

# Integration Model for Multiple Types of Spatial and Non Spatial Databases

Mustafa Man<sup>1</sup>, Mohd Shafry Mohd Rahim<sup>2</sup>, Mohammad Zaidi Zakaria<sup>3</sup>,  
and Wan Aezwani Wan Abu Bakar<sup>4</sup>

<sup>1</sup> Department of Computer Science, Faculty of Science & Technology, Universiti Malaysia Terengganu (UMT), 21030 Kuala Terengganu, Terengganu, Malaysia  
mustafaman@umt.edu.my

<sup>2</sup> Department of Computer Graphics & Multimedia, Faculty of Computer Science & Information Systems, Universiti Teknologi Malaysia (UTM), 81310 UTM Skudai, Johor Bahru, Malaysia  
shafry@utm.my

<sup>3</sup> Department of Fisheries Science, Faculty of Agrotechnology & Food Science, Universiti Malaysia Terengganu (UMT), 21030 Kuala Terengganu, Terengganu, Malaysia  
zaidi@umt.edu.my

<sup>4</sup> Pre Sea Marine Engineering Department, Training & Education, Malaysian Maritime Academy (ALAM), 21030 Kuala Terengganu, Terengganu  
aezwani@alam.edu.my

**Abstract.** Integration process of a various information in various database types requires a thorough understanding to carry out data extraction process in terms of its scheme and the structure. Due to this, a new model should be developed to resolve the integration process of this heterogeneous information in various database types and in various scattered and distributed locations. SIDIM is a model which covered processes such as pre-integration, scheme comparison, algorithm and intermediary software (middleware) development process and as well as post-integration. Emphasis are administered in algorithm development by using hybrid approach based on CLARANS approach's combination, abstract visualization and Catch Per Unit Effort (CPUE) to enable to achieve the required processed data or information in a quick, trusted and reliable manner. SIDIM will become a new engine to process information in various database types without changing any of the existing (legacy) organization system. To verify this model credibility, the case study related to fishing industry in Malaysia and artificial reef project are being made as a foundation for SIDIM efficiency testing.

**Keywords:** Spatial Information Databases Integration Model (SIDIM), Integration, Database, CLARANS Model, Abstract Visualization, Catch per Unit effort (CPUE).

## 1 Introduction

Recently, many research on information-seeking were more diverted on merging and unifying the spatial and non-spatial databases. The method of information searching is very much crucial among consumer. The consumer may requires information on some

location and the updates on the change of that location are the major concern besides continuing to make a comparison through a visualization method which is based on map. The combination of both spatial and non spatial data is indeed important and crucial [1]. Database technology is refers to wide range of approving technique for individual, industries or even government for daily operation. About 200 to 300 research papers were presented about similar issues in seminar, conference or even journal. Database technology and Database Management System are always in change. Online access becomes popular among users because it is faster and quicker. And today, data models for spatial and temporal data are most in need [2]. A broad spectrum of data is available through Web in distinct heterogeneous sources, stored under different formats and different type of databases and at different location or server: a specific database vendor format, length of data, image format, Step (CAD/CAM data), etc. Their integration is a very active field of research and development. To enable a specific tool to manipulate data coming from various sources especially different types of database format must be screened through a translation phase for example : the data (in the source format) needs to be mapped to the format expected by the application [3]. There are several challenges that need to be faced in the process of merging and integrating the database from different sources and locations. It is required that these data, i.e., data from different sources to be in priority and can be selected prior to continuing the integration process. The objectives of our research;

1. How to produce one model to enable integration process various information data spatial and non spatial from various data base that the type different?
2. How the model that proposed afford to carry integration process various information data spatial and non spatial which differ the data base structure and in water environment which differ for environment on line?
3. How the proposed model can improve in terms of integration time?

