International Journal of Advances in Computer Science and its Applications – IJCSIA
Volume 4 : Issue 1 *[ISSN 2250 – 3765]*
Publication Date : 09 January 2014

Designing Spatial Information Databases Integration Modelling using Z Spec.

Mustafa Man
Department of Computer Science, Faculty of Science & Technology,
Universiti Malaysia Terengganu (UMT), 21030 Kuala Terengganu.
mustafaman@umt.edu.my

Abstract—**Theoretically Formal Specification (FS) approach can reduce the overall software development time. FS can be used to provide an unambiguous and precise supplement to natural language descriptions. Focus to this point, it can be rigorously validated and verified leading to the early detection of specification errors or debugs during software development process. Correcting errors at the early development stage is much cheaper than modifying a delivered system. FS could be verified and validated in order to explore the consequences of this specification and potentially find mistakes. Prior to these, we have greater confidence on the testing of such system against the actual user requirement specification. This paper proposes the use of Z spec approach in designing Spatial Information Databases Integration Model (SIDIM) as a case study of artificial reefs (AR) assessment project. This model serves as a basis for the integration task, and declarative rule languages for specifying integration. Translation from source to target format is achieved by importing data from the heterogeneous sources to the middleware model, translating it to another middleware representation that better fits the target structure, and exporting the translated data to the target system. SIDIM is an integrated spatial and non spatial database model designed to enable easy access to users. It also enables evaluation of AR based on the growth of phytoplankton and zooplankton for each dedicated artificial reefs areas. SIDIM is considered as a method, model or new idea that ensures the effectiveness and efficiency of AR development projects.**

***Keywords*—SIDIM, Database Integration, Z Specification, Spatial Information System.**

I. INTRODUCTION

A formal software specification is a specification expressed in a language whose vocabulary, syntax and semantics are formally defined. This need for a formal definition means that the specification languages must be based on mathematical concepts whose properties are well understood. The branch of mathematics used is discrete mathematics and the mathematical concepts are drawn from set theory, logic and algebra [1].

In the 1980s, many software engineering researchers proposed that using formal development methods was the best way to improve software quality. They argued that the rigor and detailed analysis that are an essential part of formal methods would lead to programs with fewer errors and which were more suited to users' needs. They predicted that, by the 21st century, a large proportion of software would be developed using formal methods. Clearly, this prediction has not come true [2].

The use of formal methods is increasing in the area of critical systems development, where emergent system properties such as safety, reliability and security are very important. The high cost of failure in these systems means that companies are willing to accept the high introductory costs of formal methods to ensure that their software is as dependable as possible.

A. *Z Spec.*

Formal specification is a complete argument of mathematical representation and it is used to validate statement about system description. Usually, formal specification, verification and validation can be done manually. But nowadays, it can be done with the formal method support tools such as Z/EVES and VDM++.

Z/EVES is a tool that implements syntax checking and proof the specification. Regularly, developers cover a long time and looping process for syntax checking and proving, so there might be a greater possibility of mistakes. The proofs are efficiently when it's been presented in a user-friendly approach. Nevertheless, a lot of the proof that involved in software validation is naturally detailed, low-level and repetitious. So we can briefly state that it is unsuitable for human checking. Thus, formal proving supported by tool, which is not only reduce the possibility of mistakes but also not totally removes it [3].

The Z specification language is a way of decomposing a specification into small pieces called schemas. Each piece can be linked with a commentary that gives explanations informally the importance of the formal mathematics. A schema is essentially the formal specification analogous to programming language subroutines that are used to structure a system, where the schemas are used to structure a formal specification. The Z is physically powerful on sets and functions. Generally, Z notation is used for sequential situation. We interested in using the Z notation because it is a mature technique for model-based specification [4].

B. *Spatial Information System*

A spatial information system or geographic information system (GIS) is commonly regarded simply as a software application, but is really much more. It comprises a number of components that together make up a functioning system. These include: software, hardware, data, people, procedures and communications.

The data may include the spatial features (the spatial representations of real-world elements in point, line or polygon form) and attribute information (qualitative and quantitative information) that is linked to the spatial features. For example, a housing lot will have a number of attributes attached to it, including lot number, address, area, street address, owner, etc. Data linked to the spatial features may be contained within the system, or linked to it using database connections.

