Development of the Red Sea Biogeographic Information System

Fahmi Machda[1,*], Norú Ichim-Moreno[1,*], Chandra Prasetyo Utomo[2], and Christian R. Voolstra[1]
[1]Red Sea Research Center and [2]Mathematical and Computer Sciences and Engineering Division,
King Abdullah University of Science and Technology (KAUST), Thuwal, Kingdom of Saudi Arabia
*authors contributed equally

***Abstract*-Marine studies, surveys, and observational activities are continuously generating new and diverse data, which are hard to keep track of with tables and spreadsheets. Integrated data and information management systems that collect, analyze, and combine data are needed in order to provide a comprehensive picture of marine environments under study. For these reasons, we started to develop the Red Sea Biogeographic Information System (RBIS) at King Abdullah University of Science and Technology (KAUST) as a web application utilizing the most updated Web 2.0 technologies. RBIS is designed to have an easily accessible interface that is able to host and display research activities conducted in the Red Sea. Its data model is designed to deal with any kind of marine data. For its data structure, RBIS is organizing the data into three main categories: biological data, physicochemical data, and human activities. Spatial distribution of these data is visualized on a Google-Maps mashup. Dynamic charts are used to visualize the statistics of the data. With these functionalities, data model, and data structure, RBIS is able to organize, visualize, and do instantly combined analyses of research data from the Red Sea. The current version is accessible at http://www.kaust.edu.sa/rbis.**

I. Introduction

Research itself is complex as it integrates multiple levels of information and data. Particularly in marine biogeographic studies, it is common to employ biological, geographic, and oceanographic data that all are relevant to the outcome. Effective data processing is necessary to get the most out of these studies that assess spatial and temporal distributions of organisms and influencing factors [1].

Comprehensive biogeographic studies also need a common geographical interpretation of species observations, movement patterns, and behavior. Visual interpretation will ease the formulation of hypotheses that will aid in the understanding and utilization of marine data, which will ultimately help protect the marine environment as one of the arguable most important ecosystems. Therefore, gathering and integrating available data into an open access data center and information system will lead to a more comprehensive understanding of marine resources and environmental condition [2].

Traditionally, a data center has roles in long-term stewardship of data, integrating datasets, documenting, redistributing, and developing standards under a standard operational procedure. These traditional roles still remain important. But the trend is moving away from the traditional data depository function to a role of a data center that is more service-orientated. There is a need to create data and information software and implementations for wider audiences than just data managers and scientists. According to the data center model, databases should enable knowledge-based decision-making for policy makers and society in general. Databases should also facilitate and foster collaboration and peer review. For biological data, there is an increased need to access species databases completely and in an easy way. This data management should also incorporate a function that can easily measure biodiversity. Some well-known initiatives are the Ocean Biogeographic Information System (OBIS) and the Ocean Biogeographic Information System Spatial Ecological Analysis of Megavertebrate Populations (OBIS-SEAMAP). In general, for these systems the challenge is to logicistically employ and analyze biological data, physicochemical records, and information about human related activities and make these data available for integrative and instant analysis [3]. The involvement of an academic institution is also important because the creation of such a common data center is only feasible when a sense of community exists among marine researchers, data holders, administrators, and users of such a system [4].

The development of web 2.0 technologies and concepts has given possibilities to develop such new marine data management systems. A definition of Web 2.0 is the philosophy of mutually maximizing collective intelligence and added values for each participant by formalized and dynamic information sharing and creation. One of its concepts is treating data as a core competence [5]. Databases are valuable and growing naturally in value if they are constructed and used correctly. In turn, it can generate knowledge by linking the data in unprecedented ways. One important aspect that cannot be served by databases in the historical sense is the visualization of data. This is very important because humans are not particularly adapted to read huge tables of numbers. However, visualization becomes more and more important as data becomes more and more complex. In terms of usability and ease of data interpretation, visualized transformation of underlying data is a key to make data comprehensive to a wide community of users.

The Red Sea Biogeographic Information System (RBIS) is developed as a database to overcome, integrate, and make use of associated challenges, technologies and concepts to store, analyze and display marine research data. In particular, it is developed to support the Red Sea Research Center (RSRC) at KAUST. RSRC aims to develop a comprehensive understanding of coral reef ecosystems and their

978-1-4244-5222-4/10/$26.00 ©2010 IEEE

oceanographic context involving natural and anthropogenic factors. RBIS is presently in development and accessible online at http://www.kaust.edu.sa/rbis.

