A Decision Support System for Marine Applications

<u>Ralf Denzer</u>[a], Reiner Güttler[a], Sascha Schlobinski[a] and Jonathan Williams[b]

[a] Environmental Informatics Group (EIG) Saarbrücken, Germany, www.enviromatics.net,
Ralf.Denzer@enviromatics.org

[b] Marine Tech South, Southampton, United Kingdom, md@marinetech.co.uk

Abstract : The paper presents i-MARQ (Information System for Marine Aquatic Resources Quality), a project funded under the 5th European Framework Program. i-MARQ started in May 2002. i-MARQ focusses on the use of advanced integrated decision support for Marine applications. The project is developing enhanced decision support paradigms which can optimally exploit diverse data resources, and present probabilistic results in the geographic formats required by diverse decision-makers. Application of data fusion techniques permits forecasting of parameter best estimates and uncertainties, based on the underlying data statistics. This determination of information robustness is considered an essential ingredient of advanced environmental decision support systems.

Keywords: *Environmental Decision Support Systems (EDSS), EU FP 5, IMARQ, Marine Information System*

1. INTRODUCTION

Management of environmental risks in coastal and marine waters is increasingly important, as regulatory policy moves from discharge compliance alone towards impact minimisation. Decision-making (for operational and remediation management) requires access to best quality information on key environmental quality parameters. However, current data resources are fragmented and generally sparse both spatially and temporally. They also possess varying uncertainty.

The i-MARQ project is developing enhanced decision support paradigms which can optimally exploit diverse data resources, and present probabilistic results in the geographic formats required by diverse decision-makers. Application of data fusion techniques permits forecasting of parameter best estimates and uncertainties, based on the underlying data statistics. This determination of information robustness is considered an essential ingredient of advanced environmental decision support systems.

i-MARQ aims to combine 3 core components in a distributed environment: decision support tools, distributed, real-time sensor networks and state-of-the-art data fusion. This paper presents an early view of an anticipated system for decision support in the marine environment and will focus on the overall view of what is planned and needed.

1.1 Decision Support Systems

Environmental Decision Support Systems (EDSS) are usually defined as Environmental Information Systems (EIS) which include at least on decision support decision component. Decision support usually means the use of various models and scenarios which help users to assess how their decisions may influence the state of the system. For general overviews on EIS and EDSS, seen from the software point of view, see Swayne et al. (2000) and Lam and Swayne (2001).

From the software point of view, EDSS are very complex systems which require the integration of data management, metadata systems, models, geographical information systems (GIS) and other components like image processing or (in the case of the system presented in this paper) data fusion . This integration can be very costly and concepts are sought to make it easier and more flexible.

1.2 Distributed Sensor Networks

There is growing interest in distributed sensor networks (DSNs) which can make use of recent growth in telemetry infrastructures. Such networks can be used to create a spatially rich data resource, offering greater resolution and robustness than discrete sensors. In a marine context, large numbers of drifting buoys have been deployed in major programmes such as the Global Ocean Observing System (GOOS). Further developments in low-cost sensor technologies (eg biosensors) and durable fixed sensor platforms

will enable proliferation of DSNs to manage marine and coastal waters.

It is recognised that growing numbers of DSNs will create problems of data management. Sensor fusion will become essential, to convert large data volumes into succinct information within a decision support context.

1.3 Phenomenological Models

Sophisticated coastal modelling suites (eg the OSIS and PROTEUS suites) have been subject to intense development over the past 5 years. Present capabilities include modelling of solid-phase (sediment) settling, contaminant phase partitioning (eg between oil & water phases) and toxic reactivity (eg biodegradation and toxic uptake).

i-MARQ will not attempt to extend this capability, but rather will explore how to apply these capabilities within a data fusion paradigm. In essence, the ability to characterise the processes governing and describing key water quality parameters will determine the ability to perform the state estimation required for data fusion. Therefore, the integration of modelling and fusion competencies is an important aspect of i-MARQ, offering the prospect of a substantial advance in capability.

