Metadata in Coastal Areas – Perspectives and Experiences

Wassilios Kazakos

FZI Forschungszentrum Informatik

Abstract

Catalogues have been discussed for centuries in libraries and it is generally accepted that finding relevant books in any science without a well organised catalogue is almost impossible. Nowadays the idea of a library with its catalogue entries is transported to all kinds of data, including books, maps, geographic layers, models. The correspondence to the catalogue entries are metadata, i.e. descriptions of the data. In this paper we present the general idea of cataloguing data electronically making them this way searchable for a broader audience. Additionally we present briefly three metadata system for coastal areas developed within the last four years, namely CoastBase, EUROSION and NOKIS.

1   Introduction

As described in the recent reflection paper of the European Commission "Towards a European ICZM Strategy" (1999), coastal areas and their natural resources (marine and terrestrial) have a strategic role to play in meeting the needs and aspirations of current and future European populations. Important examples of functional uses include: tourism, shipping, industry and energy, fishing and mariculture, coastal defence and natural development. It is essential that these functions can develop in a sustainable way to support conditions for human health, employment and regional development and environmental quality. The physical and administrative units related to coastal systems and its environmental problems may vary from local up to transboundary and supranational.

Authorities responsible for the management of these areas need to take both socio-economic and environmental aspects of these functions into account. Human pressures pose the risk of destroying habitats and the resource base of the coastal zones, and with them, the ability of the coastal zone to perform many of its essential functions. The EU Demonstration Programme for ICZM and above mentioned Reflection Paper suggest that an integrated coastal management is needed, including communication between all actors representing different sectors and between administrations in different levels.

In this approach access to all relevant information is essential. This information is in general dispersed among coastal institutes and organisations and neither readily available nor accessible. This lack of availability and accessibility of information hampers effective planning and decision making; the need to improve information finding in a distributed and highly heterogeneous environment is the motivation for the development of catalogue systems.

1.1   Outline

The paper is organised as follows. In Section 2, we introduce the basics. In Section 3, 4 and 5 we present briefly three main metadata projects for coastal areas, i.e. CoastBase, EUROSION and NOKIS. An outlook on future work conclude the paper in Section 6.

2   Basics

2.1   Catalogues

In order to react upon complex environmental problems, it is necessary to have information of various application fields at hand which in most cases will be located at different sites. Only if the respective information is sufficiently available it is possible to find complex interrelations, and to mutually use available information. Both the recollection of data that already has been collected and keeping data without any further use because of missing knowledge about the data can be avoided.

The answer to a certain question requires to learn *which* data is available, *where* the data are managed, *how* this data can be obtained, and *how* to interpret the data   correctly. The information that is necessary to obtain above mentioned information is called *metadata or meta-information*. This information can be compared to the information of classical index cards in library catalogues, which describe books but are not the books themselves. Due to the present explosion of the volume of data, it becomes even more important for the user to rely on information about existing data in order to find what he or she needs. Equivalent to the library catalogues which contain index cards, meta-information systems or *catalogue systems* are the more general electronic form. Typical examples of environmental catalogues are the German/Austrian Umweltdatenkatalog (UDK) and the Catalogue of Data Sources (CDS) of the European Environmental Agency. Both deal with descriptive information about environmental resources (Kazakos et al. 1998; Swoboda et al. 1999).

2.2   Information Integration

Integration of information from heterogeneous sources has been a major topic in the database related research for several years. Roughly speaking, there are two possible approaches (Widom 1996):

➤ The materialised (or Data Warehousing) approach. From each source, information that may be of interest is extracted/exported from each source and after filtering, harmonisation and fusion with information from other sources stored in a (logically) centralised repository (the data warehouse. User queries are evaluated at the central repository without connecting to the individual sources.

➤ The virtual (or mediated) approach. When a user poses a query, this query is sent directly (after necessary transformations) to the appropriate sources for evaluation. Their results are then filtered, harmonised and fused and presented to the user.

Whereas the materialised approach allows to optimise the response times for certain applications and the access reliability, it has serious drawbacks considering the updating of the repository when the information sources change in their content. Here the autonomy of the information sources has to be restricted such that an update policy can be put to work. The virtual approach on the contrary allows to leave the content at its original place. Changes to the information sources are immediately reflected in the query results.

Prominent virtual integration projects like TSIMMIS (Hammer et al. 1997; Garcia-Molina et al. 1997), Information Manifold (Levy 1998), or MIX (Baru et al. 1999) can be more or less accurately divided according to the I³ reference architecture's terminology (Arens et al. 1995) into three main logical components:

➤ Wrappers overcome the technical and syntactical heterogeneity of the individual sources.

