Interoperability for Geospatial Analysis: a semantics and ontology-based approach

Zarine Kemp[1], Lei Tan[1] and Jacqueline Whalley[2]

[1]Computing Laboratory, University of Kent
Canterbury, Kent CT2 7RY, U.K.

[2]School of Computer and Information Sciences, Auckland University of Technology,
Private Bag 92006, Auckland 1020, New Zealand

Z.Kemp@kent.ac.uk

Abstract

Information extraction and integration from heterogeneous, autonomous data resources are major requirements for many spatial applications. Geospatial analysis for scientific discovery involves identification of relevant information resources, extraction and fusion of requisite subsets of the information, application of spatial analytical techniques and visualization of the results in an appropriate form. The motivating application domain underlying this research is marine environmental management, although the principles discussed apply to a wide range of scientific disciplines. The research discussed in the paper focuses on integration of data sources, data exploration and interactive data analysis. A knowledge base is used to capture the semantics of the spatial, temporal and thematic dimensions at a domain level, and the context-aware framework exploited to meet the requirements of a varied and distributed user community with differing objectives.

Keywords: information fusion, geospatial analysis, knowledge base, ontologies, visualization.

1 Introduction

Information technologies such as the Internet and Grid computing have revolutionized the way that data resources are discovered and shared. In application domains dependent on geospatial and scientific information, reuse, sharing and dissemination of data is a major requirement. These information repositories are maintained by autonomous organizations, are heterogeneous in structure and semantics and are used by researchers and decision-makers in various contexts and from different perspectives. Interoperability of data and services underpins the next phase of the World Wide Web. Research in distributed databases, integration of structured and semi-structured data and technologies for mediator and information brokers have enabled *syntactical* and *structural* heterogeneities to be overcome. Issues relating to *semantic* heterogeneity are also being tackled using metadata, ontologies and thesauri to express salient concepts and knowledge within a domain of discourse.

In this paper we describe the architecture and framework of a system for environmental information systems. We suggest that in the context of geospatial information systems a data integration approach based on a global monolithic view of data, and a foundational ontology, is not an appropriate solution. We propose an architecture that provides interoperability, querying and analysis capabilities for a community of researchers while maintaining the autonomy of participating data sources. The middleware framework uses an adaptable, scalable knowledge base to accommodate semantic heterogeneity and provide analysis services.

The next subsection describes a motivating application and the data sources in the test bed. Section 2 discusses system requirements and related work. Section 3 presents the system architecture and details of the knowledge base. Section 4, illustrates the interaction model using example queries and section 5 concludes the paper.

1.1 Motivating Application

The system discussed in this paper is based on a platform for marine research and decision support but the requirements and principles are equally applicable to a wide range of application areas. It is intended as a research hub for a community of scientists who pool their information resources and use analytical and visualization tools for monitoring and understanding the marine ecosystem. For example, users may wish to retrieve detailed information about the fishing industry, study phenomena such as algal blooms, explore the changes in biodiversity in a particular part of the ecosystem, retrieve applicable legislation or investigate the effects of anthropogenic activities on particular marine species.

We discuss, briefly, the content and structural characteristics of the data sets in the research test bed emphasizing the geo-referenced attributes of the information stores.

Industrial activity data: the two main activities are fisheries and aggregate dredging for the building industry.

Management of fishing activities is regulated by the Common Fisheries Policy (CFP) legislation of the European Union using sea areas defined by the International Council for the Exploration of the Sea

(ICES). Quotas are allocated by country, species and marine area; these are *ICES Divisions* defined in *vector* format. Data relating to fishery harvesting activities are held in national databases by haul, spatial reference, date/time and species/weight. The spatial reference type in this case is by *ICES statistical rectangle*, a standard grid defined for all EU waters, forming *a hierarchical subdivision* of the quota divisions (illustrated in Figure 2).

Aggregate dredging: these are *vector-defined areas* where licences have been granted for extraction of material from the seabed. Environmental impact assessment reports and research papers may be associated with these activities.

Research data: Annual surveys are conducted by research centres based in different countries. The data sets consist of environmental information such as sea surface temperature, salinity, seabed type and biomass abundances by species. The location of sampling stations (*geo-referenced point*) is stored with the primary data sets to enable variables to be subsequently interpolated over the spatial extent of interest using an appropriate interpolation technique.

Ad hoc surveys: for example of benthic fauna provide data sources at fine spatial resolutions and are stored as *point samples* in the database.

Other related data: Legislation applicable to activities, species and habitats in marine environments. The statutes refer to areas directly using geographic coordinates or, indirectly, by reference to habitats for endangered species.

Base data of the geography of the research area including coastlines, ports and rivers are held in *vector format* using standards such as ESRI .shp files (ESRI).

In addition to the data sources, marine researchers and managers subscribe to domain-related ontologies. We have included two typical global ontologies: an ontology of marine species which consists of a tree-structured biological taxonomy and a more complex marine habitat multilevel classification that is becoming a European standard (Connor *et al* 2004).

2 Requirements and Related Work

The primary role of the middleware is to provide the abstractions and services that enable the development and deployment of user-level applications in a heterogeneous, distributed, computing environment. It must also be geared to the geospatial requirements of marine environment research communities as described by Tsontos and Kiefer (2003). From the computational perspective, the system should:

- support discovery of, and access to distributed, heterogeneous information sources
- provide tools for representation, manipulation and visualization of spatiotemporal and scientific data
- be adaptable to enable data sources, semantic information and services to evolve according to the requirements of the research community.

