights the selected taxon within the degrees of separation links view if it is already displayed.

Based on requests from users, an "Export to Excel" feature was implemented for the web and taxon list views. This launches Microsoft Excel with the table data contained in the list where the button was used. For web lists, averages and standard deviations are automatically calculated for web statistics.

EcoLens was implemented in C# and runs on any standard Windows PC. To visualize each food web, it uses TreePlus (Lee et al., in press), a reusable component we developed to visualize networks. The web habitats and degrees of separation links views are implemented with Piccolo.NET, a shared source toolkit that supports scalable structured 2D graphics (Bederson et al., 2004). EcoLens accesses the MySQL server using a MySQL data provider, which links a data source and .NET code. To make each window dockable, it uses the DockPanel Suite (Luo, 2006), an open source docking library. EcoLens is now available for download at http://www.cs.umd.edu/biodiversity/#EcoLens.

2.4 TreePlus

TreePlus is an interactive graph visualization component based on a tree layout approach. It transforms a graph into a tree plus cross links (i.e. the additional links that are not represented by the spanning tree) and visualizes the tree instead of the graph. TreePlus uses a guiding metaphor of "Plant a seed and watch it grow." This allows users to start with a node and expand the graph as needed.

TreePlus reveals the missing graph structure with visualization and interaction techniques while preserving good label readability. It highlights and previews adjacent nodes when a node is focused by a single click (Fig. 3). TreePlus updates the tree structure when a node is opened by a double click. TreePlus carefully animates the transitions[1]

---

[1] See video demonstration at http://www.cs.umd.edu/hcil/treeplus.

so that users can follow changes. The color of the node background and arrows indicates the link direction relative to the focus node. TreePlus uses the color blue for outgoing links, red for incoming links, and purple for bidirectional links. For example, in Fig. 3, the red node (Homo sapiens) eats Todarodes pacifus while Todarodes pacifus eats blue nodes (e.g., Sergia lucens). TreePlus also provides users with the option to show preview bars representing how fruitful it would be to go down a path. Color bar graphs placed below the nodes represent how many organisms are reachable in each direction.
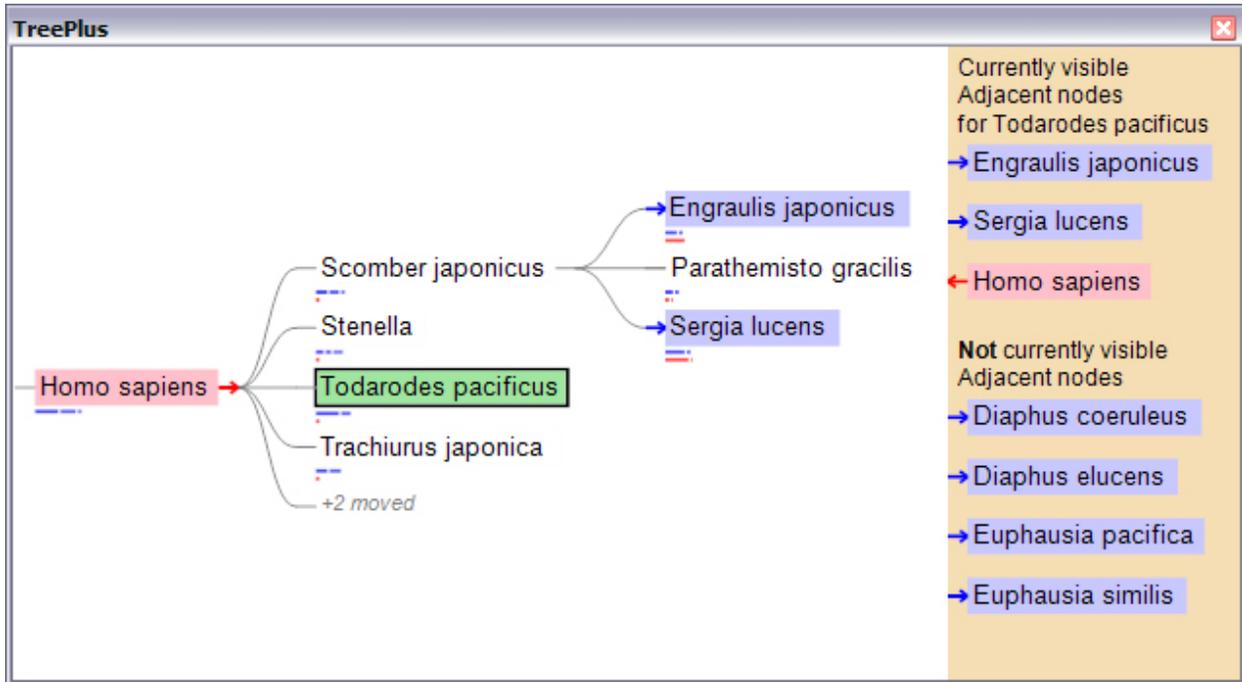
Figure 3.. Homo sapiens was set as the root, and users selected Scomber japonicus which added all its adjacent nodes to the tree. A single click on Engraulis japonicus gives it the focus and shows a preview of its adjacent nodes in the preview panel on the right. Red or blue color indicates the direction of the link.