II. System Architecture

Freely available technologies were used in the development of RBIS as a web application (Fig. 1). On the server side, a Linux server is used with Apache HTTP Web Server. MySQL is used as Relational Database Management System (RDBMS). PHP is used as the server-side scripting language and to handle database connectivity. Symfony PHP framework is used with Doctrine as the Object Relational Mapper (ORM). On the client side, CSS is used to style the HTML web pages. To implement AJAX, the jQuery library is used to improve the interactivity of the web page. XML is used to transport and store data generated when user interacts with the web page [6]. Google Map API web service is used for map rendering and drawing the geolocations of the data [7]. A javascript library for clustering points on the Google Maps is also employed [8]. Charts are made by the implementation of Flot, a jQuery based javacsript plotting library [9].

III. Marine Data Modeling

A. Data Modeling

Data modeling is the process of reviewing and capturing data elements and their relationships to each other [10]. For marine biogeography, it can be used to discover basic data elements, determine how they relate to each other at the present time, and define them so that they can be recognized and interpreted in the future. Entity Relationship Diagram is the method in data modeling used for developing RBIS (Fig. 2).

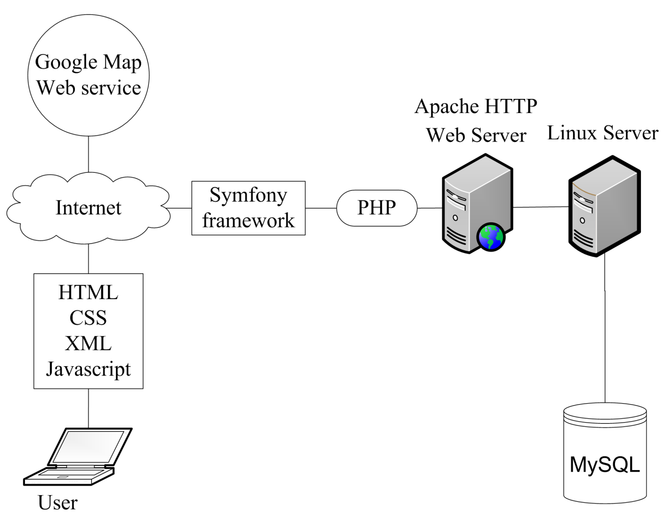

Fig. 1. RBIS Architecture displaying web application architecture utilizing Google Map API Web Services.

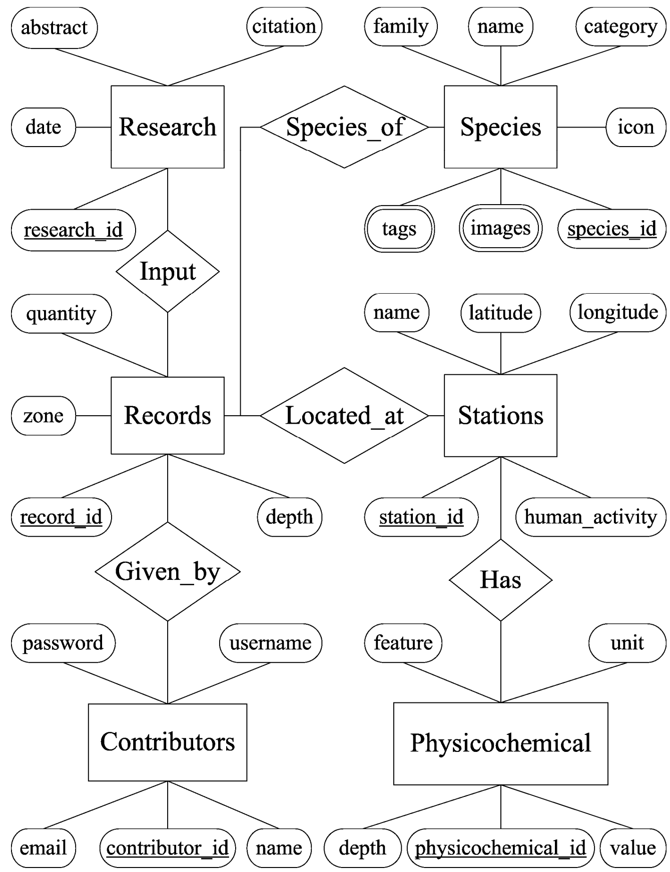

Fig. 2. Entity Relationship Diagram for RBIS displaying the interconnection of different database entities. Entities are displayed by squared corner rectangles, attributes are displayed by rounded corner rectangles, and relationships are displayed by diamonds.