Halevy *et al* (2003) discussed the notion of *interoperability* across the *structure chasm,* that is, over sources that encompass *structured* and *unstructured* data. More recently, the notion of *dataspaces* has been proposed as a data management abstraction with associated *DataSpace Support Platforms* (DSSPs) to provide the required services (Franklin *et al* 2005). The middleware described in this paper encompasses several requirements of dataspaces using capabilities of extended database management systems. Interoperability is based on XML-based mediation techniques for data sets in relational or object-relational databases (Wiederhold 1999). The knowledge base enables links between data sources to be represented and supports keyword-based information retrieval and querying.

A major characteristic of computation in the geographic sciences domain is the pervasiveness of the *space-time-theme* composite. Understanding phenomena in geo-scientific domains requires queries and analyses to be predicated in terms of these three dimensions. Theories underlying these dimensions and their representation in data management systems have been discussed by researchers including Buckland and Lancaster (2004) and Smith and Mark (1998). A consequence of this is that interoperability platforms have to incorporate an understanding of, and mappings between, different conceptual views of space. Details of these are beyond the scope of this paper but reference may be made to international standards for geospatial data such as ISO 19115 (2003) and OGC (2003) and various classifications of space such as the object and field view space or the vector-raster views of space, reflecting alternative conceptual spatial representations. In the marine domain classifications of space may also involve complex hierarchies such as the Joint Nature Conservation Committee habitat classification (Connor *et al*, 2004). Figure 1 illustrates a small subset of this classification.

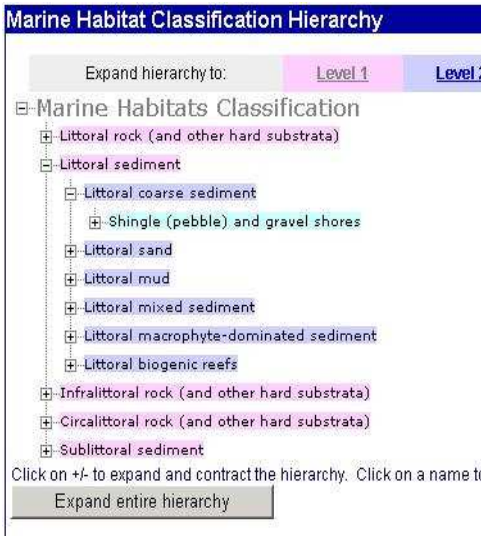

Figure 1. JNCC habitat classification (subset)

Similarly, multiple representations also arise in the case of the temporal and thematic dimensions. The middleware platform in this research uses ontologies in the knowledge base to provide mappings between

alternative spatiotemporal classifications as discussed in Kemp and Frank (2005).

Users in a research community use their expertise and experience to guide and inform the data they collect and the analyses they undertake. Bouquet *et al* (2004) suggest that application domain knowledge may be included by 'contextualizing ontologies'. We propose extending global ontologies to incorporate community (local) or regional context using the knowledge base, as illustrated in the following example.

A marine research community in the UK may include data on fishing activity which refers to regulatory areas for quota allocations and recording of catch statistics. These areas are specific instances of a generic vector-defined marine area feature. Extending the spatial ontology to include this contextual knowledge makes this semantic information explicit, facilitates data interoperability and enables users to query and analyse the information in a meaningful way. The map in Figure 2 shows a spatial containment hierarchy of marine areas referred to in section 1.1. It shows ICES Divisions (large non-uniform spaces), ICES rectangles (cells within each division) and the fine scale research grid (Eastern English Channel).

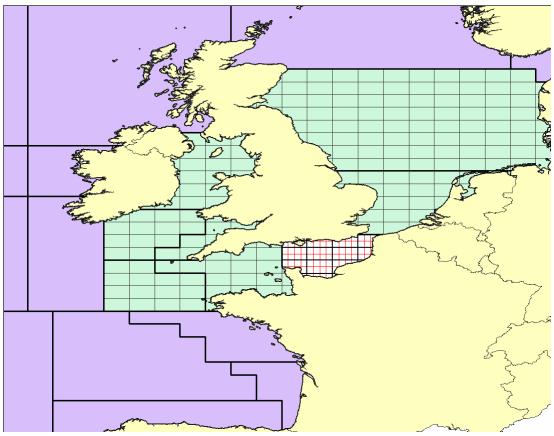

Figure 2: A nested hierarchy of marine areas

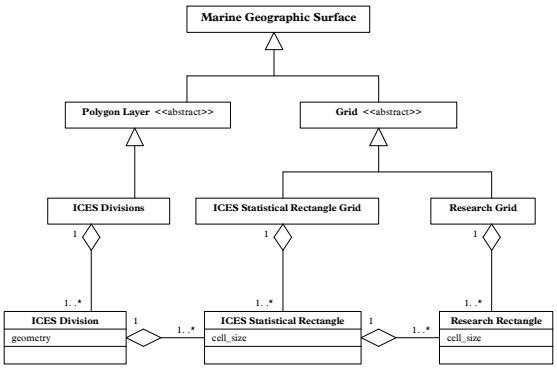

Figure 3. Part of the spatial ontology

Figure 3 shows part of the spatial ontology in the knowledge base (using the OGC standard) which has been *extended* at the lowest level to include additional semantics of marine space of relevance to the research community.