3 RESULTS

3.1 Using EcoLens to Explore Food Web Data

These results about the database were obtained using EcoLens and simple spreadsheet manipulations. Trends identified here could subsequently be analyzed statistically but that is beyond the scope of the present exercise.

Example 1: Obtaining an overview of the integrated data
Using EcoLens, it is possible to quickly get an overview of the integrated dataset (Table 1). The most frequently studied habitat in our database is plant substrates, including data from 13 countries involving 543 distinct "taxa." 28% of the 4594 distinct taxa are found in more than one study. It is obvious that studies have increasingly included more taxa and links, and that early webs appear to need more de-aggregation than later webs (Table 1). Across webs, a wide variety of taxa have been studied. "Detritus" is the most often included "taxon" in food webs (N=140). Humans appear in 20 different food webs, and two of these are prehistoric food webs.

Table 1. Summary of integrated data in this study. There were 4594 distinct taxa overall, due to overlap across studies. Deaggregation involved breaking original nodes into multiple taxa as appropriate, but also involved collapsing nodes where age and size classes for the same taxa had been reported separately. Original data recompiled from Cohen, 1989; Dunn et al., 2006; Jonsson et al., 2005; Myers et al., 2006; Vazquez, 2005.

Original data source (publication years)	N webs	Original N nodes	Original N links	Mean % species per web (SD)	De-aggregated N nodes	De-aggregated N links	Distinct "taxa"
EcoWEB (1923-1988)	213	4064	8295	39 (31)	4503	11967	2512
Webs on the web (1989-2000)	15	1366	9412	53 (29)	1373	12056	1149
Interaction Web Database (1998-2003)	26	2185	9882	55 (8)	2139	9882	479
Animal Diversity Web (N/A)	n/a	713	3025	80	711	2165	890
Tuesday Lake 2005	2	41	127	82 (3)	101	510	71
Totals	256	8369	30741		8827	36580	5101

Example 2: Answering specific predator-prey questions

What organisms are trophically linked to blue crabs, <u>Callinectes sapidus</u>? This species appears in five data sources which report similar though nearly exclusive suites of predators and prey (Table 2). Some organisms are both predator and prey though each case is reported by only one study.

Table 2. Predators and prey for blue crab (Callinectes sapidus) across five studies (Baird and Ulanowicz, 1989; Christian and Luczkovich, 1999; Kemp et al., 1977; Odum and Heald, 1975; Peterson, 1979) after recompiling and mapping to taxa in the EcoLens database. All organisms were mentioned in only one study except for those mentioned in two studies, with (2) after the name.

Organism type	Predators	Prey
Bivalves	Mya arenaria, Crassostrea virginnica	Bivalvia (2), Modiolus demissus, Urosalpinx cinerea, Mya arenaria, Crassostrea virginnica, Macoma, Brachiodontes exustus
Snails		Littorina littorea, Chione cancellata, Crepidula fornicata
Crustaceans	Crustacea	Crustacea (2), Neopanope texana (2), Alpheus normani, Palaemonetes floridanus, Cumacea, Decapoda, Hippolyte zostericola, Libinia dubia, Pagurus, Pagurus maclaughlinae
Polychaetes	Polychaeta, Nereis	Polychaeta (2), Nereis
Plankton		zooplankton, phytoplankton, microzooplankton
Fish	Actinopterygii, Sciaenops ocellatus	Actinopterygii, Floridichthys carpio, Lophogobius cyprinoidea, Penaeus duoarum
Birds	Aves , Larus argentatus, Larus atricilla Sterna forsteri, Larus delawarensis	
Nematoda		Nematoda
Cnidarian		Chrysaora quinquecirrha
Ctenophora		Ctenophora
Other	other suspension feeders	benthic invertebrates, other suspension feeders, meiofauna

