Indian Journal of Marine Sciences
Vol. 34(1), March 2005, pp. 120-127

IndOBIS, an Ocean Biogeographic Information System for assessment and conservation of Indian Ocean biodiversity

*Vishwas Chavan[1], C. T. Achuthankutty[2], Edward Vanden Berghe[3] & Mohideen Wafar[2]

[1]Information Division, National Chemical Laboratory, Dr. Homi Bhabha Road, Pune 411008, India
[2]Biological Oceanography Division, National Institute of Oceanography, Dona Paula, Goa 403004, India
[3]Flanders Marine Data and Information Centre, Flanders Marine Institute, Vismijn, pakhuizen 45-52, 8400 Oostende, Belgium
*[E-mail: vishwas@ems.ncl.res.in]

Received 1 April 2004; revised 10 January 2005

Compilation of inventories of components of coastal and marine biodiversity of Indian Ocean is hampered by several factors: low effort by some countries, preference to certain taxon, dwindling taxonomic expertise, low infrastructure of Information Technology, databases that are scattered and often non-interoperable, inconsistent reporting and a marked reluctance to share data and information. The creation of the Indian Ocean node of the Ocean Biogeographic Information System is meant to overcome some of these constraints. Benefiting from progress in Information Technology and building on the global efforts on understanding what lives in our seas, the IndOBIS aims to grow into a self-sustaining and collectively supported process of information collation, analysis and dissemination, serving the countries of the region and the international scientific community.

[Key words: Information systems, distributed database systems, biogeography, biodiversity informatics, Indian Ocean, Ocean Biogeographic Information System, IndOBIS.]

[IPC Code: Int. Cl^7 G11C 13/00]

Introduction

The Indian Ocean is the third largest ocean in the world, occupying 73.44 million km^2, equivalent to 21% of the world`s sea area. It is a significant contributor to the production of living marine resources with an estimated annual yield of 8 million tons of capture fisheries and 23 million tons of culture fisheries, equivalent respectively to 10 and 90% of world production. The tropical nature also endows the Indian Ocean countries with vast stretches of sensitive coastal and marine ecosystems and a high diversity of fauna and flora. For example, the coral reefs of the Indian Ocean account for 30% of the global reef cover and the mangroves, for 10%, with the world`s second largest mangrove forest lying in the Gangetic delta of the Indian sub-continent.

The Indian Ocean region includes 36 littoral and 11 hinterland states, accounting for 30% of the earth`s population. With most of them remaining economically under-developed or still developing, it is understandable that the coastal and marine biodiversity elements of this ocean region remain still relatively less known and that they are also heavily impacted, both from unsustainable patterns of resource exploitation and unregulated discharge of effluents into coastal waters.

What we know of the components of coastal and marine biodiversity of the Indian Ocean region or changes in them over time are highly scattered and often are not readily accessible to all. Although much research has been carried out, or is currently ongoing, many initiatives are unrelated, data are difficult to trace and even more difficult to integrate. What is available is often in non-interoperable formats, compounded by uncertain taxonomic identities or absence of a host of secondary information. This has severely constrained our ability to understand the true species diversity of our seas, let alone be in a position to predict what could live there eventually. The Ocean Biogeographic Information System (OBIS) was conceived to address these issues and the Indian Ocean node of this (IndOBIS) has specifically been designed to improve coordination between domain experts and information technology managers and strengthen communication and information exchange between the countries of the region, which so far has been characterized by a marked reluctance.

Census of Marine Life (CoML) and Ocean Biogeographic Information System (OBIS)

The Census of Marine Life (CoML, http://www.coml.org/coml.htm)[1] is a growing global network of researchers in more than 70 nations engaged in a ten-year initiative to assess and explain the diversity, distribution, and abundance of life in the ocean and changes in it over time. The Ocean Biogeographic Information System (OBIS, http://www.iobis.org)[2] is the information component of the Census of Marine Life, and the marine component of the Global Biodiversity Information Facility (GBIF, http://www.gbif.org)[3]. It is a growing network of researchers in more than 45 nations engaged in a 10-year initiative to assess and explain the diversity, distribution, and abundance of life in the oceans - past, present, and future.

OBIS is a web-based provider of global geo-referenced information on accurately identified marine organisms. OBIS contains expert species level and habitat level databases and provides a variety of spatial query tools for visualizing relationships among species and their environment. It also strives to assess and integrate biological, physical, and chemical oceanographic data from multiple sources.