Ontologies have been used in many applications to enable shared concepts to inform the research. Gangemi *et al* (2002) describe a detailed ontological framework for fisheries. This FAO (United Nations Food and Agriculture Organisation) initiative provides a platform for unifying different thesauri, topic trees and taxonomies to provide a formal, integrated ontological framework for the fishery application domain. The semantic framework from this initiative is similar to the ontology of the thematic dimension included in our knowledge base. In our system, spatial conceptualizations have to be integrated with other dimensions, temporal and thematic. Wadsworth *et al* (2005) identify a problem with using ontologies, namely that of reconciling, alternative overlapping conceptualizations. They describe a semantic-statistical methodology for quantifying overlaps to resolve the problem.

In many information systems, the ontologies, i.e. the semantics underlying retrieval and querying of data are completely hidden from the user. However, In some situations scientists need to traverse the concept hierarchies to enable them to specify the parameters for tasks that constitute the workflow. In this system we have enabled a navigational form of querying where some level of semantic information has to be identified. A simple example involves querying a data set using a variable that is an element of a hierarchical classification. The global taxonomy of marine species is such a hierarchical classification. If the user wishes to query a data set at the 'genus' level, then the user can indicate that aggregation level for the analysis. The system deduces that the concept 'species' in the data sources is subsumed under the more general category 'genus' and returns the aggregated values as required. The design of the knowledge base also enables representation of the semantics of different types of relationships.

3 Architecture of the Framework

3.1 Overview

In this section we present an overview of the prototype system that was implemented for this research. Much of the functionality provided by the framework consists of dynamic composition of data and services. Typical services consist of:

- flexible extraction of subsets from the heterogeneous resources, dependent on user-specified parameters
- data discovery at different levels of abstraction
- sub-sampling, reclassification and re-gridding of extracted data, if required
- processing data by applying computational models
- visualization of output in textual, tabular, graphic or cartographic format.

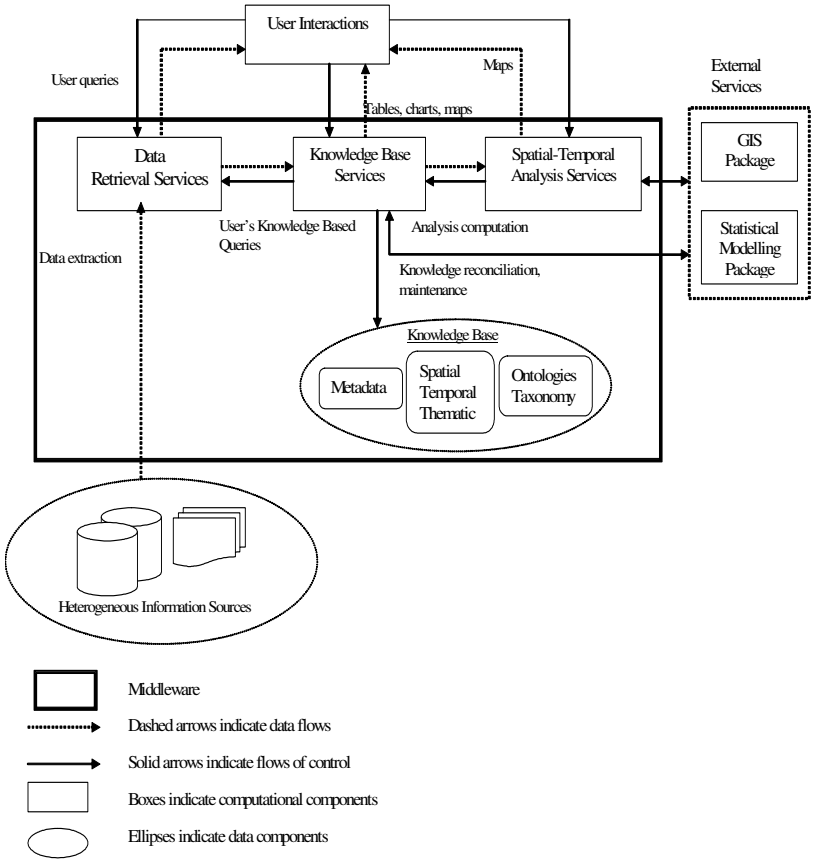

Figure 4. Architecture of the framework – system components interaction

The diagram in Figure 4 provides a high-level view of the system architecture showing the interaction between the computational and data components of the system:

- The *Data Retrieval Services* provide the functionality for querying different data sources. Users' queries are accepted and mediator services used to retrieve and combine the required information from the data sources.
- The *Knowledge Base Services* provide more advanced searching and functionality. These services include ontology-based searching, linking data sources using thesauri and spatial-temporal indexing.
- The *Analysis Services* provide computational capabilities including linking to external software packages for enviromental analyses using the data extracted by the *Knowledge Base Services*. The external capabilities include a GIS (Geographical Information System) and a statistical modelling package.
- *User Interaction Services* provide users with a graphical interface to enable specification of parameters and present results in appropriate formats.

The system components identified in Figure 4 have been implemented as interacting packages. Each package consists of services that implement the functionality provided by that package. The main system components include: the QueryGUI package that encompasses the interaction model of the framework, the MediatorWrapper package that provides the functionality for linking disparate data sources as required by specific tasks and the XMLDataLink package that provides the utility classes to interpret XML documents.

3.2 User Interaction Services

In the system framework, an XML format file is designed to represent users' queries, and data access services are implemented as a set of Java objects. An example XML query file is shown in Figure 5. The XML files are parsed according to the DOM (Document Object Model) standard using the JAXP (Java API for XML Parsing) (JAXP, DOM , Ungerer and Goodchild 2003). The links to relational databases are implemented by the JDBC to ODBC driver provided in the Java library. Thus the data sources can by distributed over a network environment.