In parallel, work has progressed in the field of proxy parameter applications, in which a phenomenological model is used to link one or more measurable parameters (eg dissolved oxygen, chlorophyll) to a non-measurable parameter (eg algal productivity). Such models are necessarily probabilistic, and a critical aspect of their usefulness is the ability to model the uncertainty within the resulting parameter estimate using fusion techniques.

1.4 Data Fusion Approaches

Data Fusion is the process of combining or integrating data from sets of disparate data sources (including phenomenological models, tables, human reports, etc) or sensors (1-4 D) into a single and more accurate estimate or picture with an associated measure of the belief in the outcome.

The field of data fusion has mushroomed over the past 5 years, building on the early foundations created by Laplace and von Neumann. Many of the major advances in data fusion techniques have occurred in the defence and robotics areas

(Manyika and Durrant-Whyte, 1994, Harris et al., 2002) for problems that frequently associated with tracking problems. There are various approaches to Data Fusion including expert and blackboard systems, intelligent agents, data based learning algorithms such as neuro-fuzzy, through to model based algorithms such as the Kalman filter (Harris et al., 2002). The Kalman filter or information filter has been the most successful method over the past decade including approaches such as the measurement fusion and state fusion. Applications embrace a wide range of needs, from military target tracking (missile and aircraft tracking), vehicle collision avoidance systems, condition monitoring in manufacturing, transport and health care and in the environment such as remote sensing via SAR/satellites and weather prediction.

The common element of these manifestations of data fusion is the merging of multiple data inputs to maximise the quality of information describing particular parameters of user interest. The application of these techniques to sensed data (sensor fusion) is of primary relevance to i-MARQ.

Data fusion can be applied at three levels:

- Low-level fusion to maximise quality of data from each individual sensor module (essentially signal conditioning);

- Medium-level fusion to extract information on specific features of interest using multiple data streams (as applied to satellite images) (essentially feature extraction);

- High-level fusion to create information describing specific behaviours in a decision-making context, embracing both spatial and temporal fusion.

High-level fusion will be applied by i-MARQ to generate wide area datasets of water quality parameters required by users, utilising the data streams described above. Spatial and temporal fusion processes will integrate a wide range of knowledge resources including databases, look-up tables, hydrodynamic modelling algorithms and constraint fields. The 'fuzzy' application of fusion at this level, to data streams having widely different spatial and temporal characteristics, is at the leading edge of data fusion techniques. Here the grand challenge is to find a common data/uncertainty representation for a variety of data sources.

To date, most of the RTD effort on data fusion has been directed towards military (target identification and tracking) or navigation (eg 'fly-by-wire') applications. This work has created a variety of techniques for state estimation and fusion using non-linear process models that are primarily derived from the data souses via on line learning algorithms allowing us to deal with highly nonlinear and time varying data sources. i-MARQ will adapt this expertise to generate GIS-enabled datasets from distributed non-homogeneous environmental sensor networks. This field of advancement targeted by i-MARQ will represent a significant and valuable advance in the state of the art.

2. OVERVIEW OF THE i-MARQ SYSTEM

The system architecture for i-MARQ takes the form of a client-server-client structure. This supports access and consolidation of data from a variety of distributed data providers, processing of this data to add value, and distribution of resulting map-based information to a wide range of users, as illustrated in the system schematic (Fig 1).

When a local partner has new data to offer, it will normally be placed on a local server where it will be identified by meta information system (MIS). The MIS will signpost these data streams for use by other i-MARQ modules.

A key module is the fusion & modelling engine which will generate best-estimate values of output parameters and their associated uncertainty levels.

Fused information will in turn be signposted by the MIS, so it can be accessed by the GIS module. These applications will publish map-based information to users through the internet, and also supply information alerts.