The OBIS Portal provides data content, information infrastructure, and informatics tools - maps, visualizations and models – to provide a dynamic, global facility in four dimensions (the three dimensions of space plus time). Potential uses are revelations of new spatial/temporal patterns of species distributions, generation of new hypotheses about the global marine ecosystem and guidance of future expeditions. The scope of OBIS offers new challenges in data management, scientific cooperation and organization, and innovative approaches to data analysis. Maintaining the principle of open access, the digital atlas developed by OBIS is expected to provide a fundamental basis for societal and governmental decisions on how to harvest and/or conserve marine resources. A more recent description of the system is reported elsewhere[4-7].

OBIS history

The initial idea for OBIS was developed from a CoML-sponsored Benthic Census Meeting held in October 1997. Recommendations from this meeting led to the establishment of a prototype OBIS website at Rutgers University (USA) in 1998 to demonstrate the initial OBIS concept. In June 2001, OBIS joined Global Biodiversity Information Facility (GBIF, http://www.gbif.org/)[3] as an Associate Participant. GBIF is a data system for worldwide biological data, OBIS will grow in concert with GBIF to become the major component for ocean biogeography and systematics. The major contributors to OBIS, as of December 2004 are:

- BATS Zooplankton
- Biogeoinformatics of Hexacorals
- Biotic Database of Indo-Pacific Marine Mollusks
- CSIRO Marine Research
- CephBase
- FAO Catch Statistics and Aquaculture Production
- FishBase
- Fishnet
- Gulf of Maine Biogeographic Information System
- History of Marine Animal Populations
- The Huntsman Marine Science Centre
- Integrated Taxonomic Information System
- Maritimes Region Homepage (Canada)
- National Ocean Data Center (NODC)
- OBIS-SEAMAP
- SeamountsOnline
- Species 2000
- ZooGene

A broad range of other affiliations has been formed which includes:

- Intergovernmental Oceanographic Commission (IOC)
- Scientific Committee on Ocean Research (SCOR)
- DIVERSITAS
- International Council for the Exploration of the Seas (ICES)
- International Association of Biological Oceanographers (IABO)
- International Union of Biological Sciences (IUBS)
- United Nations Environmental Program (UNEP)
- World Conservation Monitoring Centre (WCMC)
- UNESCO Man-And-the-Biosphere (MAB)
- Global Ocean Observing System (GOOS)
- Global Biodiversity Information Facility (GBIF)

OBIS structure

OBIS is structured as a federation of organizations and people sharing a vision to make marine biogeographic data, from all over the world, freely available over the World Wide Web through the OBIS Portal. OBIS partners agree to develop and promote standards and interoperability according to the standards and protocols being developed for other environmental data systems around the world. It is not a project or programme, and is not limited to data only from CoML-related projects. OBIS is not incorporated, does not employ staff, own equipment, or apply for funding. Organizations involved in OBIS take on these responsibilities. Any organization, consortium, project or individual may contribute to OBIS.

OBIS is a distributed system, making full use of recent developments in technology, such as XML. Data from individual providers are not uploaded in a central database; instead, dynamic links are created with individual datasets, and data are consulted directly from the original location of the dataset. At this original location, an 'OBIS provider' has to be installed. An 'OBIS portal' connects to one or several of these providers, and sends visitors to the portal site information from these providers, combined in a single web page.

For communication between the OBIS portal and the OBIS provider, the 'Distributed Generic Information Retrieval (DiGIR)' protocol is used. DiGIR is an open source software, developed in PHP, and is designed to use open protocols and standards (HTTP, XML and UDDI – Universal Description, Discovery and Integration of web services); both the request for data and the results are passed between portal and provider as XML structures[8]. One of the advantages of DiGIR is really generic: the semantics are completely decoupled from the protocol and software used. The actual content of the data being exchanged is specified in an XML schema: a list of data fields with names, descriptions, and format notes. Both OBIS and GBIF have developed schemas. The one used by GBIF is known as the Darwin Core, now on version 2. The list of fields specified in the OBIS schema is a superset of the fields used by GBIF; this implies that all OBIS providers can also act as GBIF providers, but not the other way round. Part of the process of installing DiGIR will be to map the fields in one's database to the schema of one's choice.