➤ Mediators overcome the semantic differences (schema/information model) between the sources and fuse equivalent information artefacts.

➤ Facilitators select the sources needed to satisfy a given user need and combine them appropriately.

3   CoastBase

CoastBase started in January 2000 with 11 partners from 8 European countries. It aims to improve European marine and coastal management. The Project was supported by European Commissions IST programme (5th framework). The project is described in more detail in (Kazakos et. al. 2000) and

(Kazakos et al. 2001). One central part of the project was a virtual catalogue. It is designed to be flexible enough to be used not only in other coastal and marine applications, but also in other application domains.

The overall CoastBase system is based on a 3-tier multi-server architecture. The three tiers are the CoastBase client tier, the CoastBase server tier and the data abstraction tier. Client and server are further divided into sub-components according to the functional decomposition of CoastBase. These are catalogue, data access and manipulation, and feedback. In the data abstraction tier, the wrappers for the inventory data, as well as for the local repositories are implemented.
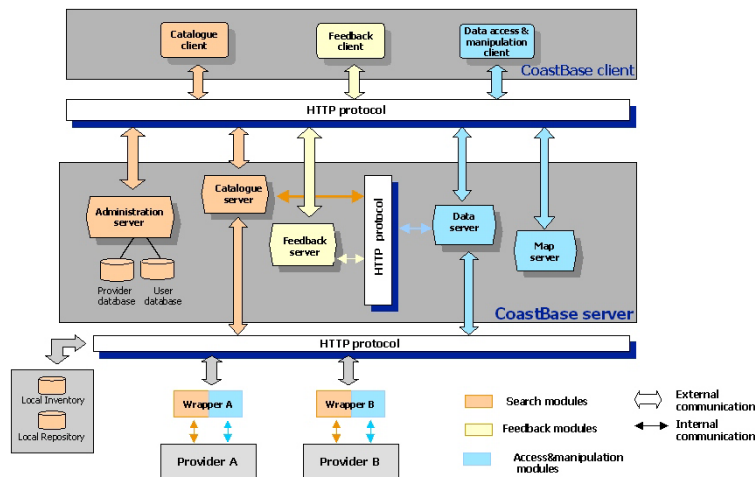

Figure 1: CoastBase general architecture.

The virtual catalogue of CoastBase covers the search and feedback modules. They cover three main functional units needed in every similar type scenario. The administration authenticates and authorises the user according to user groups. The query dispatching and retrieval of distributed inventory data is done by the catalogue server. The feedback server contains a update module, allowing the creation and actualisation of inventory data.

One of the major design goals of the overall system and especially of the virtual catalogue component was the extensive use of XML and XML-based techniques, in order to achieve maximum interoperability with other systems and the reusability of several parts of the system. The user interface for the catalogue client is in HTML and can be used by all typical browsers. The catalogue server implements the main functionality of the virtual catalogue that is able to access distributed, but homogeneous sources with the same domain model.

In order to build a scalable system that makes it easy to integrate new sources, a virtual integration approach was chosen. Except for a local repository into which products generated by CoastBase users are uploaded, there is no centralized database. In fact there is neither a centralized database for the data nor for the metadata. Instead, the existing Internet access points for the sources to be connected are used. The user queries are dispatched at run-time to the sources. In order to overcome the technical heterogeneity, wrappers that translate the queries into queries native to the source, adapt the access protocol, translate the results into XML and transform the structure to fit the CoastBase domain model are employed.

4 NOKIS

NOKIS (North See and Baltic See Coastal Information System; http://nokis.baw.de) is a joint project of KFKI (Kuratorium für Forschung im Küsteningenieurwesen) and BAW (Bundesanstalt für Wasserbau) with the main objectives to establish a metadata information system for the German

North Sea and Baltic Sea coastal regions according to the concept of an open system that permits participation of additional partners at any time.

Seven participating federal and state offices compile local metadata bases related to their data archives and integrate them via the internet. To support them, NOKIS provides a local prototype meta database which is installed in every participating institution and which is integrated via the internet. FZI developed the system for the NOKIS consortium.

The overall architecture of the NOKIS system consists mainly of local meta data bases and a central server located at BAW. In the local data bases, data sources are maintained by the participating institutions. They differ considerably in the data base technology applied to handle data sets and also in the availability of related metadata. There is a need for adding search techniques to the distributed data archives.