Spatial Information Database Integration Model (SIDIM) is one of our prototypes which can integrate spatial and non-spatial data from various databases format and location. In principle, our model in this study is based on location coordinate technique in order to merge the predicted results based on location. The integration approach from various databases were used to give an impact analysis of information especially among spatial and non spatial data. The algorithm is developed using location integration technique in order to fasten the extraction process of spatial and non spatial data.

## 2 Related Work

Research on data integration is one of the most important element and is a '*hot issues*' in the research of spatial data. Previous researches were more concentrated on developing a common form of information that is easier for the user to manage the database without consuming longer time. Integration is defined as a merger of various information from various sources that can provide benefits from the aspects of

collection of the best information, less in processing time, resource conservation and data-sharing to various purposes. Information which will be combined or integrated must be staying in a state of analogous (uniform), but if the state of the information which will be consolidated is unequal, thus a form or modification process has to be executed on that particular information structure without damaging the original information. To place the information in various databases and at different or scattered location may contribute to difficulty in performing integration process. Thus a research to integrate 'every single piece into one' [5] is to be initiated as a kickoff start. Earlier integration method is called MULTIDATABASE in 1980. It is then followed by a mediator called as GARLIC. Later on, INFOSLEUTH is introduced. Then integration based on ontology which known as OBSORVER took place. After that, integration is based on peer to peer known as HYPERION. The journey lasts with web based integration method which is known as ACTIVE XML [6]. The weaknesses of various databases in a common and typical organization can be prescribed as follows; High potential of incompleteness, inaccuracy of processed data, no control and monitoring of repeated data, conflict in task distribution to update data and inaccuracy of achieved data.

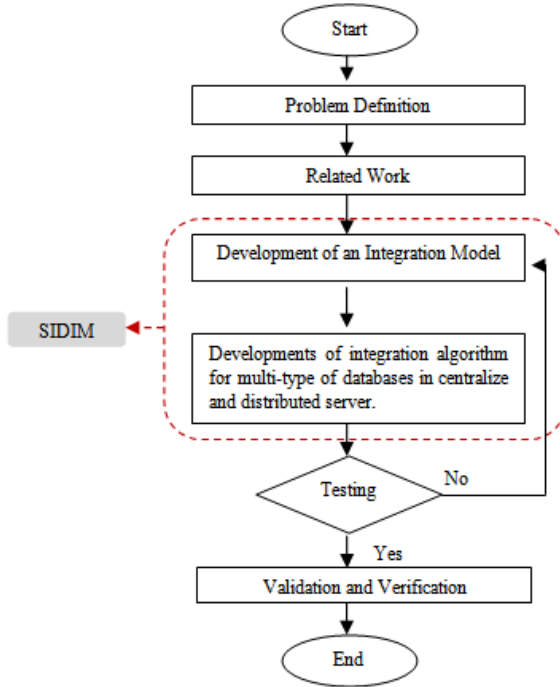
*Interoperability* is the key solution to solve the issue in integration. It allows heterogeneous data to communicate or interact from one application to another application by replying with the intended information. Interoperability is a complex feature because of the data source obtained is not in database format i.e. in email form or file attached only and external data source obtained through internet access in which the data is either unstructured or semi-structured which needs greater understanding in its semantic [7].

### 3 Research Methodology

SIDIM is an integrated spatial and non- spatial database designed to enable access and efficiency evaluation on artificial reefs based on the growth of phytoplankton and zooplankton for each dedicated artificial reefs area. The SIDIM is considered as a method, model or new idea to ensure the effectiveness and efficiency of artificial reefs development project as a place for fish population and marine park. Since the high cost incurred to evaluate the development of artificial reef by hiring a special scuba task force, then the arrival of SIDIM is fit for the time. The SIDIM enables the integration of more than one database either within a similar or a different schemes. The research methodology consists of systematic process flow which is illustrated in figure 1. A few processes which have been subdivided that need to be taken prior to integration as in the 4 levels as depicted in figure 2 as follow:

- i. Pre-integration: Process to assess database environment are used like Oracle example, MYSQL, ORACLE, and MS ACESSS and from other database.
- ii. Scheme Comparison: Scheme comparison or structure for each this database is needed to facilitate integration process conducted. In the early stage, this process is made manually.