How often do two taxa shown to have a trophic link in one web have a trophic link when they co-occur in another web? We examined datasets from Tuesday Lake collected in two different years, 1984 and 1986 (Jonsson, 2005). The datasets had 51 and 50 taxa respectively, but together they comprise 71 distinct taxa, so only 30 taxa occur in both datasets. We systematically examined the taxon list from the 1984 Tuesday Lake dataset to determine in how many cases a taxon occurs in some other dataset available to us or has a close relative (e.g., congener or less-well resolved taxon such as genus). Eight of 21 taxa not found in the other Tuesday Lake study were found in other studies in our dataset. Of those remaining, congeners were found in the companion dataset (N=5) or in other datasets (N=4). Thus, only four taxa found in the 1984 Tuesday Lake dataset are not found at least as a congener in another study. For closer examination, we chose one taxon in the 1984 Tuesday Lake dataset, Daphnia pulex, because it had many trophic links and it or congeners appeared in many other studies. Daphnia pulex links in the 1984 Tuesday Lake web are corroborated in the 1986 Tuesday Lake dataset by links among congeners as often as links among the exact taxa (Table 3). While exact links do not appear in other datasets, links among congeners are often reported. The large number of cases where congeners co-occur but are not linked suggests that these links were overlooked, that precise identifications may be necessary, or that trophic links depend on factors other than simple co-occurrence.

Table 3. Corroboration of trophic links of <u>Daphnia pulex</u> in Tuesday Lake 1984 dataset in a companion dataset (Tuesday Lake in 1986) and in other EcoLens datasets. "Same taxa" means both nodes in the trophic relationship are as identified in the 1984 Tuesday Lake dataset (e.g., <u>Daphnia pulex, Chaoborus punctipennis</u>). "Co-occuring congeners" indicates that one or both of the nodes in the trophic relationship are not the specific taxa in the 1984 Tuesday Lake dataset but are instead either a congener (e.g., <u>Daphnia rosea</u>) or genus level identification (e.g., <u>Chaoborus</u>). Other datasets were all of those including <u>Daphnia pulex</u> or a congener.

Corroboration type	Tuesday Lake 1986	Other datasets (N=7)
Same taxa are linked	11	0
Same taxa co-occur but are not linked	0	1
Co-occuring congeners linked	11	12
Co-occuring congeners not linked	0	26

Example 3: Assessing data quality and integration
Compared to a traditional database front-end, we believe that EcoLens provides an easier way to identify problems in an integrated dataset. While conducting the second analysis in Example 2, we found ten taxonomical spelling errors in while looking for related species in different datasets. For another example, a user might discover that there are two webs from the same publication, year, and locality. In a traditional database, several SQL queries would be needed to compare the species lists and links for each of these dataset. In EcoLens, a double click on the food web entry immediately brings up both a species list and a graphical view of the links. Indeed, one of our evaluators found that two of the webs had been erroneously duplicated because the same webs occurred in two different compilations.

Researchers have raised serious concerns about uneven taxonomic resolution of early food web studies (Cohen, et al. 1993; Jordan, 2003; Paine, 1988; Pimm, 2002; Polis, 1991). Indeed, EcoLens clearly shows that many of these early webs are poorly resolved to species and can support many of the conclusions reviewed in Dunne (2005). However, 40 (19%) of the 213 EcoWEB webs have over 75% of nodes resolved to species, only slightly less than the more recent webs (23%) in our compilation. Browsing the 2512 distinct taxa in the EcoWEB corpus indicates that trophic information for more than half of them (N=1596 taxa, including 1349 identified at least to genus) is not yet available from any other source.

How are the datasets consistent or not consistent? Using TreePlus, we discovered that one metastudy (Dunne et al., 2006) reports flows of energy from organisms to detritus with the same kind of link that other metastudies use only for feeding relationships (Cohen, 1989; Vazquez, 2005).

Our data includes trophic links from an online encyclopedia, ADW. This source is unusual because it is not a food web study, constrained by time and place and research project, but rather a broad survey of the literature on what links are possible. Of the 890 organisms whose feeding relationships are provided by this source, 710 are not found in any other source.