The complete schema for OBIS has over 70 fields. However, only seven fields are required to constitute a basic biogeographical record (Table 1). An optional field is a URL for the record, allowing a link directly back to the relevant record in the provider's database.

A typical request from a user would result in the portal interrogating several providers (Fig. 1). The end user initiates the process by interrogating the portal

Table 1—List of required fields in the OBIS schema, plus Record URL

Name	Required	Type	Description
Date Modified	Last Y	Date, Time	ISO 8601 compliant stamp indicating the date and time in UTC(GMT) when the record was last modified. Example: "November 5, 1994, 8:15:30 am, US Eastern Standard Time" would be represented as "1994-11-05T13:15:30Z" (see W3C Note on Date and Time Formats - http://www.w3.org/TR/NOTE-datetime).
Institution Code	Y	Text	A "standard" code identifier that identifies the institution to which the collection belongs. Can be a full name or an abbreviaton if it is commonly understood in your discipline.
Collection Code	Y	Text	A unique alphanumeric value that identifies the collection within the institution. Can be a full name or an abbreviaton if it is commonly understood in your discipline.
Catalog Number	Y	Text/ Numeric	A unique alphanumeric value that identifies an individual record within the collection. It is recommended that this value provides a key by which the actual specimen can be identified. If the specimen has several items such as various types of preparation, this value should identify the individual component of the specimen
Scientific Name	Y	Text	The full name of lowest level taxon the Cataloged Item can be identified as a member of; includes genus, specific epithet, and subspecific epithet (zool.) if applicable or infraspecific rank abbreviation, and infraspecific epithet (bot.) Use name of suprageneric taxon (e.g., family name) if Cataloged Item cannot be identified to genus, species, or infraspecific taxon.
Longitude	Y	Numeric	The longitude of the location from which the specimen was collected or in which the sample/observation/record event occurred. This value should be expressed in decimal degrees (East & North = +; West & South = -). GPS-derived data should be referenced to the WGS/84 datum.
Latitude	Y	Numeric	The latitude of the location from which the specimen was collected. This value should be expressed in decimal degrees (East & North = +; West & South = -). GPS-derived data must use the WGS 84 geodetic reference system (http://www.wgs84.com/).
Record URL	N	Text	Gives the web address of the page where more information on this particular record (not on the whole dataset) can be found.

site. The portal checks the list of providers it knows about, and sends an xml-formatted request to each of these. Each of the providers receives the request, and uses it to check for relevant data in the database. In this process, the OBIS schema is used to match database fields with OBIS fields. Possible hits are packed in a XML file and returned to the portal. The portal collects all these responses, combines them in a single HTML page, which is returned to the end user's browser.

Each database can feed several providers (Fig. 2). When it is useful or desirable to split the available data in separate sections, this can be queried separately through separate providers. Likewise, several portals can exist, each with its own specificity. It is perfectly possible, for example, to build a series of regional portals, or to build portals specific for a taxonomic group. Once a provider has been installed, the underlying data become available for any portal site that might want to make use of it.

In order to contribute data to the OBIS community, a DiGIR provider should be installed on a web server and the fields of the local database mapped to the names of the OBIS Schema. The first step clearly requires a HTTPD server (such as IIS or Apache), and also that the PHP system in which DiGIR is developed, is running. Apache, PHP and DiGIR are all open source software and available freely. Only some expertise, obtainable from OBIS development team, would be needed to make all these components functional. Alternatively, data can be made available to the OBIS development team (or in principle to any other organisation that has an OBIS provider running) and brought online this way.

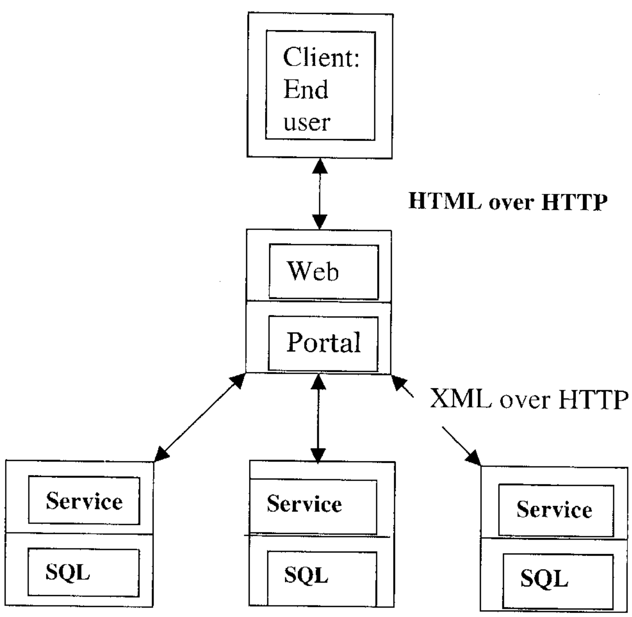

Fig. 1—Several DiGIR Providers make information from their database available to a single portal, which can be consulted by the user.