Figure 5. XML file representing a query.

The QueryGUI package provides the functionality for user interaction. Users interact with the user interface of package to submit queries to the underlying data sources via the MainFrame object. The AddQueryButton adds a new QueryPanel to the MainFrame. The RemoveButton will remove the last QueryPanel from the MainFrame. The user can add as many QueryPanels to the MainFrame as required. The QueryPanel enables users to specify the data sources, concepts (i.e. variables) and logical conditions involved in a query. Options are also provided to so that users can specify the style and format of the displayed results.

3.3 Data Retrieval Services

The main purpose of this interface is to transform the XML query to the underlying data source query language. Each information source implements a wrapper interface called *DataSource* that must be registered with the *Mediator*. Wrappers consist of the structure specification of a data source and an understanding of the transformation between XML documents and the underlying data.

The MediatorWrapper package implements part of the wrapper-mediator methodology for interoperable data sources as described by Wiederhold (2000). When the user clicks on the "Execute" button on the user interface, the query is generated as an XML file and sent to the *Mediator*. The *Mediator* interprets the XML file, sends it to the relevant data source wrapper according to the content of the XML file, and combines the results returned from each data source for the user. The mediator does not have a global mediated schema that is shared by all the participating data sources. It is only aware of the data sources that are registered currently. It also gets relevant information from the XML query file to rearrange and merge the results returned from the data sources. The results returned from each data source are *ResultSet* type objects; they are combined into an *ArrayList* of *ArrayLists* object by the *combineResult2* method in the *Mediator* class.

The *XMLDataLinkUtility* package provides the utility classes that can be used by other classes, such as the *XMLDataUtility*; it contains methods to interpret XML documents. The *DataConcept* class represents the thematic concept on which a user's query is based, such as 'year', 'species', 'abundance'. There are classes in this package to map the terminologies used by users in the *DataConcept* objects to the equivalent terms used in the individual data sources. The *DataLink* classes also have the necessary knowledge of the underlying data sources, schemas or structures to form a query. If the underlying data source is a relational database, the associated *DataLink* object holds information about relation names and attribute names in each relation, and the links between relations. If the underlying data source is in XML format, the *DataLink* object is aware of the document type definition associated with each XML file.

3.4 Spatiotemporal Analysis Services

The Spatial-Temporal Analysis Services add computational capabilities for the specialized analytical tasks required by marine scientists. The example modelling and simulation task used in the prototype involves calculating Habitat Suitability Indices (HSI) for different marine locations for a species. These indices are then used to arrive at a classification of habitats for that species, presented to the user in cartographic format.

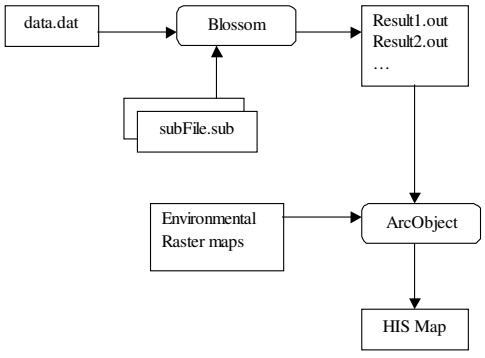

Figure 6. HSI Workflow diagram

Figure 6 illustrates the workflow involved in the analysis process in this component. The user interacts with the main interfaces to specify parameters for the analysis: environmental variables, temporal range and species. The software extracts relevant data from the data source, generates an ASCII text file in the format required by the statistical package Blossom (2005). The application software interacts with Blossom commands to carry out the analyses. The output returned by the Blossom package is interpreted and used to produce Habitat

Suitability maps. Several map generating and manipulation functionalities are used in producing the final and intermediate maps: interpolation methods are used to generate raster maps for the environmental variables, and map algebra techniques are used to combine the separate environmental maps into the Habitat Suitability Index maps.

3.5 Knowledge Base Services

The Knowledge Base, discussed further in the subsection 3.5.1, maintains a repository of ontologies that represent users' understanding of relevant domain concepts. The metamodel incorporates semantics and links between the global ontologies (biological taxonomies, habitats etc.), ontologies for dimensions space, time and theme and the underlying data collections. The Knowledge Base Services provide the functionality that enables users to express queries in terms that are relevant to them. The code in this component is similar, at the design level, to the services in the *QueryGUI* package (section 3.2).

The user interaction model is an important aspect of this component. Its interface includes browsing capabilities for two reasons: first, it reveals to users the classification schemes and hierarchies available to allow for semantics-based querying (illustrated in section 4.1) and second, it enables them to identify the level of the hierarchy required. This selection prompts the system to generate aggregate variable values at the required level of abstraction. This feature is illustrated in the example in section 4.2.

Another important aspect of the Knowledge Base services functionality is to resolve semantic differences between data sources. The case study in section 4.3 serves to illustrate this feature using the example of inconsistency in the classification of the environmental variable *sediment.* The knowledge services use a reconciliation method to reclassify the hierarchy in the data resource and map it to the global domain classification for marine habitats.