Example 4: Finding sample datasets for further analysis
Recent attempts to use a machine-learning approach to predicting food webs using biological taxonomy (Parafiynk, in prep) required us to pick a sample food web for initial software development and testing. A good sample web would be from a recent study (presumably high quality), with a large number of links, whose nodes are largely resolved to species, and in a habitat that has been well-studied historically so that training data would be available from other webs. The web with the largest number of links, Little Rock Lake in Wisconsin, is from a habitat for which we have many studies from 1955 to 2005. It has a low number of nodes identified to species (27%), however, consistent with other lake studies. Forest habitats might also provide good test set, though the webs are smaller in number of links. One of them (El Verde) fits many of our requirements. Indeed, the complex El Verde web also became our choice as we further developed and tested TreePlus, the graph drawing view (Lee et al., in press).

3.2 EcoLens Evaluation
Evaluators included four biologists: two post-doctoral fellows and two professors who spent 2 to 4 hours working with the program. While it is not appropriate with such a small sample to report statistical results, we highlight here the trends to illustrate strengths and weaknesses of our application. Overall, evaluators had a favorable response to the features of EcoLens, with scores generally falling between 7 and 9 on a 10-pt Likert-scale. Notable exceptions include only moderate interest in TreePlus preview bars (scores of 5 and 6). Other low scores include one respon-

dent who gave a 5 to degrees of separation links view, and one gave a 6 to the taxon list view when asked whether these views would be useful. Finally, we asked "Would you ever use a megaweb constructed by linking taxa from all these studies together?" During our design process we had decided that this would not likely be of interest to this community and indeed, only one of the four respondents expressed interest in it.

We also received qualitative comments on the nature of the datasets that illustrate what was and was not clear in the EcoLens representation. All of the biologists examined closely the webs that their own research had contributed to the compiled dataset. One discovered that two webs appeared more than once (having been contributed by more than one source). Another pointed out that one of the webs is actually not a single web but a composite across a number of different lakes. De-aggregation and standardization of taxa received praise from three of the biologists. The fourth does not focus on this aspect of food webs in his research and did not mention it. One biologist expected to see graphical representations of link attributes (e.g., strength, fluxes) or node attributes (e.g., body mass) which we do not currently show.

The biologists recommended adding interaction features, such as allowing categories other than habitats to be visualized in the bar chart. However, they were mostly interested in seeing additional data. In addition to the quantitative attributes for nodes and links mentioned above, they requested additional statistics for webs (e.g., connectance), more categories for webs (e.g., different habitat scales), and the ability to handle interactions other than trophic networks (e.g., pollination, parasitic, or seed disperser networks) where the sign of the interactions might be positive or negative.

4 DISCUSSION

With respect to the integrated dataset, we have several conclusions. Mapping food webs to standardized taxonomies remains a significant challenge. We accomplished this task only partly aided by custom software. However, once most of the food web nodes were mapped, benefits became apparent. It became possible to easily and fairly compare the taxonomic resolution of studies and also to consider the change in taxonomic resolution and scope of studies over time. Typographical errors in taxonomic data became much more obvious. We can more easily determine which studies are comparable based on their taxa and which are not and we can assess the degree of overlap between two studies. We can also assess similarity in datasets in a more detailed way by manually considering whether their nodes are mapped to near relatives. This similarity assessment problem is of broad interest in the ontology and semantic web research communities (e.g., Ehrig et al., 2004).

While EcoLens is more than a simple database table viewer, it is not an analysis tool per se. EcoLens makes it easy to identify promising trends that merit statistical examination, and to export at least some of the relevant data. Thus EcoLens illustrates the kind of exploration functionality that would be a valuable complement to management and analysis tools.

We do not claim that EcoLens provides the ability to handle all Food Web study tasks. It is ideal for sorting, filtering, and viewing subsets of the data. It does not provide, on its own, clustering capabilities. Identifying topological motifs such as cycles is not supported because its current graph visualization is TreePlus, which is designed for label-oriented and local topology tasks. Overview-oriented graph visualizations would complement the EcoLens approach.

