

Fusing Scientific Modelling with Knowledge Management Systems

S. Kininmonth^{a,c}, T. Donovan^b and S. Edgar^a

- a. Australian Institute of Marine Science, Townsville,, Queensland Australia
- b. CRC Reef, Townsville, Queensland, Australia
- c. Corresponding author Ph. +61-7-4753443 s.kininmonth@aims.gov.au

Abstract: The influence of scientific problem solving on policy makers and the general public is a direct function of the perceived relevance and clarity of communication output. However environmental modelling is complex and this diminishes the ability of traditional media to clearly represent model output. Simplified brochures and news articles struggle to convey the complexity of the network of abiotic and biotic interdependent systems. Conversely detailed reports and scientific papers are rarely read. The optimum delivery of science involves the viewer dynamically determining the level of detail for a topic of interest. Static media such as newspapers and non-digital television have no provision for information expansion. This paper presents a knowledge management model focusing on integrating GIS into dynamic web pages to deliver a scalable information product. The web site called www.ReefFutures.org uses dynamically generated web pages (Coldfusion) to incorporate GIS (ArcIMS). Using textual search engines (Verity) within a highly graphical and dynamic website permits the viewer to fully explore a topic of interest. The interactive mapping technologies permit the viewer to zoom in and fully investigate the scientific complexities surrounding an environmental issue. For many viewers this will be the first time to examine the raw data used by scientists. The first issue being addressed is coral bleaching on the Great Barrier Reef. Spatial data depicting recent coral surveys can be viewed in the full context of satellite derived sea surface temperatures. Hyperlinking offers a mechanism of connecting images of the corals from surveys to the map elements with metadata describing various modelling techniques used. The ability to deliver complex modelling results through a sophisticated communication delivery mechanism will significantly enhance the influence of science.

Keywords: *science communication, knowledge management; dynamic web page, webmap, coral bleaching*

1. INTRODUCTION

The transition of knowledge generated by scientific modelling to knowledge consumed by society is often ineffective and cumbersome. Yet most scientists are driven by a passion to influence the society in which they live. This passion is particularly prominent for natural resource scientists who study threaten environments. Ironically the field of environmental modelling is often hidden from environment-dependant society due to research complexity and ambiguity.

To influence the policy makers and the general public requires a high degree of perceived relevance and clarity of communication output (Freyfogle and Newton 2002). Traditional media, with concise and targeted articles, struggle to convey the complexity of the intertwined network of abiotic and biotic interdependent variables. Formal scientific publications such as papers and

reports serve to provide credence and relevance to the scientific community but are rarely read by the wider society. The optimum delivery of science involves the viewer dynamically determining the level of detail for a topic of interest. Static media such as newspapers and non-digital television have minimal provision for information expansion. Complex issues, such as coral bleaching, are represented as simplistic articles like “Bleaching has reef in hot water” (The Australian, 20th August 2001, page 7).

Knowledge Management Systems (KMS) offer the potential to convey complex research results to a wider audience. The definition of KMS is broad and confused (Hlupic et al. 2002) but can be defined as “technologies which enhance and enable knowledge generation, codification and transfer”(Hlupic et al. 2002) p.91). Knowledge and information are often confused and this directly relates to the complicated and non-linear creation of knowledge from information (Styhre 2002). Information is data that creates a change in

the receiver's knowledge and then becomes obsolete or non-information (Styhre 2002). As Styhre (2002 p.230) states the repetition of non-informative data, such as advertising, that makes no difference "may even be frustrating or somewhat annoying". Knowledge, however, is not the accumulation of information but rather the intelligent use of information (Hlupic et al. 2002, Styhre 2002). To complicate matters further, only knowledge relevant to a specific decision making process is desirable, and designing filters is a fundamental problem for KMS (Fink 2002). To be effective, KMS requires the strategic implementation of knowledge management tools. Hlupic et al (2002, p.95) outline three categories of tools that can be utilised within a KMS:

- "Knowledge Generation – the creation of new ideas, recognition of new patterns, the synthesis of separate disciplines, and the development of new processes.
- Knowledge Codification – the auditing and categorisation of knowledge.
- Knowledge Transfer – the forwarding of knowledge between individuals, departments and organisations."