3.5.1 The Knowledge Base

Georeferenced digital libraries and web-based search engines as described in Janee and Frew (2002) are frequently underpinned by carefully curated ontologies and gazetteers. In the case of our system framework, the data schemas, diverse ontologies, classifications, taxonomies and thesauri all represent relevant information. Unifying a wide collection of semantic fragments into a definitive well-crafted knowledge base is a major challenge as discussed by Frank and Kemp (2001). Another characteristic of the diverse data resources is that they overlap in their content to varying degrees. In order to accommodate the structural and semantic diversity and to provide links between the data sources in the application and the scientific domain related concepts we provide layered conceptual *domain knowledge mode*l. In addition to articulating the semantics at various levels of abstraction, our framework encapsulates the associated services required for interoperability in multidimensional, hierarchical information spaces (Kemp and Lee 2000). Zaslavsky *et al* (2003) describe a similar system based on the Open GRID Services Architecture as a community cyberinfrastructure.

The knowledge base consists of a layered structure that hides the complexity and diversity of the information resources from end users. It consists of three types of objects: metadata objects, dimension ontologies and global or domain ontologies.

The *metadata layer* consists of metadata objects (one per information source) that encapsulate collection or document level information about each data source conforming to standards prevailing in the marine geoscience community. They contain administrative and access information, details of the provenance of the data sources, lineages and approximations of the spatial and temporal extents of the underlying data. These coordinates enable quick 'first pass' searches over the data sets available in the information base. Metadata objects also include information on the format/data type of the spatial and temporal attributes in the collection to determine the appropriate level of spatial integration when data are extracted from more than one resource. Many geoscientific data portals provide metadata views for tasks such as data discovery, access services and indication of fitness-for-use. Our knowledge base enables each metadata object to be linked with one or more of the types in the dimension ontologies to enable access to a range of spatial and temporal services.

The components of the *dimension level metadata* objects perform two functions. They articulate the domain concepts that enable users to specify the spatial, temporal and thematic parameters relevant to queries. They also provide links to the underlying information sources to enable transparent interoperability over the different data sets. As most queries in environmental analysis examine attributes with reference to the space-time-theme composite, three ontologies have been provided at this level: the spatial hierarchy, the temporal hierarchy and the thematic classification. Figures 2 and 3 (in section 2) illustrate part of the spatial ontology specialized by community-related context. Each spatial class in the ontology (Figure 3) is represented by its type_name, textual label, textual description and structural and functional specification. The classes in the bottom layer of the ontology instantiate the aggregation semantics. Thus, for example, a particular ICES Division may be identified by its code (VIId), its complete or part textual description (Eastern English Channel), its defining coordinates (MBR: minimum bounding rectangle) and by direct interaction at the user interface. The ontology specification also enables the aggregation of aspatial attributes of spaces *contained_in* the specified area.

Similarly, the temporal hierarchy can provide several perspectives on time. For example, a linear temporal view enables investigation of phenomena using operators based on temporal logic such as *overlap, touch, disjoint* and so on. An alternative classification may be based on seasons as shown in Figure 7.

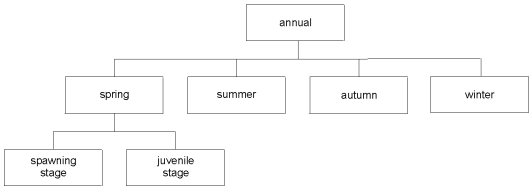

Figure 7. Temporal seasonal classification

The seasonal classification is relevant for the framework and could be implemented using either lookup tables or functions depending on the spatial extent of the analysis. In the current implementation, temporal attributes are identified simply as *time points* and *temporal intervals* with associated operators.

The thematic ontology consists of a hierarchy of textual terms classified according to the main categories of information sources in the testbed: 'fisheries', 'legislation', 'research', 'benthos' and 'dredging'. Concept terms link to other domain related concepts, using *hierarchical* (*broader/narrower term*), *associative* and other relationships. Some of these concepts are also linked to the global ontologies in the knowledge base. For example, linking the concept 'species' to the global taxonomy for marine species, provides access to the semantics of the biological taxonomy.

The *reference ontology layer* contains the global semantics applicable to the marine domain. Information at this level refers to global repositories such as those supported by the Global Biodiversity information Facility (GBIF), Taxonomic Databases Working Group (TDWG) and the Ocean Biodiversity Information System (OBIS). In our prototype system we have included a structured biological taxonomy of the species that occur in the marine area of interest and are used for various tasks such as providing the infrastructure for associating different common names for scientifically identified species, aggregating data at various levels of the hierarchy and indirectly enabling the underlying data sets to be linked for ad hoc analysis. A more complex example of an ontology at this level is the detailed classification of marine habitats (illustrated in Figure 1). This ongoing European initiative on defining a classification for marine spaces, starts with fairly coarse classifications based on a few major physical parameters and proceeds through successive levels of refinement to include topographic features and biotic communities associated with the ecological units. In the current version of the project the main use of this ontology is to enable users to define habitat suitability indexes for relevant species depending on the variables available in the underlying data sets.

4 Geospatial information retrieval: case studies

In this section the capabilities of the research framework are illustrated using typical queries and analysis tasks.

4.1 Interoperability of data sources

The first example illustrates retrieval of data from two heterogeneous fishery databases. The query parameters specify retrieval of:

- Theme: *fishery*, subtheme *catch*
- Theme: species *Solea solea*
- Time: *temporal interval* (calendar dates)
- Space: *ICES rectangle* (the second level of the spatial hierarchy, illustrated in Figure 2)
- Display mode: *map*

Figure 8 illustrates the output showing the requied variables from two separate national databases.