There are six entities (squared corner rectangles) in the diagram:

1. The *Species* entity has seven attributes: species_id (auto-increment), name, family, category, icon, images, and tags. Species_id, name, category, and family will store the id of species, latin name, category name, and family name of the species. Icon will save the icon filename of the species, images will save all the images filename of the species, and tags will save all the species' labels. Images and tags are multi-value attributes so each species can have more than one image and label.
2. The *Stations* entity has five attributes: station_id (auto-increment), name, latitude, longitude, and human activity. The attributes station_id and name will store the id and name of the station. Latitude and longitude will save the coordinates of the station and human_activity will save the information about activities on the station.
3. The *Physicochemical* entity has five attributes: physicochemical_id (auto-increment), feature, depth, unit, and value. Physicochemical_id, feature, and depth will save respectively the id, the kind of physical and chemical features, and the depth where the

measurements were taken. Examples of features are dissolved oxygen, temperature, and salinity. Unit and value attributes will save the information of unit measurements and obtained values.

4. The *Research* entity contains the properties of research. It has four attributes: research_id (auto-increment), date, abstract, and citation. Each *Research* entry has many *Records* entries associated with it.
5. The *Contributors* entity has five attributes: contributor_id (auto-increment), name, email, username, and password.
6. The *Records* entity integrates *Species*, location, *Research*, and *Contributor*. It has four attributes: record_id (auto-increment), quantity, zone, and depth.

Between the entities, there are five relationships (diamonds) in the diagram:

1. *Input* is a relationship between *Research* entity and *Records* entity. This relationship connects each research with all data produced by it and consequently connects each dataset with related research, which produced that dataset.
2. *Given_by* is a relationship between *Records* entity and *Contributors* entity. This relationship connects each record with the contributor who provided the record data and consequently connects contributor to all record data that they contribute.
3. *Species_of* is a relationship between *Species* entity and *Records* entity. This relationship connects each record with the species on that record and consequently connects each species with all recorded data of it.
4. *Located_at* is a relationship between *Records* entity and *Stations* entity. This relationship connects each record with the location of that record and consequently connects each station with all associated recorded data.
5. *Has* is a relationship between *Station* entity and *Physicochemical* entity. This relationship connects each station with all physicochemical properties that it is associated with and consequently connects all the physicochemical data with the related station.

B. Data Structure

RBIS is organizing marine biogeography data in a hierarchical fashion consisting of three main groups: biological data, physicochemical data, and human activities. This organization is designed to allow combined analysis integration, for example, human-induced changes affecting organism abundance in the Red Sea. Biological data are divided in seven subgroups, which are microbes, primary producers (containing seaweeds and plants), invertebrates, vertebrates (fish), marine mammals, marine reptiles, and marine birds (Fig. 3). This data structure is designed to handle the complexity of biological data in a simpler way than taxonomic classification for faster data exploration in the system.

C. Sample Data Set

KAUST collected data from a coral reef vertical zonation near the proposed desalination plant outfall. The data was collected by transect-belt method in 2008. The transect-belt method is used in biological studies to investigate the distribution of organisms in relation to a certain area, such as the seashore [11]. The survey was conducted at six stations near KAUST. Collected data from that survey [12] was used as a sample data set for integration into RBIS.