The overall concept of NOKIS is also suited to aid the local maintenance personnel and responsible parties who are in close contact with the region and its peculiarities to improve local data handling and documentation by creating appropriate metadata for efficient automated search operations. NOKIS participants are not expected to change their methods of archiving their data, documents, images, etc. They are, however, provided with uniform graphical user interfaces which are used for global search on the central server as well as for local search on their own data bases.

Locally created and maintained metadata bases are uploaded to the central NOKIS metadata server http://nokis.baw.de whenever the participating organizations consider it necessary to update the central metadata base. Any proprietary information is filtered out during this replication process. The contents of the central metadata archive represents the collection of all public meta information provided by participating organizations.

The essential task at the beginning of the project was to decide on the type of metadata which should be applied in a sustainable web portal for the coastal zone. Many proprietary attempts, which provide limited and very specific "data about data" are used in intranet environments. It is often argued that rather comprehensive international standards require too many entries to be successfully applied. However, efficient search facilities, which are needed to handle the information abundance require intensive preparatory effort for satisfactory performance. A number of standards has been developed during the last decade, which are in use world-wide for different purposes. They differ in the granularity of meta information which is an important issue for the functionality that can be built upon metadata.

In co-operation of more than 30 nations, the international metadata standard ISO 19115 has been developed. The Committee Draft 3 released in 2000 was taken as the basis for the NOKIS metadata model. The recommended core metadata for geographic datasets specified by ISO 19115 consists of only 12 mandatory and 11optional elements for information on responsible party, access, quality and description of datasets. The NOKIS metadata model declares 23 elements of the ISO metadata model mandatory to also accommodate information relevant for GIS applications. Thus the NOKIS metadata model is fully compliant with the ISO 19115 standard and guarantees that sufficient meta information is available for documentation and intelligent search methods.

The overall system is developed by following an XML-extreme approach as proposed in (Kazakos et al. 2001), i.e. using XML-technologies as far as possible, to achieve maximum reusability and easy configuration. We further extended our approach, and set up a system architecture that generates all components of a typical metadata repository out of an XML schema definition (XSD).

5   EUROSION

EUROSION is a project commissioned by the General Directorate Environment of the European Commission, which will result in policy recommendations on how to manage coastal erosion in Europe in the most sustainable way.

One quarter of the European Union's coast is currently eroding despite the development of a wide range of measures to protect shorelines from eroding and flooding. The prospect of further sea level rise due to climate change and the heritage of mismanagement in the past - such as inappropriate infrastructure - imply that coastal erosion will be a growing concern in the future. This is why the Directorate General Environment of the European Commission tendered the EUROSION project in 2001, which was won by a consortium led by the National Institute for Coastal and Marine Management of the Dutch Ministry of Transport, Public Works and Water Management. The implementation of the project started in January 2002. The project is expected to achieve its objectives by the end of 2003.

Through supporting the Integrated Coastal Zone Management Practitioners Network and facilitating access to relevant data and information, EUROSION offers a follow-up to the EU demonstration program on Integrated Coastal Zone Management - with an emphasis on pilot projects which focused on erosion management - and is consequently biased towards ICZM strategies.

The overall objective of EUROSION is to provide the European Commission with a package of recommendations for policy-making and information management practices to address coastal erosion in Europe, after thorough assessment of knowledge gained from past experiences and of the current status and trends of European coasts. However the project also aims at producing results of immediate value for policy makers and managers on other administrative levels.

Although EUROSION is a content oriented project one major concern was the publication and dissemination of the result through an internet based platform. disy and FZI are developing currently this platform based on the technologies of CoastBase and NOKIS projects. The general architecture is thus similar to CoastBase. In addition to CoastBase special focus was on incorporating descriptions (metadata) about geographic information, models and reports. To achieve this a metadata model was developed based on the ISO standard for geographic meta-data, ISO/DIS 19115. The system allows search and retrieval based on these meta-information and provides together with NOKIS one of the first Web-based metadata editors for ISO 19115 conformant metadata.

The overall system is developed by following in the main lines an XML-extreme approach as proposed in (Kazakos et al. 2001), i.e. using XML-technologies as far as possible, to achieve maximum reusability and easy configuration. We further extended our approach, and set up a system architecture that generates all components of a typical metadata repository from an XML schema definition (XSD). In addition we are now able to support multi-schema editing providing a customized metadata-editor for each type of information, i.e. from simple links to ISO-19115 compliant metadata for geographic information.