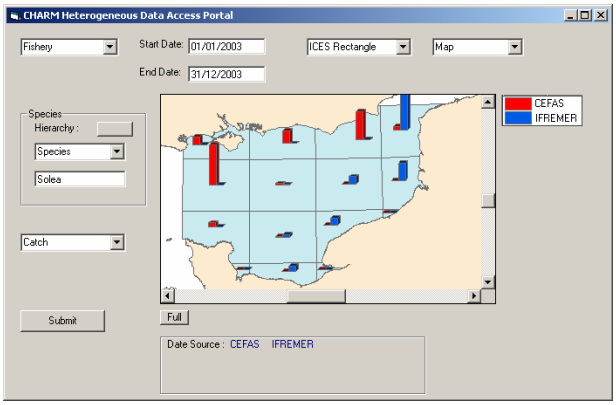

Figure 8. Data from heterogeneous databases

4.2 Aggregation of data at user-specified level of concept hierarchy

This example illustrates the interaction of an information source with one of the semantic hierarchies in the knowledge base, the biological taxonomy. The user has navigated the taxonomy and selected the genus *Loligo* as the aggregation level for thematic information to be retrieved. The abundance values refer to the data collected in annual research surveys. Figure 9 shows the abundance of this genus (all species aggregated), from the identified input source, in cartographic format. In this case, the visualization of the *spatial* dimension is in *point* form, with the size of the icons of the sampling locations reflecting the relative values of the abundance. The ICES grid is superimposed on the map for visual reference, for example for industrial catch of the same genus.

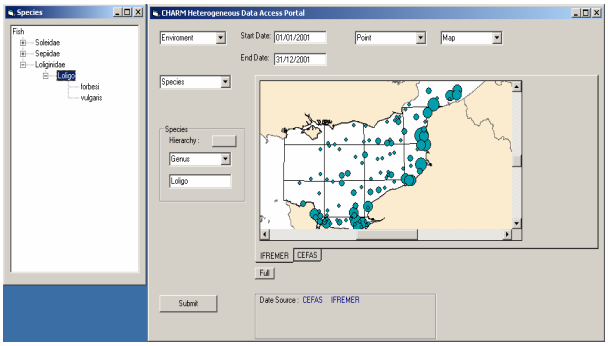

Figure 9. Aggregated abundance of selected genus

4.3 Visualization based on ontology-defined classification

This example illustrates the use of a domain level (global) ontology, the habitat classification (Connor *et al,* 2004), to reclassify an alternative classification in one of the data sources. The ontology includes seabed sediment classes in its definition of marine areas at level 2 (see Figure 1). The first frame, Figure 10 (A), shows a small subsection of this classification where the class 'Littoral sediment' (LS) is further subdivided into three subclasses, 'Littoral coarse sediment' (LCS), 'Littoral sand' (LSa) and 'Littoral mud' (Lmu). This particular research survey database uses an alternative sediment classification system (Larsonneur *et al* 1979), which contains four subclasses at this level: 'Coarse sand', 'Fine sand', 'Gravel and pebbles' and 'Mud' as shown in Figure 10 (B). When this data set is used locally, this classification is appropriate. However, when it is integrated with data sets from other national data sets, the global ontology is used to *reclassify* it to achieve semantic consistency. Figure 10 (C) shows the same data in map form where the original categories 'Fine sand' and 'Coarse sand' have been merged for equivalence with the ontological class 'Littoral sand'. Thus the framework enables individual databases to maintain local heterogeneity and also provides a reclassification service, when required, for global interoperability.

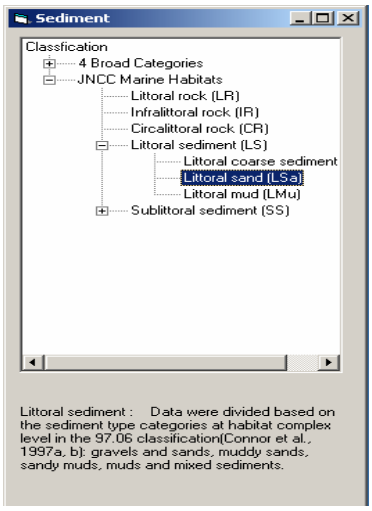

Figure 10(A) Global ontology

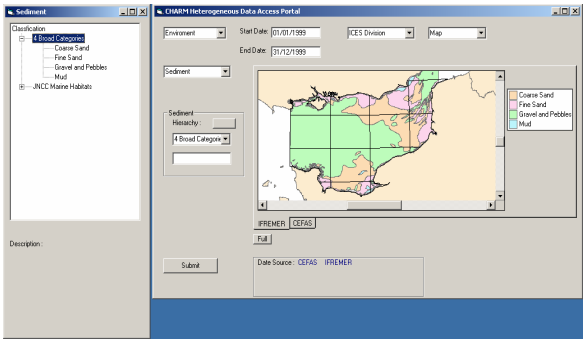

Figure 10(B) Local classification (4 sub classes)

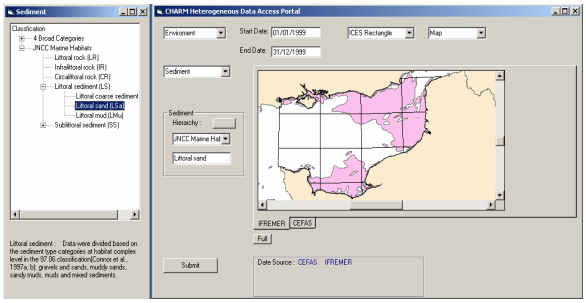

Figure 10(C). Reclassification of sediment types (3 subclasses)

4.4 User-specified spatial search and multiple thematic retrieval

This example illustrates the discovery of multi-theme data and related information. User interaction in this example starts with an interactively specified rectangle of interest as shown in the upper window of Figure 11.