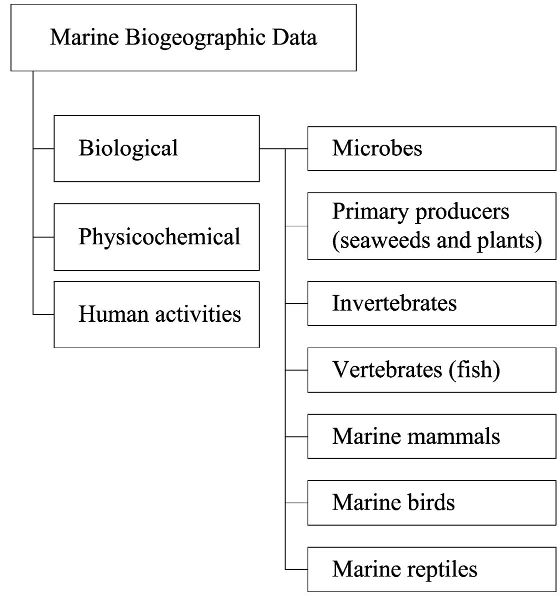

Fig. 3. Hierarchical structure of RBIS data organization

IV. Application and Discussion

A. Interface Design and Visualization

The RBIS interface consists of three sections. The first section is a data bar containing a data report to the query, the second section is a right-side column for data browsing, and the third section is the Google Maps mashup module (Fig. 4). The report module's main function is to show the quantity of species in each specific observational location as bar charts. The right-side column section contains the tree structure of the data that exists in the system. This tree structure allows the user to browse the existent data and select the parts of it that are of interest to the user. The selected elements will be used as the query that will trigger the functionality of the other modules. The function of the Google Maps module is to show the distribution of data locations. On the map, points are dynamically generated from the data that the user selects. The points generated in the map can be clustered according to the zoom level.

Through this interface design, RBIS is able to visualize the original survey data and allow users to analyze the spatial coverage of the observations. Multiple data on the right side tree can be selected. The comparison between species quantity in the different location is shown as generated charts and spatially on the map. RBIS is also able to combine data related to different species to show differences of biodiversity in each location. These visualization features are a distinct feature of RBIS in comparison to other biogeographic data services such as OBIS and OBIS-SEAMAP.

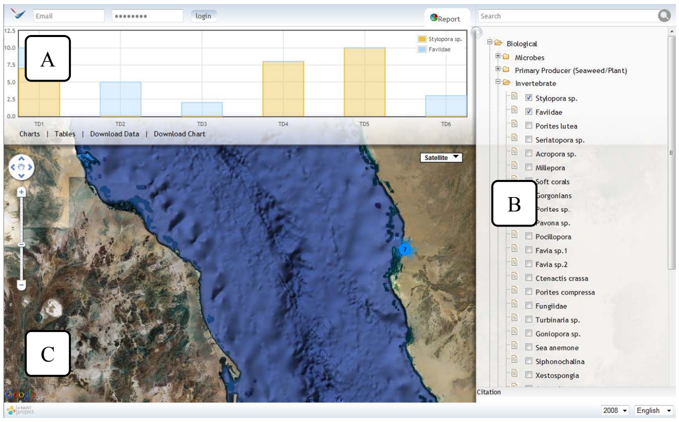

Fig. 4. RBIS user interface showing: A) the first section that contains data report, B) the second section displaying the tree structure for data browsing, and C) the third section that contains Google Maps mashup module.

B. Gap Analysis

Gap analysis is a type of analysis that is useful to compare the distributions of organisms with the distribution of areas of interest [13]. The implementation of the interface design and visualization function in RBIS allows the user to easily conduct Gap Analyses. One apparent use case is the mapping of Marine Protected Areas (MPAs) to the abundance and distribution of species of interest. Please note that in comparison to OBIS and OBIS-SEAMAP, RBIS is able to show the distribution of multiple species, and hence ecosystem diversity.

The distribution of our sample data set shows that each location has a different biodiversity. The chart bar dynamically displays the different quantities of data in regard to the different sample sites (geolocations) (Fig. 5). Gap analyses provide an intuitive and fast way of assessing data distribution.

In regard to KAUST, gap analyses can contribute to interpretation of data such as where the most important and biologically diverse areas to be protected are located. In regard to our sample data set, RBIS can help assess where the outfall of the purposed desalination plant should be located and where the next research should be conducted to investigate the impact of the outfall.

C. Data Referencing

Since RBIS is designed to be a data aggregator for a region that involves many institutions, data referencing is necessary to attribute the original data owners and give the correct citation of displayed data. To avoid difficulties in keeping track of the original source of data when many data points are selected, RBIS has one citation box that will collect citations of all displayed data (Fig. 6). The citation box facilitates and enables users to go back to the original reference with ease.