A critical feature of this architecture is its open system approach to data sources and its resilience to data failure. If new data owners wish to make their data available to i-MARQ (ultimately on a commercial basis) they can port this onto an internet-accessible server with meta information according to the i-MARQ protocol. This will then
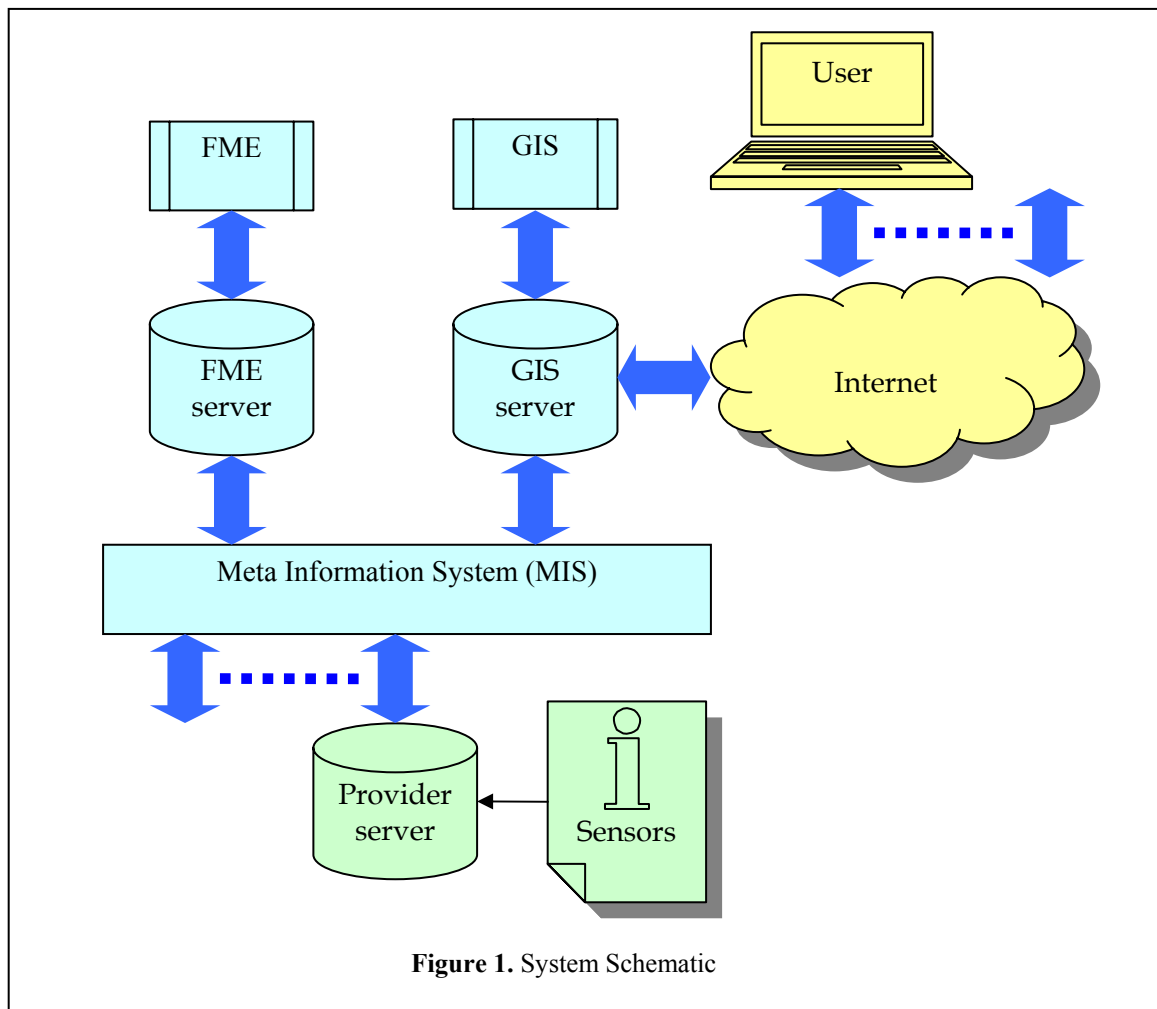

Figure 1. System Schematic

be accessed by the i-MARQ MIS and contribute to the regional information published by i-MARQ.