For many scientists the tools of knowledge generation (ie statistical modelling, data mining) have been the focus at the expense of knowledge transfer tools. Work programs at research institutes often restrict human resource capacity to scientists with knowledge generation skills with scientific papers being the preferred and limited knowledge transfer tool. To address a need for wider communication, libraries now focus beyond codification to metadata management with search engines and online retrieval systems that facilitate knowledge asset management (Williamson and Liopoulos 2001). Digesting formalised knowledge is often tedious and science communication staff are used to assist re-engineering of scientific papers into popular media. Using journalistic techniques these staff are able to extract clarity and relevance from research modelling. The World Wide Web provides a rapid and convenient means of knowledge dissemination (Boechler 2001). To display and arrange condensed articles on the web requires web site managers. In many cases this system of scientists, librarians, communicators and web site managers is fully functional and satisfies the clients needs. For example the Long Term Monitoring Program at Australian Institute of Marine Science has developed a web site (<http://www.aims.gov.au/pages/research/reef-monitoring/reef-monitoring-index.html>) that facilitates rapid transfer of monitoring information with derived knowledge concerning

the ecological status for the Great Barrier Reef (Sweatman et al. 2001). New technologies in web site design and function have created a new array of transfer tools (Fischer 2001). In particular the use of interactive web-based mapping (*webmaps*) has added a spatial dimension to the textual and static picture displays of the past.

The spatial dimension is the prime focus of Geographic Information Systems (GIS) and has been the domain of specialists since inception (Talen 2000). With the design of internet mapping servers the web viewer has the ability to directly request map images composed in real time from stored data. This data can be the actual data used by the scientists in their modelling systems. Although presently limited in analytical functionality the spatial servers do offer query and display functionality. The simple ability to view collateral datasets for an area of interest at a suitable scale is particularly powerful. Interactive maps by themselves do not have sufficient context to be valuable in knowledge transfer (Boechler 2001). Dynamic web pages with search capability interlinked to the webmaps provide knowledge transfer capacity.

The principal technological development has been the design of web application servers. Contrary to web servers, the application servers are able to interact with databases, deliver customised information on user preferences and validate user actions (Macromedia 2002).

This paper describes the KMS at the Australian Institute of Marine Science with particular emphasis on the knowledge transfer tools. The first section will outline the knowledge generation and codification tools across the institute while the second section will focus on delivering a knowledge transfer tool that is targeted towards disseminating coral bleaching research.

2. SYSTEM ARCHITECTURE

The Australian Institute of Marine Science (AIMS) "was established by the Australian Commonwealth government in 1972 to generate the knowledge needed for the sustainable use and protection of the marine environment through innovative, world-class scientific and technological research" (www.aims.gov.au/pages/about.html). Thirty years of accumulated knowledge reside within the institute and this requires management to optimise future knowledge creation. Issues of integration across research groups combined with data quality and accessibility have highlighted the need for a more sophisticated system (Kininmonth 2002). A KMS is under construction and consists of three integrated elements (figure

1); an Enterprise Geographic Information System (EGIS), a data centre (ADC) and knowledge transfer servers (Web servers).

2.1 Implementation

The EGIS is a client focused GIS that utilises the functionality delivered by ESRI's suite of software utilities. Figure 2 provides a diagrammatic overview of the system within AIMS. Pivotal to the success of EGIS is the centralisation of quality assured spatial data combined with standardised metadata and multiple access opportunities (Kininmonth 2002). Spatial data within AIMS is stored in a multi-user geodatabase using ESRI's Spatial Data Engine (SDE). Strict data-naming conventions are employed to keep the data layers organised, while a system of user privileges allows fine-grained control of clients accessing the data. Security considerations aside, the geodatabase model has further advantages in speed of data access (Zeiler 1999). Large-scale images are stored as seamless raster datasets, and data retrievals are automatically constrained to the client's map extent. Image pyramids - a pre-calculated resampling of the images - are used to further improve drawing performance. Vector data is also extremely fast, as the geometry is stored internally, and an internal system of grid-based indexing tables are used in spatial queries.