Most spatial information systems employ a similar approach to the thematic layering of maps-that is, the grouping of data into thematic layers (feature classes, levels, layers, coverage). This information may be referenced or tied geographically, which allows the overlay of various layers of information on top of each other.

Spatial information systems that are well integrated with other systems in an organization can provide information quickly and in desired form to managers and key professionals. This provides efficient and effective decision support.

II. DATA INTEGRATION

Research on data integration is one of the most important element and is a 'hot issue' in research of spatial data. Previous researchers concentrated on developing a common form of information that is easier for the user to manage the database and is less time consuming [4].

The history of integration method started in early 80's which is known as MULTIDATABASE, followed by a mediator called "GARLIC". Later, INFOSLEUTH was introduced and updated into a method of integration based on ontology which is known as OBSORVER. The integration method were then developed using peer-to-peer known as HYPERION. The journey of integration method led to current web based integration method which is known as ACTIVE XML [5].

Integration is defined as merger of various information from several sources that provide benefits in terms of collection aspect, processing time, resource conservation and data sharing for various purposes. Information that one wish to combine or integrate must be staying in a state of that analogous (uniform). However if the state of information to be consolidated is unequal, a modification process have is necessary.

To place the information in various databases and at different or scattered locations will cost difficulty in performing the integration process. Thus research to integrate '*every single piece into one*' should be initiated.

According to [6] concrete solutions to data integration is not available. In this study, the following methods were used:

1. *Manual Integration* - resolve interface which differ and various demand language, consumer need to know by detail on location, data presentation logical and semantic.
2. *Same interface* – consumer must be prepared with interface that uniform to facilitate information-seeking that require with usage one navigator.
3. *Integration through application* – application which enables information source and give revenue that is wanted by consumers.
4. *Integration through middleware (intermediary software)* – intermediary software reuse the function that can solve problem of data integration. Intermediary software using SQL command for data queries. Intermediary software difference needs to be combined so that the software will be allied.
5. *Standard Data Access* – logical data integration could be achieved globally although located in different places, hence virtually connected.
6. *Reserve the same storage utilization* – Integrating physical data may be done by the same reserve storage partnership. Local storage will process the integrated information faster.

III. SPATIAL INFORMATION DATABASES INTEGRATION MODEL (SIDIM)

The normal method of integrating data from one format to another is by writing a specific program for each translation task. Examples are the Oracle and MS Access to MYSQL Format. Writing such a program is often complicated by numerous technical aspects of the specific data sources that are not really relevant to the integration process (e.g. specific DB access protocol). A sound solution for data integration task requires a clean abstraction of the different formats in which data are stored, and means for specifying the correspondences between data in different world and for translating data from one world to another.

To achieve the above objective, a *middleware* databases integration model call SIDIM was introduced. This model serves as a basis for the integration task, and *declarative rule languages* for specifying integration. Translation from source to target format is achieved by (1) importing data from the sources to the middleware model, (2) translating it to another middleware representation that better fits the target structure, and (3) exporting the translated data to the target system.

SIDIM is an integrated spatial and non spatial database model designed to enable easy access to users. It also enables evaluation of AR based on the growth of phytoplankton and zooplankton for each dedicated artificial reefs area. SIDIM is considered as a method, model or new idea that ensures the effectiveness and efficiency of artificial reefs development projects. More importantly SIDIM is reduced in cost because it does not require hiring a special scuba task force. SIDIM enables the integration of more than one database either in same or different schemes.

This method is one method regarded as a new idea to ensure effectiveness of artificial reefs development project can be valued his effectiveness. As evaluation process previously requires fairly high cost required one special scuba unit to assess artificial reef levels of development and development

International Journal of Advances in Computer Science and its Applications – IJCSIA
Volume 4 : Issue 1 *[ISSN 2250 – 3765]*
Publication Date : 09 January 2014

by diving method. SIDIM is the proposed model for information combination of two or more databases type which possess same and different scheme.