The most difficult part of the implementation of the visualization was determining how to accommodate the complex schema in logically coupled views. For example, which of the bipartite schema elements should be available for a bar chart? In PaperLens (Lee et al., 2005), the authors chose a topic of papers and showed the number of papers in each topic, and in EcoLens we chose a basic unit of studies and showed the number of habitats. We made careful choices as to the direction of the coupling so that one view could drive another but that made it difficult if we wished the coupling to go the other direction (for example, to run down a list of selected taxa to see which other webs it appeared in).

Our approach will not be truly general until it is possible to easily map new schemas to the coupled elements. Lessons learned from EcoLens reported here informed development of a more generic and feature-filled application, NetLens (Kang et al., in press). In NetLens, the direction of coupling is under user control and bar charts can be constructed from most attributes of the bipartite schema. However, the complex mapping of nodes to taxa and taxa to studies with habitats and localities is too complex for NetLens to handle.

Future tool development should include displaying and exporting trophic link data. Matrix and list representations may be helpful for trophic link data. Support for set operations among lists (e.g., Kim et al., in press) would assist in comparisons. Practical mechanisms are needed for uploading new data (either from a whole web or collections of known trophic links). Support is also needed for the challenging process of mapping nodes to taxon names, though

improved taxonomic rigor in food web studies will eliminate the need for this mapping process. Users have expressed interest in being able to browse food web taxa graphically from a phylogenetic point of view, such as with TaxonTree (Lee et al., 2004). Similarly, habitats and biomes also occur in nested hierarchies so a way to view data at different scales of interest would also be useful. Finally, adding a geographic viewer is desirable, as localities can be mapped to Latitudes and Longitudes.

Further database development will involve support for other taxon and link attributes (flows, other measures of link strength). Addition of this information as well as additional link presence/absence data will depend on improved methods for discovering and integrating data. The SPIRE project is currently developing tools that use semantic web technology to automate the process of harvesting and integrating food web and natural history data (Parr et al., in press).

5   CONCLUSION

We have described the development of a large food web database and EcoLens, a tool that facilitates rapid exploration of it. Using this tool we obtained an overview of the integrated dataset, answered specific questions about the data, assessed the quality of the data and its integration, and showed an example of selecting a subset of data for further analysis. We found that many habitats are not well represented in our large database. We confirm earlier results about the small size and lack of taxonomic resolution in early food webs but find that they and a non-food-web source provide trophic information about a large number of taxa absent from more modern studies. Corroboration of Tuesday Lake trophic links across studies is usually possible, but lack of links among congeners may have several explanations. As the complexity of integrated datasets grows, so does the importance of an effective exploration tool. EcoLens is but one step towards fostering larger scale studies.

ACKNOWLEDGEMENTS

REFERENCES

Baird, D. and Ulanowicz, R.E., 1989. The seasonal dynamics of the Chesapeake Bay ecosystem. Ecological Monographs 59, 329-364.

Bederson, B.B., Grosjean, J., and Meyer, J., 2004. Toolkit design for interactive structured graphics. IEEE Transactions on Software Engineering 30, 535-546.

Christian, R.R. and Luczkovich, J.J., 1999. Organizing and understanding a winter's seagrass foodweb network through effective trophic levels. Ecological Modeling 117, 99-124.

Cohen, J., 1989. Ecologists' Co-Operative Web Bank. Version 1.00. Machine-readable data base of food webs. New York: The Rockefeller University.

Cohen, J.E., Beaver, R.A., Cousins, S.H., DeAngelis, D.L., Goldwasser, L., Heong, K.L., Holt, R.D., Kohn, A.J., Lawton, J.H., Martinez, N., O'Malley, R., Page, L.M., Patten, B.C., Pimm, S.L., Polis, G.A., Rejmanek, M., Schoener, T.W., Schoenly, K., Sprules, W.G., Teal, J.M., Ulanowicz, R.E., Warren, P.H., Wilbur, H.M., and Yodzis, P. 1993. Improving food webs. Ecology 74, 252-258.

Dunne, J., Williams, R., and Martinez, N., 2006. Webs on the Web. Available at: http://www.foodwebs.org.

Dunne, J.A., 2005. The network structure of food webs. In: M. Pascual and J.A. Dunne (Editors), Ecological Networks: Linking Structure to Dynamics in Food Webs. Oxford University Press, Oxford, 27-86.

Ehrig, M, Haase, P., and Stojanovic, N., 2004. Similarity for ontologies - a comprehensive framework. In Workshop Enterprise Modelling and Ontology: Ingredients for Interoperability, at PAKM 2004. Available at http://citeseer.ist.psu.edu/ehrig04similarity.html.