Mapping Service (ArcIMS) which provides access through a common web browser. Importantly the metadata that describes the spatial data is also served via a web interface powered by ArcIMS. Researchers can blend and manipulate data to suit their modelling requirements. The EGIS is especially critical at AIMS where data storage exceeds several terabytes and data mining techniques are required (Koperski et al. in press). This system will greatly aid the knowledge generation tools currently being utilised (see Wooldridge and Done, this edition).

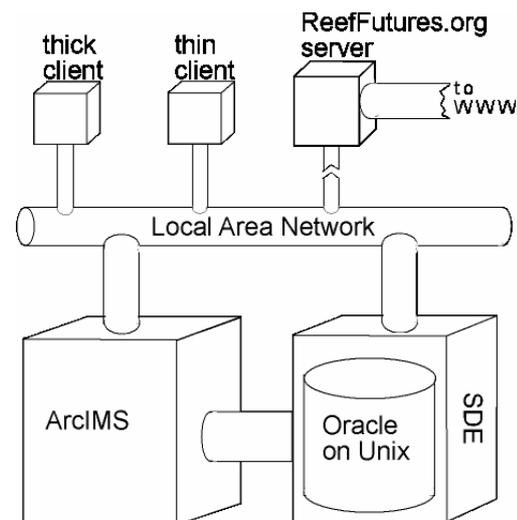


Figure 2. The structure of EGIS

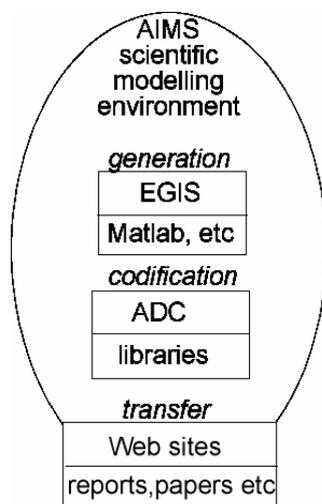


Figure 1. The Knowledge Management System at AIMS with the generation and codification systems embedded within the scientific modelling environment. The transfer systems create a link between this environment and the wider world.

Access is through a number of interfaces including the ArcInfo software (thick client), webmaps and customised Java application programming interfaces (thin clients). The webmaps are serviced by the Arc Internet

The AIMS data centre (ADC) uses a warehouse approach and is an effective system for collecting, extracting, transforming and cleaning organisational data (Chiu 2003). The ADC will facilitate the integration of scientific, financial, human resource and corporate datasets. The EGIS will provide spatial warehouse tools to complement the ADC. An Oracle relational database is at the hub of the AIMS data centre. The ability to use a commercial relational database management system (DBMS) for all the data storage requirements is particularly important in an enterprise system (Zeiler 1999). The DBMS is a repository for all data in the system, including metadata. Set up in this way, there is less overhead involved in maintenance and updating of AIMS datasets, due to the centralised storage. Any changes to the underlying datasets are done once, in the DBMS, and the updated data is immediately available to all clients. Backups of the data are likewise simplified. The GIS software used integrates extremely closely with the DBMS, and is able to automatically update spatial metadata as the spatial data changes. Multiversioning of datasets is also possible, where different users can access different versions of the same data. This can be

easily implemented using the current system, although this has not been necessary at this stage.

To publish the results of the scientific modelling in a manner that satisfies the viewers interests requires dedicated software tools. At AIMS several dedicated servers were built expressly for this purpose. Their configuration is quite complex with multiple application servers interacting as requests are received from viewers (figure 3). Initially the request is interpreted by the Apache web server. If the script contains Coldfusion elements then the Coldfusion server interprets the code otherwise Tomcat Servlet Engine interprets the code. Coldfusion by Macromedia Pty Ltd is a powerful internet server that provides rapid deployment of interactive web sites. Both ColdFusion and Tomcat can send requests to the ArcIMS application server for specific geospatial information such as map graphics and data arrays.