- iii. Intermediary software development (middleware): an application shall be developed for integration process the data base workable.
- iv. Post-Integration: Integration process assessment was being conducted from credibility process aspect and “interoperability”.



**Fig. 1.** Research Methodology Flow Chart for Integration Process of Multi Database Types

The focus of the research is on the algorithm development which enables multi database types in a distributed environment can be integrated into another database through online based on three-tier architecture.

**A. Pre-integration Processes → 1<sup>st</sup> step**

Like those mentioned previously, process of pre integration requires a study on present database environment. The researcher has already developed an application called ARPOS, funded under eSciencesFund (MOSTI) research grant in 2006 [8]. ARPOS database is in MYSQL and WiFISH database in MSSQL based. The system development uses .php software which is accessible through wired internet and wireless internet. Both of the databases are owned by different agencies namely Department of Fisheries (DOF) and LembagaKemajuanIkan Malaysia (LKIM) and were located in different location.

B. Scheme Comparison → 2<sup>nd</sup> step

The evaluation results of the two databases are found that the location (position) of artificial reef can be equal with the location of fish catches conducted. Due to this, the location based technique is a core and as a fundamental to determine the effectiveness development level of the artificial reef project development at a certain location and also within the timeline. The evaluation will be made to test for efficiency level and reliability of developed middleware in integrating process. The evaluation is carried out at the system user level. The process time is recorded to obtain and justify the information processing speed and interoperability is fulfilled. Other than processing time factor, evaluation of Catch Per Unit Effort or Ability Rate (CPUE) is also encountered in order to assess the whole project’s efficiency for each dedicated artificial reef location. CPUE is defined as the level and efficiency degree between the catch in tan (t) and the ability of catch (frequency of fish catch) is being conducted [9],[10]. This method is based on variance and CPUE average CPUE. CPUE with non zero catch is modelled in the following formula:

$$\ln(\text{CPUE})_{i,j,k,l} = \mu + \alpha_i + \beta_j + \gamma_k + t_l + \varepsilon_{i,j,k,l} \tag{1}$$

where  $\mu$  is the variable,  $\alpha_i$  is the year factor,  $\beta_j$  is the month factor,  $\gamma_k$  is the zone factor,  $t_l$  is the rate factor, and  $\varepsilon_{i,j,k,l}$  is the random error component. Then the yearly value of CPUE is calculated in tone / total travelling trip that relates to statement below:

$$\ln\left(\frac{P}{1-P}\right)_{i,j,k,l} = \mu' + \alpha'_i + \beta'_j + \gamma'_k + t'_l + \varepsilon'_{i,j,k,l} \tag{2}$$

Besides, if ‘p’ is defined as efficiency evaluator, then the yearly value of CPUE will relate to the calculation as follows:

$$\text{CPUE}_i = \exp(\mu + \alpha_i + 0,5\sigma^2) \exp(\mu' + \alpha'_i + 0,5\sigma'^2) \tag{3}$$

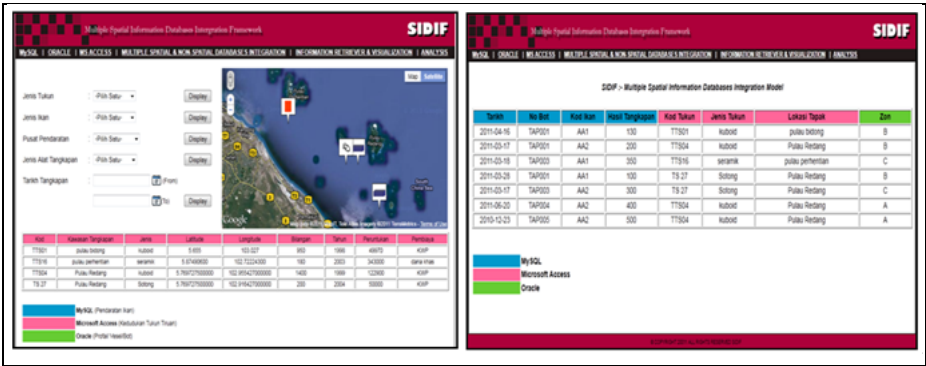
Assumption is made where an increase in CPUE trips with the catches can be influenced by the changes of boat’s efficiency and the catch resources which represents either the fish catch is a success or otherwise. The result of integration process can determine the fish types with the catch area.