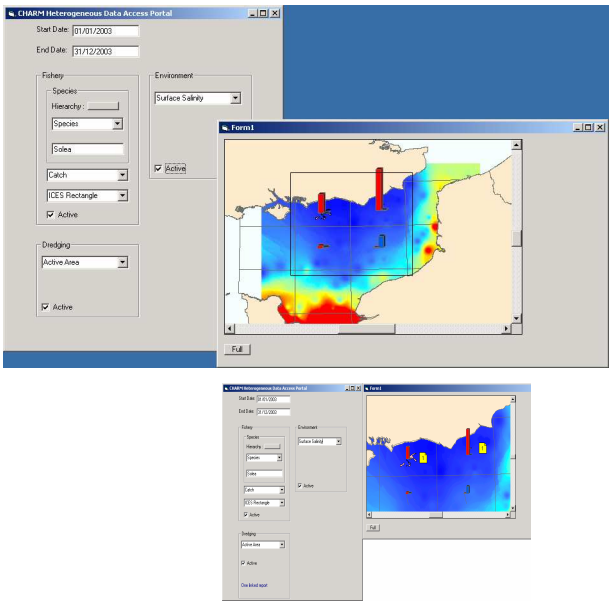

Figure 11. Multi-theme output and retrieval of linked information

The system reveals the data sources relevant to the search space and enables the user to refine the query by specifying required parameters. In this example, the environmental variable *surface salinity* is displayed as a raster map (interpolated from point samples in the research database), overlaid with information relating to *fishery catch* data for species *Solea solea* by *ICES rectangles*. Unstructured information in the user-specified search rectangle referring to active dredging areas is also displayed in the lower window. The knowledge base enables the system to discover that the dredging areas have *association* links with text documents (research reports). The existence of the documents is indicated on the map using document shaped icons which also indicate

the number of relevant documents discovered (1 in this case). Clicking on the hyperlinks displays the contents of the documents.

5 Conclusion and future work

Initial results of this research are promising. Flexible and open-ended support for scientists and decision-makers can be provided by enabling interoperability across dispersed heterogeneous information sources coupled with appropriate metadata and semantic knowledge.

The future of the World Wide Web will involve scientific domains with a large number of existing metamodels and ontologies (Costello and Vanden Berghe 2006). There will also be increasing requirements to extend or contextualize existing ontologies and map between different ontological specifications. A related requirement is to enable the ontological resources (knowledge bases) to evolve and be easily updated, as data sets and metadata models are added to the resources for a research community (Reinoso-Castillo *et al* 2003). In our system, the collection level metadata and the ontologies are part of the knowledge base which functions as a community resource at a central hub. This makes it easier to extend the range and type of information resources and related semantics available to researchers in a scientific domain.

There are several interesting directions for future work.

Ontologies and reasoning: Investigation of formal ontology specification languages to enable reasoning with multiple ontologies in complex scientific domains.

Real-time response: Many environmental monitoring tasks such as early warning systems for natural hazards require real-time responses. Wiederhold (2000) has suggested that real-time simulations should be an integral part of decision-making frameworks. A future extension of this research will consider the design and performance implications of this requirement.

Platform for creating and maintaining the knowledge base: It would be useful to include an interface to enable users to discuss and update the knowledge model as a collective activity. This could be similar to a 'blackboard' component in decision support systems. Such a facility would encourage cooperative decision-making and thus be of interest in areas such as the marine domain where there exist recognized conflicts of interest between user groups such as the fishing industry and marine biologists.

Ontologies and reasoning: Investigation of formal ontology specification languages to enable reasoning with multiple ontologies in complex scientific domains.

Distributed framework: The design of the framework and the underlying technologies assume that the data, users and computational framework are distributed. To achieve a disciplined model for the framework, standards such as WSDL, UDDI and SOAP should also be investigated as well as the promise of the GRID architecture for a federated infrastructure as discussed by Watson (2005).

6 References

Blossom: Blossom Statistical Software. US Geological Survey. http://www.fort.usgs.gov/products/software/blossom/blossom.asp Accessed 8 Aug. 2006.

Bouquet, P., Giunchiglia, F., van Harmelen, F., Serafini, L. and Stuckenschmidt, H. (2004):Contextualizing ontologies. *Journal of Web Semantics* 1(4): available online: http://www.websemanticsjournal.org/ps/pubs/2004-20

Buckland, M. and Lancaster, L. (2004): Combining Place, time and Topic. *D-Lib Magazine* **10**(5), ISSN 1082-9873.

CFP: The European Union Common Fisheries Policy.http://ec.europa.eu/dgs/fisheries/index_en.html

Accessed May 2006.

Connor, D.W., Allen, J.H., Golding, N., Howell, K.L., Lieberknecht, L.M., Northen, K.O. and Reker, J.B., (2004): The Marine Habitat Classification for Britain and Ireland, Version 04.05 JNCC, Peterborough, ISBN 1 861 07561 8.

Costello, M. and Vanden Berghe, E. (2006): 'Ocean Biodiversity Informatics': a new era in marine biology research and management. *Marine Ecology Progress Series* **316**: 203-214.