D. Point Clustering

RBIS can display different information according to the selected zoom level. When a RBIS user zooms out of the map, RBIS will cluster overlapping points that show the total numbers of different observations in terms of data types and location. When the user zooms into the map, the points will be

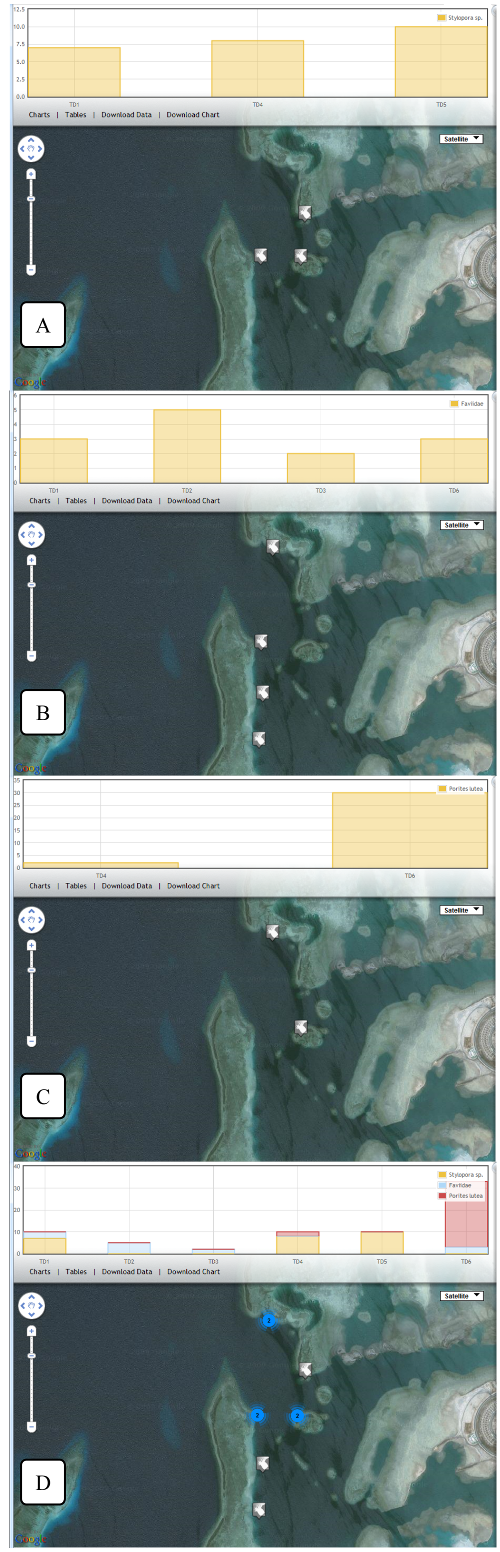

Fig. 5. Gap analysis implementation in RBIS: shown are the spatial distribution and abundance of the coral species A) *Stylopora sp.,* B) *Favidae sp.,* C) *Porites lutea* and D) the combination of *Stylopora sp, Favidae sp,* and *Porites lutea.*

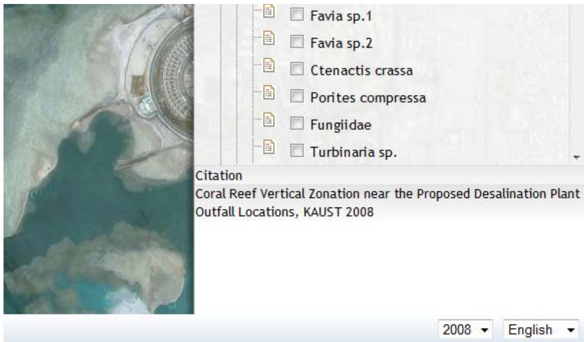

Fig. 6. The data referencing feature in RBIS allows researchers to go back to the original data source in the same interface as where the data is selected.

un-clustered showing more detailed observation points. This function makes it easy to spot yet-to-be-explored regions in the Red Sea, as well as to have an instantaneous visual and statistical information summary (Fig. 7).

E. Future Work

The first phase of the project has been mainly focused on visualization and data organization. In the following steps, we plan to integrate more data of Red Sea related research. Data from recent researches and cruise expeditions conducted by KAUST will be collected on spreadsheet templates and uploaded to RBIS. Our long term anticipated plan is to broaden the inclusion of users and data to the maximum extent possible, engaging with government agencies such as Presidency of Meteorology and Environment (PME) of the Kingdom of Saudi Arabia, Regional Organization for the Conservation of the Environment of the Red Sea and Gulf of Aden (PERSGA), and other related institutions.

V. Conclusion

RBIS stores, combines, and visualizes marine biogeography data of the Red Sea. The first step of the development has been focused on the visualization, data organization, and data integration for biogeographic studies incorporating state-of-the-art Web 2.0 technologies and concepts. At this point, RBIS is able to relate different data instantly and visualize its distribution on Google Maps and by dynamic bar charts. This feature is highly useful in Gap Analyses. RBIS is also able to compute point clustering according to the zoom level. This implication is useful to look at data in a hierarchical and organized manner and to easily identify underexplored regions. Data referencing is in the form of a citation box in the same interface selected data is displayed, so that the user has instantaneous overview of data sources. The next steps are the development of functions that increase and facilitate the incorporation and upload of data sets by users.