Globalizing OBIS: Establishment of Regional OBIS Nodes (RON)

The scope of OBIS offers new challenges in data management, innovative approaches to data analysis and scientific cooperation and management. When OBIS was conceived, many biologists were hesitant to make their data publicly available through a central portal. Many of the international authorities on the taxonomy of marine organisms (including those who maintain the Global Biodiversity Information Facility's Catalogue of Life) are now part of the OBIS Federation. As a consequence, OBIS has gained credibility with the scientific community and provides a mechanism for quality assurance of species data. As on December 2004, OBIS is serving over 5.6 million records for over 50,000 known species from 38 data sources.

OBIS is seen as the means for development of international marine biological standards and protocols through the Global Biodiversity Information Facility (GBIF), Intergovernmental Oceanographic Commission's IODE and GOOS. This could usher in a mindset where scientists will be able to trust biological data that they have not collected themselves. As was the case with data on physical oceanography, community data sets will become the basis for scientific breakthrough in areas such as spatial ecology and relationships between biodiversity and ecosystem function.

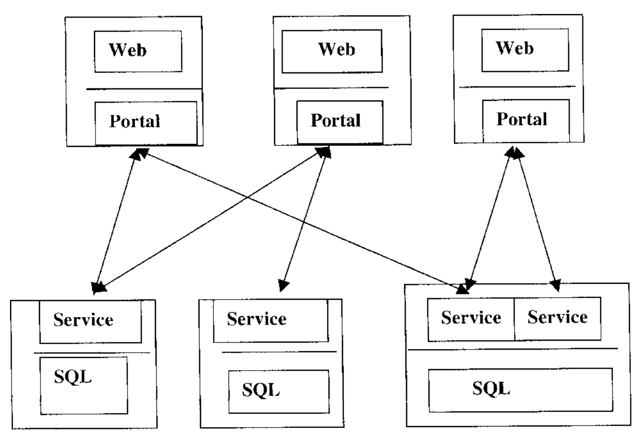

Fig. 2—One data server can support several services; one provider can be accessed by several portals

The major global data management initiatives concerning biodiversity data have established nodes that gather and supply data all around the world. These are the Global Biodiversity Information Facility "national nodes" and the Intergovernmental Oceanographic Commission's "National Ocean Data Centres". OBIS has recognized that a network of Regional OBIS Nodes (RONs) will have several strategic and practical advantages in expanding and sustaining OBIS. The most important function of RONs is to foster the on-line provision of marine biogeographic data from individual regions and areas of expertise. For example, a European RON should actively arrange for data held within European databases to be made available on-line through a European OBIS portal. Each RON will continue to populate OBIS with data and aid development of on-line data analysis and presentation tools, and will be an integral part of OBIS (Fig. 3). RONs would support the development of other RONs but will be independently motivated by the vision to 'publish' quality data on-line so as to increase the availability of data to researchers and educators around the world in the interest of good science.

In mid 2004, the Alfred P. Sloan Foundation has given a $1.5 million grant to establish a global network of organizations to support development of OBIS (www.iobis.org) with regional nodes in Australia, Canada, China, Europe, India, Japan, New Zealand, South America, and Sub-Saharan Africa.

IndOBIS: Indian Ocean Node of OBIS

During the first workshop on coastal and marine biodiversity [held at Goa, India in December 2003], the participants from Indian Ocean rim countries agreed that an Indian Ocean gateway to OBIS could be advantageous in enhancing the quality of the regional databases, promoting regional co-operation and benefiting from the scientific and technical expertise within the region so that a sense of ownership could be fostered. This was discussed in detail by the working group on data and information management during the workshop and has been endorsed as a strong recommendation.