## 4 Testing and Evaluation

Evaluation and testing process for SIDIM model is conducted by comparing between manual data with a SIDIM processed data. The testing focused on the ‘interoperability’ feature in spatial and non-spatial data processing among the different types of database within a centralized server and compared with spatial and non-spatial data processing in a different types of database within a distributed server. The result obtained (as shown in Table 1) is compared either the two different servers (centralized or distributed) having a similar results or on the other way around. The comparison done will confirm whether each algorithm deployed at each particular steps has achieved the objectives or otherwise. Figure 2 shown the user interfaces of the system.

**Table 1.** An interoperability testing between different types of database in a centralized environment

Database Types	Types of Web Server	Interoperability Testing with Multi-types of Databases		
		MYSQL	ORACLE	MS Access
MYSQL	Apache	Yes	Yes	Yes
ORACLE	IIS	Yes	Yes	Yes
MS Access	IIS	Yes	Yes	Yes



**Fig. 2.** System User Interfaces

## 5 Conclusions

SIDIM is a model design which is reusable in various types of database integration problem. The location based integration technique (longitude and latitude data) for artificial reef determination is very well conducted and can attain a satisfiable result. First they support data integration by allowing the resolution of different data representation conflicts. In addition, they permit the representation and use of query capabilities that may not be provided by any or some of the integrated GIS systems. As a result, we also tackle the problem of tools integration and extension. We hope that this tool could be extended to other domain as a solution for multiple types of database integration.

## References

- Ougouti, N.S., Belbachir, H., Amghar, Y., Benharkat, N.A.: Integration of Heterogeneous Data Sources. *Journal of Applied Sciences* 10(29), 2923–2928 (2010)
- Elmasri, R., Navathe, S.: *Fundamental of Database System*, 5th edn. Addison Wesley (2006) ISBN: 0321369572
- Abiteboul, S., Cluety, S., Milo, T., Mogilevskyz, P., Siméony, J., Zoharz, S.S.: *Tools for Data Translation and Integration*. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering* (1999)

4. Parent, C., Spaccapietra, S.: Issues and approaches of database integration. *CACM* 41(5), 166–178 (1998)
5. Ismail, W., Joseph, K.S.: Future IT trends for GIS/Spatial Information Management. *Scientific Research and Essay* 5(10), 1025–1032 (2010)
6. Patrick, Z., Klaus, R.: Three decades of data integration-All problems solved? In: *IFIP Congress Topical Sessions 2004*, pp. 3–12 (2004)
7. Kim, J., Peng, Y., Ivezic, N., Shin, J.: An Optimization Approach for Semantic-based XML Schema Matching. *International Journal of Trade, Economics and Finance* 2(1), 2010-023X (2011)
8. Mustafa, M., Yazid, M.S.M., Maizura, M.N., Khalid, S.: GIS spatial data visualization tools for artificial reefs distribution. In: *The 3rd International Conference on Mathematics and Statistics (ICoMS-3) Proceeding 2008*, pp. 25–38 (2008)
9. Raymond, T.N., Jiawei, H.: CLARANS: A Method for clustering objects for Spatial Data Mining. *IEEE Transactions on Knowledge and Data Engineering* 14(5), 1003–1015 (2002)
10. Stephans, A., MacCall, A.: A multispecies approach to sub setting logbook data for purposes of estimating CPUE. *Journal of Fisheries Research*, 299–310 (2004)