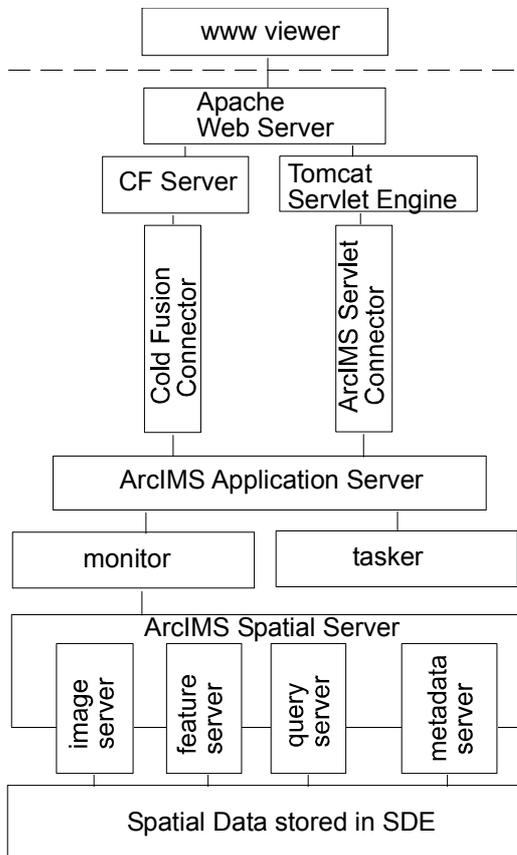


Figure 3. Architecture of the AIMS spatially enabled web sites

This web site can operate without Coldfusion instead using HTML and Javascripts to provide functionality. The internal AIMS web site and the external Reef Futures website contain maps that are constructed in this manner, however the

loading time for the Javascripts and HTML code is considerable and the ability to integrate with other pages is limited. Coldfusion servers offer a fast and rapidly customizable interface that is engineered to interact with databases and multiple scripting languages. The page content is stored in a SQL Server 2000 (Microsoft) database and is composed dynamically based on the user's requests. For the webmaps the Coldfusion code has been engineered to provide simple spatial tools so the viewer can navigate and interrogate the information on display (figure 4). Of particular note is the Verity search engine that is able to search and index multiple documents to create a collection. Text searches can include the entire document and not just the metadata.

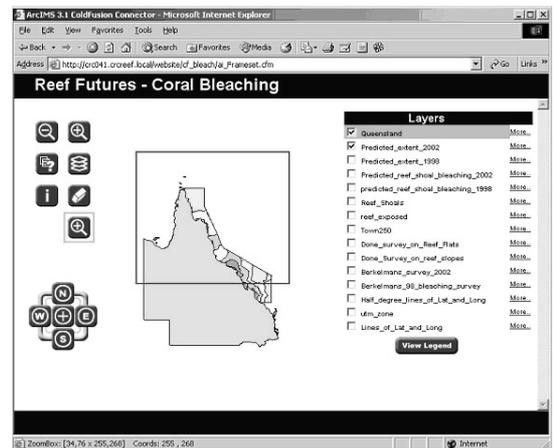
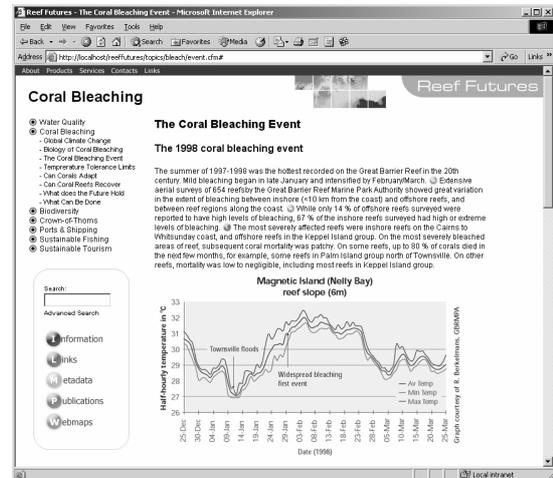


Figure 4. Screen capture of Web pages. The top capture shows the regular interface with a toolbox and navigation path on the left. The central text contains buttons that provide access to additional information such as other websites, pop up information panels and webmaps. The webmap below shows the simple tools and map layer management with links to additional metadata information.