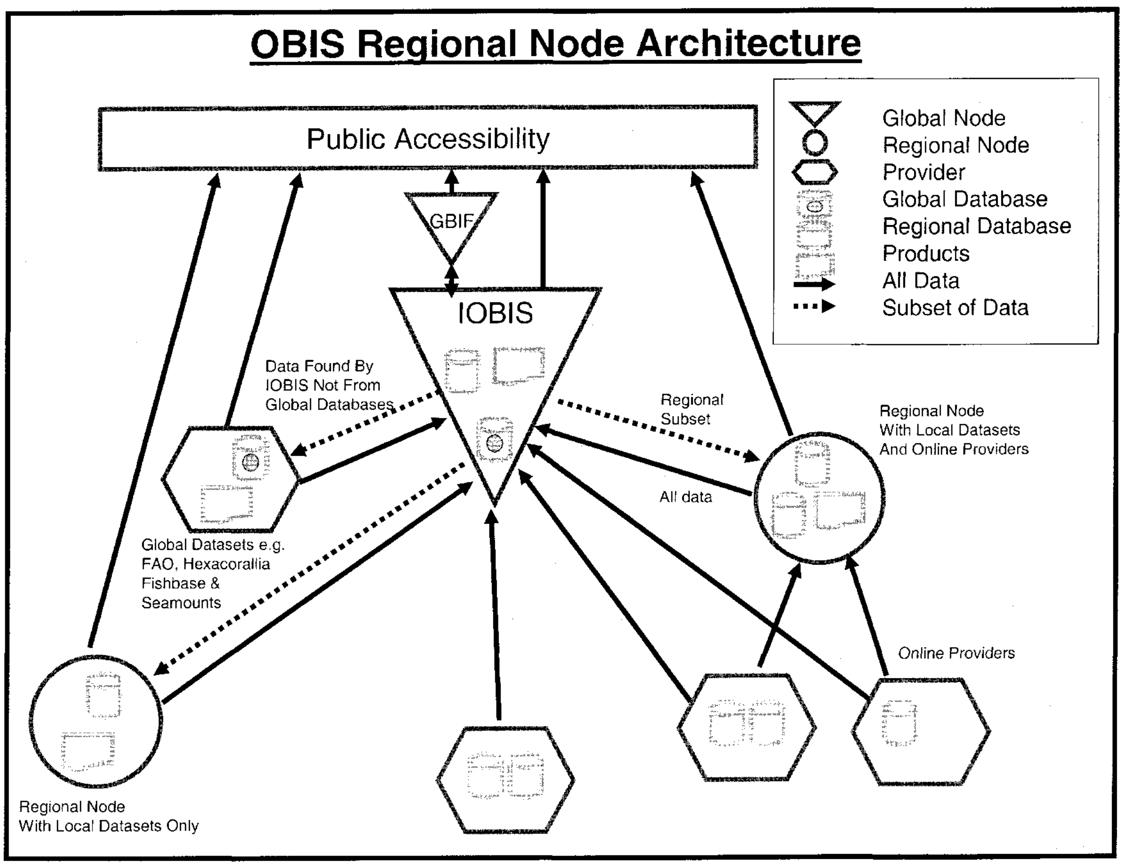

Fig. 3—Schematics of OBIS operations, including data exchange between the portal and regional nodes, and interactions with data providers, global databases, and end-users (public).

IndOBIS will contribute to the understanding of the past and the present, in order to learn about the future of life in the Indian Ocean. IndOBIS will become a prime provider of biodiversity information on the Indian Ocean, and make this information available in a multidimensional geographic context; promote communication and awareness to user groups at all levels, using the appropriate information tools; and enable informed decision making process leading to sustainable use of natural resources.

IndOBIS: Deliverables

When mature, IndOBIS would be a self-sustaining process of information collation, analysis and