Acknowledgement

We thank Professor James Luyten and the RSRC at KAUST for supporting this research as a long-term initiative to store, organize and retrieve marine research data. We also thank Abdulaziz Al-Suwailem and the Coastal and Marine Resources Core Lab (CMRCL) at KAUST. We thank Rebecca Allen and Nathan Ball for providing helpful comments.

Fig. 7. Point clustering when A) the map is zoomed out and B) the map is zoomed in

References

[1] M.E. Monaco, M.S. Kendall, J.L. Higgins, C.E. Alexander, and M.S. Tartt, "Biogeographic assessments of NOAA National Marine Sanctuaries: the integration of ecology and GIS technology," *IEEE OCEANS Conference Proceedings,* vol. 3, pp.1489 – 1498, 2003. [Online]. Available: IEEE Xplore, http://ieeexplore.ieee.org/. [Accessed: Mar. 21, 2010].

[2] O. Sergeyeva, D. Slipetskyy, V. Gorbunov, and V. Vladymyrov, "Integrated data and information management system to collect and analyze biological data for climate change research," *2nd Biannual Black Sea SCENE EC project joint conference,* 2008. [Online]. Available: http://repository.ibss.org.ua/dspace/handle/99011/696. [Accessed: Mar. 19, 2010].

[3] Jan Mees and and Edward Vanden Berghe, "Marine Data Management: we can do more, but can we do better?," *iode.org*, Jan. 23, 2007. [Online]. Available: http://www.iode.org/index.php?option=com_content&task=view&id=3&Itemid=33. [Accessed: Mar. 19, 2010].

[4] Patrick N. Halpin, et al., "OBIS SEAMAP, The world data center for Marine Mammal, Sea Bird, and Sea Turtle distributions," in *Oceanography*, vol. 22, 2009, pp. 105-115. [Online]. Available: http://www.tos.org/oceanography/issues/issue_archive/issue_pdfs/22_2/22.2_halpin.pdf. [Accessed: Mar. 19, 2010].

[5] Senoaji Wijaya, Marco R. Spruit, Wim J. Scheper,"Webstrategy Formulation: Benefiting from web 2.0 concepts to deliver business

values," in *Web 2.0, The Business Model*. Miltiadis D. Lytras, Ernesto Damiani, and Patricia Ordóñez de Pablos, Eds. Spinger, 2009, pp 103-132. [E-book]. Available: SpringerLink.

[6] Fabien Potencier, *Practical symphony*. France: Sensio SA, 2009. [E-book] Available: http://www.symfony-project.org/get/pdf/jobeet-1.4-doctrine-en.pdf. [Accessed: Mar. 21, 2010].

[7] Google, "Google Maps API," [Online]. Available: http://code.google.com/apis/maps/documentation/. [Accessed: Mar. 20, 2010].

[8] Xiaoxi Wu, "MarkerClusterer," 2009. [Online]. Available: http://code.google.com/p/gmaps-utility-library-dev/. [Accessed: Mar. 21, 2010.

[9] Ole Laursen, "flot, Attractive Javascript plotting for jQuery," [Online]. Available: http://code.google.com/p/flot/. [Accessed: Mar. 20, 2010].

[10] Sharon Allen and Evan Terry, *Beginning Relational Data Modeling, Second Edition*. Berkeley, CA: Apress, 2005. [E-book]. Available: SpringerLink.

[11] David Arnold Hill,Matthew Fasham, Graham Tucker, Michael Shewry, and Philip Shaw, *Handbook of biodiversity methods: survey, evaluation and monitoring*. Cambridge, UK: Cambridge University Press. pp. 219–222 2005.

[12] KAUST, "Coral Reef Vertical Zonation near the Proposed Desalination Plant Outfall Locations," unpublished.

[13] Michael D. Jennings, "Gap analysis: concepts, methods, and recent results," in *Landscape Ecology,* vol. 15, Netherlands: Kluwer Academic Publishers, 2000, pp. 5-20. [Online]. Available: SpringerLink, http://www.springerlink.com. [Accessed: Mar. 21, 2010].