The application server can deliver images, features, query and metadata requests to the spatial server. The spatial server then directly accesses the data to satisfy the requests. This ensures requests are always operating on the current data. Extended functionality can be built into the viewing applications through the use of the Arc Extensible Markup Language (XML). XML is a flexible consistent server-side language that focuses on information transfer. ArcXML can be sent directly to the application server with a sophisticated set of instructions largely determined by the viewer. Future developments in providing the viewer the ability to manipulate the modelling inputs (ie the temperature of oceans in the year 2050) will be through this mechanism.

2.2 User Feedback

With a website that is composed of many active elements the issue of cognitive overhead and disorientation demands attention (Boechler 2001). Cognitive overhead is defined as the “amount of cognitive resources necessary to successfully complete an informational task in hypertext” (Boechler 2001, p.27). Websites that hinder the viewers ability to plan routes through the web page, assist with understanding the contents and coordinate informational tasks will significantly degrade their effectiveness.

Hypertext disorientation is the feeling of being lost with in the structure of the website. Symptoms of this disorientation include looping, inefficient navigation, generation of query failures, disorganised screen layouts with multiple concurrent windows, and excessive back-tracking (Boechler 2001). To avoid cognitive overhead and disorientation the Reef Futures websites will have clear navigational pathways with minimal ‘clutter’.

The final website product satisfies the initial requests we received from leading scientists and management authorities. The Collaborative Centre for Research on the Reef (CRC Reef) was able to ensure stakeholders provided input to the website design. The singular most common request was to package the available information into a single site with sophisticated search facilities. Figure 5 shows a simplified version of a suggested structure with clear demarcation of function. Given the capacity of hypertext we were able to obscure the boundaries between the functions so that all four options can be accessed at any point within the website. Once a topic is selected viewers are shown a front page with an introductory statement which can be read like a brochure (option 1, figure 5). Also included are the facilities to search the metadata and reports

and then directly read those documents (option 2 & 3, figure 5). From within the brochure pages and from the side menu ‘toolbox’ viewers can request the interactive maps (option 4, figure 5).

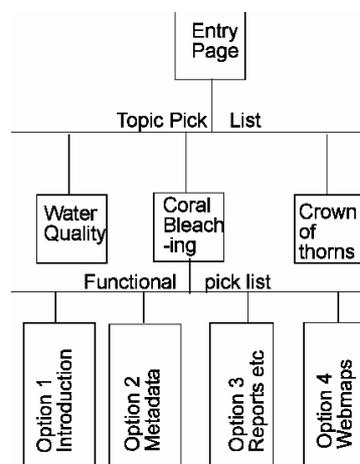


Figure 5. A simplified version of a request for website functionality.

3. CONCLUSION AND RECOMMENDATION

The capacity to delve into the scientific world is greatly enhanced through the use of the AIMS knowledge management system. The visual appeal that greets the viewer as they navigate through the multi-tiered WebPages enhances the overall effectiveness. The simplistic nature of the software tools ensures the wider public has complete access with minimal training. All elements of this KMS are growing rapidly at present and the final configuration will be a single system with a comprehensive array of functions. Through this interface scientists and the general public will be able to maximize their investigations over a wide range of topics.

While the system works well, it is dependent upon a number of complex interacting sub-systems, including web servers, Unix and Windows Operating Systems, high-speed networks, GIS software and the DBMS. Initially there is a steep learning curve, and good communication is required between the Network, GIS and DBMS administrators, particularly as different people from different departments often perform these roles. Maintaining software compatibility as components are upgraded is a continual area of frustration. For example recent changes from Coldfusion 5 to Coldfusion MX have required installation of an upgraded ArcIMS 4.01. However the Apache1.3 web server software did not fully support this upgrade and additional configuration was required.