In addition, the data fusion and modelling processes will allow the system to continue publishing data in the event of loss of access to a particular data source – albeit with declining information quality (increasing uncertainty) in the location suffering loss of data.

3. ARCHITECTURE

Integration of distributed data sources and software (e.g. models, legacy systems) requires a methodology describing the information to be integrated to grant access and availability to a wide variety of users. One way to describe data sources and applications is the usage of meta information managed by a Meta Information System (MIS).

Meta data allows to describe context related and topological characteristics as well as additional information about the syntax of data types and processes. All this is system and implementation independent (Denzer 2002) and can be applied to the i-MARQ system.

The MIS does not directly access data and computational processes, but describes the data and computational processes by metadata.

A well defined set of communication channels is required to enable proper access and flow of information. Figure 2 shows the abstract integration schema if integration is done via the MIS.

The MIS provides an inventory of available components, namely primary data sources ("xyz" in fig. 3 can be any type of primary distributed data source) and computational components. Within the MIS, meta information is available which describes these data sources. The MIS is also aware which computational components are available in the system and meta data about these systems is available as well.

A very important question is how derived data (i.e. output from the computational components) will be treated in the overall system. For a static system with a defined set of parameters, a central store of such data could be feasible. For a system which is able to grow and to change its behaviour, such a central store would quickly degenerate to a large data graveyard which is difficult to administer (in particular as the responsibilities for the different derived data sets are distributed).

For this reason, the i-MARQ architecture requires each computational component to implement it's own data store for result data. There are a number of advantages to this.

First, each component's implementers know best how to store their results in an optimal way (a lot of the data has very special formats in the first place).

Secondly, the component can easily generate meta data for results (e.g. how and when the results were generated from which input and which quality). Thirdly, each component remains a functioning component even without i-MARQ and could therefore also been used in a different context than i-MARQ.

The MIS never accesses the data stores directly. This is done via the communications channels. There is a clear separation between the meta information system and the data access. The MIS
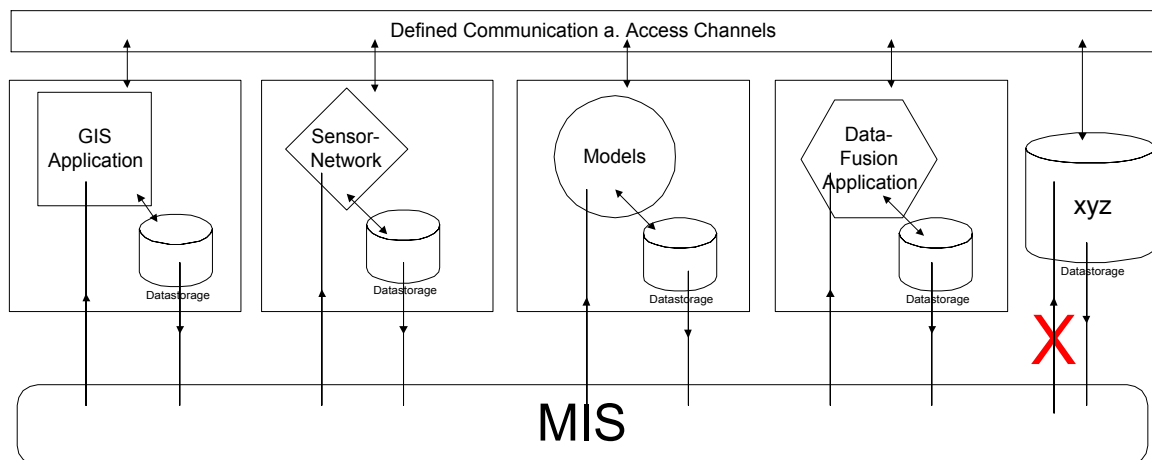

Figure 2. Architecture Diagram

has the knowledge about data sources, components and data stores, as well as the access mechanisms for these components and data.

Each component retrieves this information from the MIS. Access is then done directly through the data access channels.