DOM: Document Object Model, http://www.w3.org/DOM/

ESRI: Environmental Systems Research Institute, http://www.esri.com/ Accessed Jan. 2006.

Frank, R. and Kemp, Z. (2001): Ontologies for knowledge discovery in environmental information systems. Proc. *Workshop on Complex reasoning on geographical data (CRGD),* Raffaeta, A. and Renso, C. (Eds). December, 2001, pp 15-30, Paphos, Cyprus, 2001.

Franklin, M., Halevy, A. and Maier, D. (2005): From Databases to Dataspaces: A New Abstraction for Data Management. *ACM SIGMOD* **34**(4): 27-33.

Gangemi, A., Fisseha, F., Pettman, I., Pisanelli, D.M., Taconet, M. and Keizer, J. (2002): A Formal Ontological Framework for Semantic Interoperability in the Fishery Domain, *Proc. ISCW 2002, International Semantic Web Conference*, June 9-12, Sardinia, Italy.

GBIF: Global Biodiversity Information Facility. http://www/gbif.org/ Accessed Aug. 2006.

Halevy, A., Etzioni, O., Doan, A., Ives, Z., Madhavan, J., McDowell, L. and Tatarinov, I. (2003): Crossing the Structure Chasm. *Proc. Conference on Innovative Database Research (CIDR 2003),* Asilomar, CA, USA. January, 2003).

ICES: International Council for the Exploration of the Sea, http://www.ices.dk/ Accessed May, 2006.

ISO: International Standards Organization. http://www.iso.org/ Accessed Jul. 2006.

Janeé, G. and Frew, J., (2002): The ADEPT Digital Library Architecture, *ACM / IEEE Joint Conference on Digital Libraries*, July 2002, Portland, Oregon.

JAXP: Java API for XML Processing. http://java.sun.com/webservices/jaxp/ Accessed June 2005.

Kemp, Z. and Lee, H. (2000): A Multidimensional Model for Exploratory Spatiotemporal Analysis. *Proc. 5th International Conference on GeoComputation*, Carlisle, B. and Abrahart, R. (Eds). University of Greenwich, UK.

Kemp, Z. and Frank, R. (2005): Knowledge representation and semantic interoperability in marine information systems. In *GIS/Spatial Analyses in Fishery and Aquatic Sciences (Vol. 2)*. Nishida, T., Kailola, P. and Hollingworth, C. (Eds). Fishery-Aquatic Research Group, Saitama, Japan. 735 pp. ISBN: 4-9902377-0-6.

Larsonneur, C., Vaslet, D., J.-P. Auffret. (1979): Les Sédiments Superficiels de la Manche, Carte Géologique de la Marge Continentale Française. *Bureau des Recherches Géologiques et Minières, Ministère de Industrie, Service Géologique National*, Orléans, France.

OBIS: Ocean Biogeographic Information System. .http://www.iobis.org/ Accessed Aug. 2006.

OGC: Open geospatial Consortium. http://www.opengeospatial.org/ Accessed Jul. 2006.

Reinoso-Castillo, J., Silvescu, A., Caragea, D., Pathak, J., and Honavar, V. (2003): Information extraction and integration from heterogeneous, distributed, autonomous information sources – A federated ontology-driven, query-centric approach. In *IEEE International Conference on Information Integration and Reuse*, Las Vegas, Nevada, 2003.

Smith, B. and Mark, D. (1998): Ontology and Geographic Kinds. *Proc. 8th International Symposium on Spatial Data Handling, SDH '98*, Poiker, T. amd Chrisman, N. (Eds). Vancouver, Canada, July 1998.

TDWG: Taxonomic Databases Working Group. http://www.nhm.ac.uk/hosted_sites/tdwg/ Accessed Nov. 2005.

Tsontos, V. and Kiefer, D. (2003): The Gulf of Maine biogeographical information system project: developing a spatial data management frameworkin support of OBIS. *Oceanologica Acta* **25** (2003): 199-206.

Ungerer, J. M. and Goodchild, F. M. (2002): Integrating spatial data analysis and GIS: a new implementation using the Component Object Model (COM), *International Journal of Geographic Information Science*, IJGIS, **16**(1): 41-53.

Wadsworth, R., Balzter, H., Gerard, F. George, C., Comber, L. and Fisher, P. (2005): Quantified Conceptual Overlaps: their use for reconciling inconsistent data sets using Siberian land cover as an example. *Proc. GISPlanet 2005 Conference*, 30 May - 4 June 2005, Estoril, Portugal.

Watson, P. (2005): Databases in Grid Applications: Locality and Distribution. In *Databases: Enterprise, Skills and Innovation*. Jackson, M. Nelson. D. and Stirk, S. (Eds). Lecture Notes in Computer Science 3567, Springer.

Wiederhold, G. (1999): Mediation to Deal with Heterogeneous Data Sources. In *Interoperating Geographic Information Systems*, Vckovski, A. Brassel, K. and Schek H-J. (Eds). Springer Lecture Notes in Computer Science 1580.

Wiederhold, G. (2000): Information Systems that Really Support Decision-making, *Journal of Intelligent Information Systems*, **14**(2) 85-94.

Zaslavsky, I., Baru, C., Bhatia, K., Memon, P., Velikhov, P. and Veyster, V. (2003): Grid-enabled mediation services for geo-spatial information. *Proceedings of NG2I-03, Workshop on Next Generation Geospatial Information,* Cambridge, Massachusetts, USA, October 19-21, 2003.