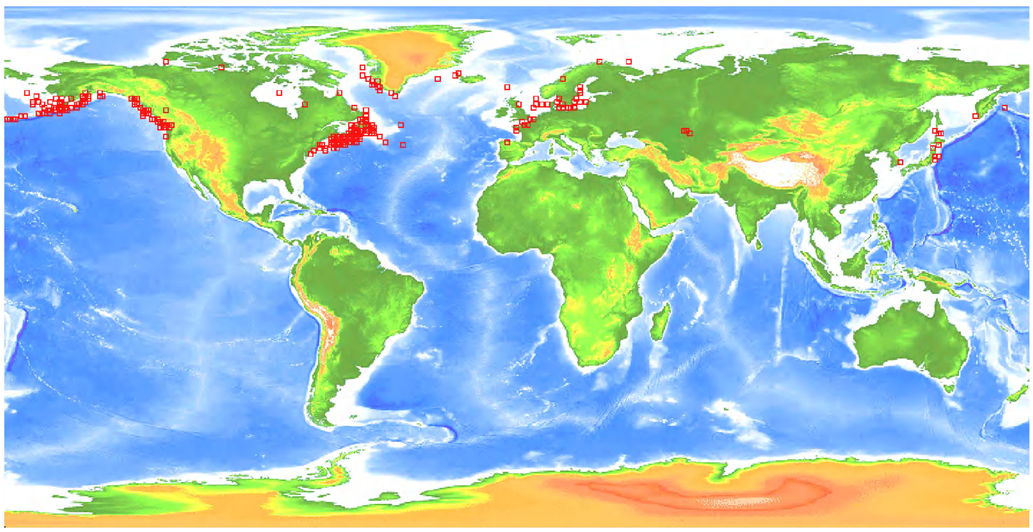

Fig. 4—Distribution of Atlantic cod, as found through the OBIS Portal, and represented by the c-squares mapper.

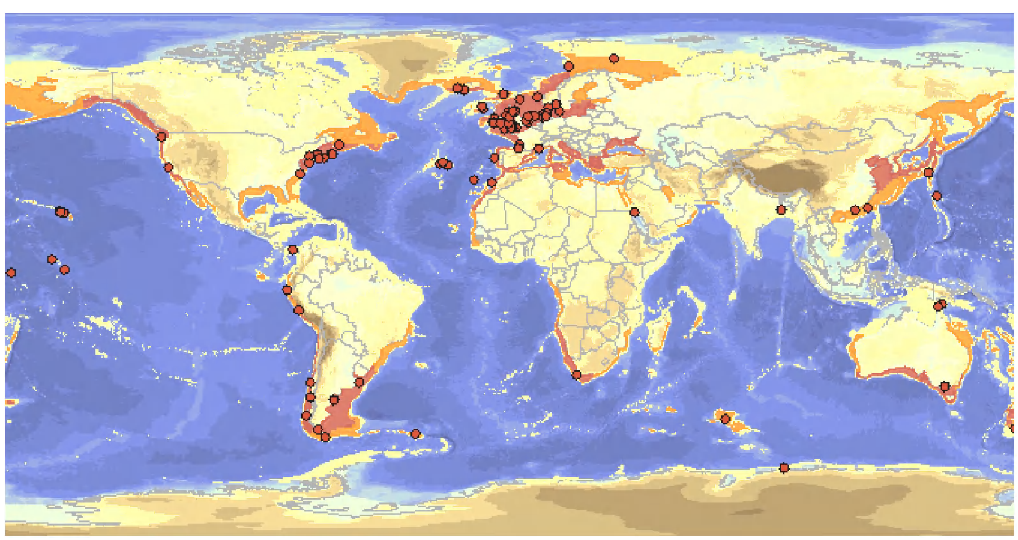

Fig. 5—Distribution of *Sagartia* (red dots), and ocean areas with similar physical characteristics (red and orange zones)

dissemination. The business plan of IndOBIS promises to deliver several tangible products by leveraging on the networking of potential custodians of the data. Some of the proposed deliverables are discussed below:

Electronic catalogue of known biota of Indian Ocean—IndOBIS would collect, integrate and collate information from sources that lie scattered among several agencies, institutions and individuals within the region in forms and formats ranging from ancient literature to modern day electronic forms such as flat files and relational databases. This would require taxonomic treatments, validation and authentication by domain experts.

Digital literature bank—Aquatic Sciences and Fisheries Abstracts (ASFA) captures abstract data on literature published since 1970. The pre-1970 data for the Indian Ocean, available in non-digital form, would be captured for abstracting or bibliographic details on them and digitised.

Digitizing collections—A large fraction of the marine biological collections from the region are housed in the museums / collections outside the region, notably in Europe and USA. IndOBIS will develop a framework to access data on these collections. In respect of museum collections within the region, IndOBIS would identify such museums and provide them with technical and data management guidance and work with them to secure funding for digitization of these collections.

Databank on taxonomic expertise—Data on taxonomic expertise of both institutions and individuals is scattered and not up-to-date. IndOBIS would develop a framework so that institutions and individuals themselves could update information on their own.