Fekete, J.-D., Grinstein, G., and Plaisant, C., 2004. IEEE InfoVis 2004 Contest: the history of InfoVis. Available at: http://www.cs.umd.edu/hcil/iv04contest.

Integrated Taxonomic Information System (ITIS), 2006. Available at: http://www.itis.usda.gov.

Jonsson, T., Cohen, J.E., and Carpenter, S.R., 2005. Food webs, body size, and species abundance in ecological community description. Advances in Ecological Research 36, 1-84.

Jordan, F., 2003. Comparability: the key to the applicability of food web research. Applied ecology and environmental research 1, 1-18.

Kang, H., Plaisant, C., Lee, B., and Bederson, B.B., in press. NetLens: Interactive exploration of Content-Actor data. Proceedings of the IEEE Symposium on Visual Analytics Science and Technology 2006.

Kemp, W.M., Smith, W.H.B., McKellar, H.N., Lehman, M.E., Homer, M.,Young, D.L., and Odum, H.T., 1977. Energy cost-benefit analysis applied to power plants near Crystal River, Florida. In: C.A.S. Hall and J.W. Day (Editors), Ecosystem Modeling in Theory and Practice: An Introduction with Case Histories. Wiley and Sons, New York.

Kim, B., Lee, B., and Seo, J., in press. Visualizing set concordance with permutation matrix and fan diagram, Posters/Videos Compendium of the IEEE Symposium on Information Visualization 2006.

Lee, B., Czerwinski, M., Robertson, G., and Bederson, B.B., 2005. Understanding research trends in conferences using PaperLens. Extended Abstracts of the ACM Conference on Human Factors in Computing Systems 2005, 1969-1972.

Lee, B., Parr, C.S., Campbell, D., and Bederson, B.B., 2004. How users interact with biodiversity information using TaxonTree. Proceedings of the 8th International Working Conference on Advanced Visual Interfaces 2004, 320-327.

Lee, B., Parr, C.S., Plaisant, C., Bederson, B.B., Veksler, V.D., Gray, W., and Kotfila, C., in press. TreePlus: Interactive exploration of networks with enhanced tree layouts. IEEE Transactions on Visualization and Computer Graphics.

Lima, M., 2006. Visual Complexity. Available at: http://www.visualcomplexity.com/vc/.

Luo, W., 2006. DockPanel Suite. Available at: http://sourceforge.net/projects/dockpanelsuite/.

Myers, P., Espinosa, R., Parr, C.S., Jones, T., Hammond, G.S., and Dewey, T.A., 2006. The Animal Diversity Web. Available at http://animaldiversity.org.

Odum, W.E. and Heald, E.J., 1975. The detritus-based food web of an estuarine mangrove community. In: Estuarine Research, Vol. 1, Chemistry, Biology and the Estuarine System. Academic Press, New York, 265-286.

Paine, R.T., 1988. Food webs, linkage interaction strength, and community infrastructure. Ecology 69, 1648-1654.

Parr, C.S., Espinosa, R., Dewey, T., Hammond, G., and Myers, P., 2005. Building a biodiversity content management system for science, education and outreach. Data Science Journal 4, 1-11.

Parr, C.S., Lee, B., Campbell, D., and Bederson, B.B., 2004. Tree visualizations for taxonomies and phylogenies. Bioinformatics 20, 2997-3004.

Parr, C.S., Parafiynyk, A., Sachs, J., Ding, L., Dornbush, S., and Finin, T., in press. Integrating ecoinformatics resources on the semantic web. Proceedings of the 15th International World Wide Web Conference .

Peterson, C.H., 1979. The importance of predation and competition in organizing the intertidal epifaunal communities of Barnegat Inlet, New Jersey, Oecologia 39, 1-24.

Pimm, S., 2002. Food Webs. University of Chicago Press, Chicago.

Polis, G., 1991. Complex trophic interactions in deserts: an empirical critique of food-web theory. American Naturalist 138, 123-155.

Storch, D. and Gaston, K.J., 2004. Untangling ecological complexity on different scales of space and time. Basic and Applied Ecology 5, 389-400.

Vazquez, D., 2005. NCEAS Interaction Web Database. Available at: http://www.nceas.ucsb.edu/interactionweb/.