6 Conclusions

One of the major challenges integrated coastal zone management has to face is the distribution and heterogeneity of data needed for the management tasks. Metadata and meta-information systems provide the means to find available data but also to identify possible gaps in the data collections. In this paper we presented briefly three major approaches to search for data and information about coastal zones, focussing on technical aspects only, like distribution and Internet technologies. We believe, that future ICZM has to tackle the issue of setting up meta-information system and agreeing on metadata standards. CoastBase and NOKIS did an important step towards this direction.

References

Arens, Y., R. Hull & R. King (eds.) (1995): Reference Architecture for the Intelligent Integration of Information, Program on Intelligent Integration of Information, ARPA, Version 2.0.

Baru, C., A. Gupta, B. Ludäscher, R. Marciano, Y. Papakonstantinou & P. Velikhov (1999): XML-Based Information Mediation with MIX, Exhibitions Program of ACM SIGMOD.

Garcia-Molina, H., Y. Papakonstantinou, D. Quass, A. Rajaraman, Y. Sagiv, J. Ullman, V. Vassalos & J. Widom (1997): The TSIMMIS approach to mediation: Data models and Languages, Journal of Intelligent Information Systems.

Hammer, J., H. Garcia-Molina, J. Cho, R. Aranha & A. Crespo (1997): Extracting Semistructured Information from the Web, Proceedings of the Workshop on Management of Semistructured Data. Tucson, Arizona.

Kazakos, W., R. Kramer, R. Nikolai, C. Rolker, S. Bjarnason & S. Jensen (1999): WebCDS - A Java-based Catalogue System for European Environment Data. In Dogac, A., Özsu, M.T., Uluzoy, O (Eds): Current Trends in Data Management Technology, 234 – 249, Idea Group Publisher.

Kazakos, W., R. Kramer & A. Schmidt (2000): Coastbase - The Virtual European Coastal and Marine Data Warehouse. In Armin Cremers and Klaus Greve, editors, Computer Science for Environmental Protection '00. Environmental Information for Planning, Politics and the Public, volume II, 646-654.

Kazakos, W., R. Kramer, R. Nikolai, C. Rolker, S. Bjarnason & S. Jensen (1999): WebCDS - A Java-based Catalogue System for European Environment Data. In Dogac, A., Özsu, M.T., Uluzoy, O (Eds): Current Trends in Data Management Technology, 234 – 249, Idea Group Publisher.

Kazakos, W., A. Schmidt & H. Paoli (2001): XML based Virtual Catalogue Module in Coastbase. In Lorenz Hilty and Paul Gilgen, editors, Sustainability in the Information Society. 15th International Symposium Informatics for Environmental Protection, volume II, 513-520.

Lehfeldt, R., C. Heidmann & M. Piasecki (2002a): Metadata in Coastal Information Systems, in Holz, KP, Kawahara, M, Wang, SY (eds) Advances in Hydro-Science and –Engineering Volume 5. Proceedings of the 5th International Conference on Hydro-Science and -Engineering, Warsaw. Abstract Volume, p. 181.

Lehfeldt, R., F. Sellerhoff & M. Piasecki (2002b): Components of Web Portals in Coastal Engineering, in RA Falconer, B Lin, EL Harris, CAME Wilson, ID Cluckie, D Han, JP Davis, S Heslop (eds) Hydroinformatics 2002. Proc. 5th Intl. Conf., Cardiff. IWA Publishing, London, 1501-1506.

Levy, A. (1998): The Information Manifold approach to data integration, IEEE Intelligent Systems, September/October 1998, 11-16.

Remssbot Project Homepage, http://www.netor.gr/remssbot/

Sluyter, R., C. Potma, T. Krognes, M. Petrakis & P. van Hooydonk (1997): AirBase: 1997 Development Status and Extensions Foreseen, Second European Workshop on Air Quality Monitoring and Assessment, Brussels, 22-23 September 1997.

Swoboda, W., F. Kruse, R. Nikolai, W. Kazakos, D. Nyhuis & H. Rousselle (1999): The UDK Approach: the 4th Generation of the Environmental Data Catalogue for Austrian and German Public Authorities, Proc. IEEE Meta-Data'99, Bethesda, Maryland, USA, April 1999, http://computer.org/proceedings/meta/1999/papers/45/wswoboda.html.

Widom, J. (1996): Integrating Heterogeneous Databases – Lazy or Eager?, ACM Computing Surveys 28(4), December 1996.

Address

Wassilios Kazakos
FZI Forschungszentrum Informatik /DBS Database Systems Department
Haid-und-Neu-Str. 10-14
76131 Karlsruhe
Germany

E-mail: kazakos@fzi.de