Database on datasets—Several databases on biodiversity of Indian Ocean have been developed in isolation and hence are inaccessible. IndOBIS would locate them and convert them to web accessible databases of data sets.

Other deliverables—IndOBIS would begin initiatives in developing products, tools, and protocols and help evolve guidelines to encourage information and communication technology implementation in marine biological data management, and also to develop capacity building in marine biodiversity informatics. These include:

1. Computer assisted taxonomy and electronic field guides
2. Data collection tools (for rescue, archival and capture)
3. Multilingual data dissemination and acquisition
4. Capacity building in marine biodiversity informatics
5. Data analysis, interpretations, visualization tools
6. Data management/exchange guidelines, policies and protocols

IndOBIS: Long-term sustenance plan

IndOBIS would be required to establish closer ties with national, regional and international agencies to ensure sustained support to its activities. Encouragement to national delegates to explore funding opportunities and provision of strategic and political support to them to raise funds is a must for this. Hence, IndOBIS would need to ensure visibility through products of national / regional significance which could later be part of OBIS and integrate with other data management initiatives such as IO-GOOS, CODATA, IODE, etc.

Discussion

Several web applications can be built making use of the information resources provided by the OBIS system. The most obvious result returned from a query to a portal is a list of species occurring in an area. Alternatively, a map can be created with all specimen records for a given taxon, such as the distribution of the Atlantic cod, *Gadus morhua* (Fig. 4). The main OBIS Portal also offers choices of mapping softwares, one of which is the C-square mapper developed at CSIRO[9]. A combination of physical oceanographic parameters with known distribution of species could be used in this software to predict where else a species might be expected to occur. using the WhyWhere algorithm[10]. Figure 5 shows both the observed distribution (as specimens known to OBIS) and the possible range of the genus *Sagartia* as calculated by the 'Dynamic Mapper' of the University of Kansas[11].

One of the great strengths of the system is its distributed nature. Data are not uploaded to a central location, but remain in the institute where they have been collected. This means that the originator /custodian of the data remains in control of the data. For many scientists, this is a strong argument in favour of making data available to the system. Another consequence of the distributed nature of

OBIS is that it becomes possible to have regional nodes, serving as gateways to the international portal. The regional nodes are more efficient in mobilising marine biodiversity information from their area than a central, international portal, by virtue of the fact that they are closer to the scientists from the region, know better what data are available, and often know the scientists who collected data in the region personally. Organising a regional node, such as IndOBIS, with also a regional portal to the biodiversity data, therefore, increases the visibility of the OBIS programme in the region, and will serve to further attract biodiversity data to the system.

References

1 Census of Marine Life (CoML)., available at http://www.coml.org/.

2 Ocean Biogeographic Information System (OBIS), available at http://www.iobis.org/.

3 Global Biodiversity Information Facility (GBIF), available at http://www.gbif.org/.

4 Grassle, J.F., The Ocean Biogeographic Information System (OBIS): an on-line, worldwide atlas for accessing, modelling and mapping marine biological data in a multidimensional geographic context, *Oceanography*, 13(2000) 5-7.

5 Paterson, G., Boxshall, G., Thomson N. & Hussey, C., Where are all the data? *Oceanography,* 13 (2000) 21-24.

6 Costello, M.J. Developing species information systems: the European Register of Marine Species (ERMS). *Oceanography*, 13(2000) 48-55.

7 Zhang, Y & Grassle, J.F., A portal for the Ocean Biogeographic Information System. *Oceanologica Acta* 25 (2003) 193-197.

8 Blum, S., Vieglas, D. & Schwartz, P.J. Distributed Generic Information Retrieval. Available at http://DiGIR.sourceforge.net/events/20011106/DiGIR.ppt (2001)

9 Rees, T., .C-squares – a new metadata element for improved spatial querying and representation of spatial dataset coverage in metadata records (abstract). EOGEO Proceedings, Ispra, Italy, 2002. [http://eogeo.jrc.it], (2002).

10 Stockwell, D. WhyWhere - answering the question "Where is it and why?" on a global scale. [http://landscape.sdsc.edu:8080/BIODI/Wiki/WebMapModellingSystem/tutorial.html/] 2003.

11 Fautin, D.G. & Buddemeier, R.W., Biogeography of the Hexacorallia: http://www.kgs.ku.edu/Hexacoral/ (2003)