There are some aspects to the existing system that could be improved. Currently the software being used imposes a publishing process on the metadata before it can be searched using the ArcIMS Metadata Explorer Service. The original metadata is integrated with the datasets, and is automatically updated as the datasets change (eg. spatial extent, projection, etc). To enable web-searches, the publishing process stores a copy of the metadata in the DBMS. This duplication of metadata is less than ideal, as metadata is not searchable in a web-based format until it has been published, and the published metadata must be re-published when the original datasets change.

The benefits for the scientific modelling community are substantial as integration of datasets mirrors the team-based research methodology. This research cohesion can generate substantial growth in the flow of knowledge both within the institute and to the wider community. Management agencies that wish to engage with scientists can begin to appreciate the complexities of the issues while scientists can ensure alignment with management priorities through feedback mechanisms. The wider community can begin to comprehend the reasons for scientific debate and uncertainty without the confusion surrounding program management and funding priorities. The influence of science should then be significantly enhanced as local management priorities are addressed within a regional environmental framework.

Future research should address the ability to engage viewers with interactive modelling tools and multi-media presentations such as video interviews of key scientists. In situ web cameras could also be strategically employed (ie underwater at a bleached site) to add a temporal dimension to the KMS. Examination of the web server log files will provide rapid feedback on the tools and pages that viewers utilised. Web site development priorities should be based on these examinations combined with feedback comments. The impact that the WWW will have on scientific KMS is only just being acknowledged and the potential to expand its functionality is immeasurable.

5. ACKNOWLEDGEMENTS

The authors would like to thank the IT and Data Centre team lead by Scott Bainbridge for technical support; Terry Done, Vicki Harriott and Louise Goggin for direction; and two anonymous reviewers for constructive comments.

6. REFERENCES

- Boechler, P. M. How Spatial Is Hyperspace? Interacting with Hypertext Documents: Cognitive Processes and Concepts. *Cyber Psychology & Behaviour* 4:23-46. 2001
- Fink, P. Improving knowledge management systems: issues and reviews of recent experiences. OECD. 2002
- Fischer, M. M. Innovation, knowledge creation and systems of innovation. *The Annals of Regional Science* 35:199-216. 2001
- Freyfogle, E. T., and J. L. Newton. Putting science in its place. *Conservation Biology* 16:863-873. 2002
- Hlupic, V., A. Pouloudi, and G. Rzevski. Towards an Integrated Approach to Knowledge Management: 'Hard', 'Soft' and 'Abstract' Issues. *Knowledge and Process Management* 9:90-102. 2002
- Kininmonth, S. J. GIS: The key to research integration at the Australian Institute of Marine Science. Pages 67-72 in J. Breman, editor. *Marine Geography; GIS for the oceans and seas*. ESRI Press, Redlands. 2002
- Koperski, K., J. Han, and J. Adhikary. Mining Knowledge in Geographic Data. *Communications of the ACM*. in press
- Macromedia. Macromedia Coldfusion MX: *Getting started building ColdFusion MX applications*. Macromedia, San Francisco. 2002
- Styhre, A. The Knowledge-intensive Company and the Economy of Sharing: Rethinking Utility and Knowledge Management. *Knowledge and Process Management* 9:228-236. 2002
- Sweatman, H., A. Cheal, G. Coleman, S. Delean, B. Fitzpatrick, I. Miller, R. Ninio, K. Osborne, C. Page, and A. Thompson. Long-term monitoring of the Great Barrier Reef. 5, Australian Institute of Marine Science, Townsville. 2001
- Talen, E. Bottom-up GIS. *Journal of the American Planning Association* 66:279-293. 2000
- Williamson, A., and C. Liopoulos. The learning organisation information system (LOIS): looking for the next generation. *Info Systems Journal* 11:23-41. 2001
- Zeiler, M. *Modelling our world*. ESRI Press, Redlands, California. 1999