4. CLIENT COMMUNICATION

Clients will access i-MARQ through both mechanisms: the MIS and the data access channels mentioned above. Fig. 3 distinguishes clients from the other components. The use of the MIS allows for generative programming: a client first accesses the MIS to explore the possibilities and then uses the MIS to retrieve the access mechanisms to data and components.

A client will therefore get information about available data sources and applications from the MIS. Access channels will enable the client component to obtain data from applications and data sources.

This scheme allows for rather intelligent clients, because the business logic of data and meta data management and access are not hard-wired.

Therefore clients can be implemented in a very flexible way because they can react on new data and services posted to the system at run-time.

Implementation of such flexible clients is clearly more difficult and uses more resources. The payback of the concept will come when the system needs to be extended or the characteristics of

components need to be changed. Such changes will carry a much lower burden on developers, as the general concept has allowed for them from the start.

5. CONCLUSIONS

i-MARQ is work in progress. At this time, the project has carried out one third of its work plan (12 of 36 months) and the main architectural decisions have been taken. In parallel, the pilot regions are defined and the science for the data fusion is developed. The first designs of the GIS front-end are under development. This paper describes the architectural approach taken for the overall system.

The implementation of the system will take another 12 months and extensive piloting is planned in the third project year.

6. ACKNOWLEDGMENTS

7. REFERENCES

Denzer R., *Generic Integration in Environmental Information and Decision Support Systems,* Proceedings iMESs conference 2002

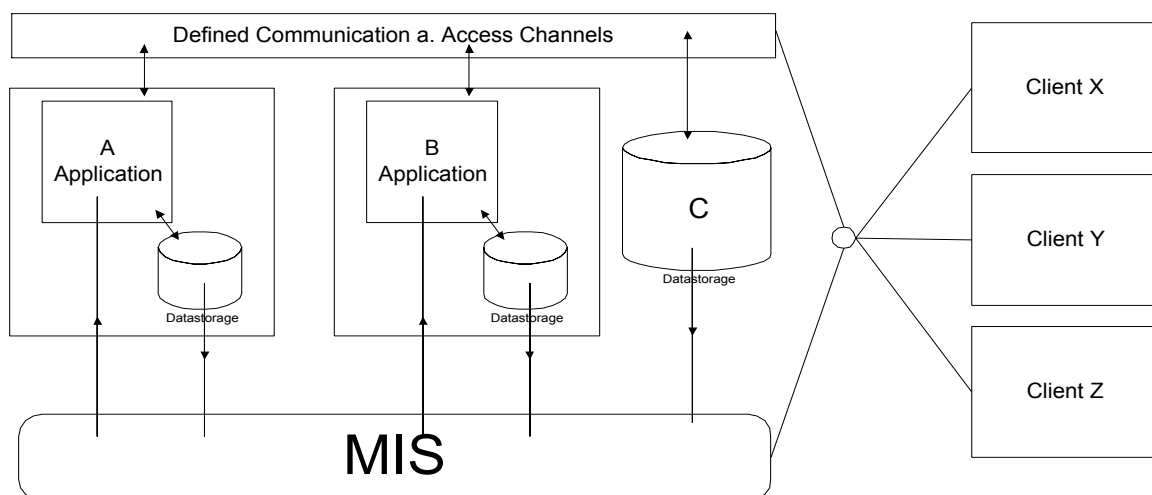Harris, C.J. Hong, X. and Gan Q., Adapative

Figure 3. Collaboration between Clients and Services

Modelling, Estimation and Fusion from Data, Springer Verlag, Berlin, 2002

Lam D., Swayne D. A., 2001. Issues of EIS software design: some lessons learnt in the past decade. *Environmental Modelling and Software,* 16 (5), 419-425, 2001

Manyika J., Durrant-Whyte H., Data Fusion and Sensor Management, Ellis, Horwood, NY, 1994

Swayne, D.A., Denzer R., et al., Environmental Decision Support Systems – Exactly what are they? *Environmental Software Systems Vol. 3 – Environmental Information and Decision Support*, pp. 257-268, Kluwer Academic Publishers, 2000