In this study, integration involved only between three databases namely artificial reef database (ARPOS) [6], fish landing database (WiFISH) [7] and Boat Profile. The evaluation results of the databases will find that the location (position) of artificial reefs can be equal to the location of fish catches conducted. Due to this, the location based technique is a core and a fundamental to determine the effectiveness development level of the artificial reef project development at a certain location and also within the timeline. An equivalent proposed assumption can be formed as follows:

> *Catch location (CL) is equivalent with artificial reef development (AR). Assessment factor is dependent on the catch yield number (CT) for each type of fish (FT) and comparison with artificial reef type (RT) which included. A formula can be set up here. CL ≡ AR where if found CT ≡ FT is high then AR is effective which depending on catch date (DT) made. Vice versa if CT low, then AR to be ineffective.*

Based on that assumption formula, a newly develop algorithm call location based technique can assess the effectiveness level of artificial reef development project. The integration of location based process and scheme comparison would be running by intermediary software (middleware) which shall be develop refer to Fig. 1 by using 3-tier architecture [8].

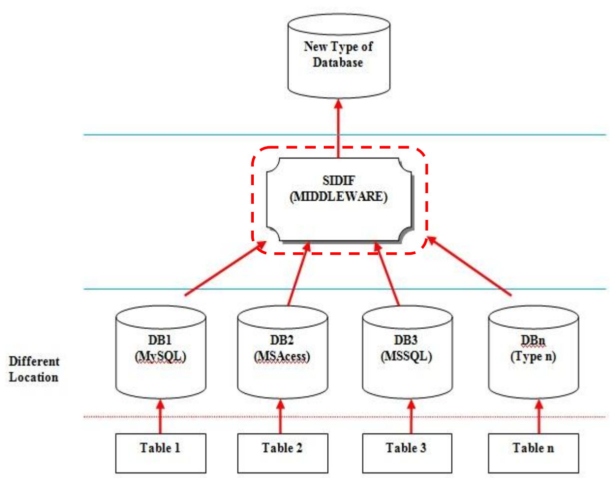

Figure 1. Overall SIDIM Architecture

Since those data in the three different databases are at different locations and in different formats, then intermediary software (middleware) is very much needed for integration purposes [9][10]. So that, the user requirement and design of the concept is needed before the development process is started. In this research we propose the formal specification approach for designing the model based on informal user requirement specification.

IV. RESEARCH METHODOLOGY

In the establishment of this research, time is mostly occupied in investigating the constructs of UML and formal methods. Below are the stages involved in this research's undertaking in detail as shown in Fig. 2.

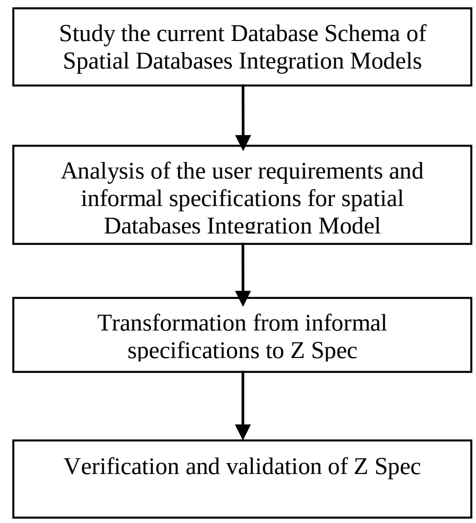

Figure 2: Research Methodology

A. Study of the current Database schema

Like those mentioned previously, the process of pre integration requires a study of the present database environment. The researcher has already developed an application called ARPOS. ARPOS database is in MYSQL and WiFISH database in MSSQL based. The system development uses open source 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 Lembaga Kemajuan Ikan Malaysia (LKIM) and were in a different location.

The evaluation results of the three databases are found that the location (position) of artificial reef can be equal to 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 format and types of each database could be found after the process was done.

B. Analysis of the User Requirements and Informal Specifications for Databases Integration Model

Formal specification begins with an informal statement of requirements to set scene in the database system. The kind of behavior that the user of the system might see is related to the operation that the user requires in the interface. The system architecture model for multiple types of spatial and non-spatial integration model is divided into 3 basic processes such as inputs, processing and output.

C. Transformation from informal specifications to Z specifications

The purpose of this stage is to prove the practicality and effectiveness of applying UML and then transform the class diagrams to Z specification using RoZ. This is translating a UML class diagram into Z schemas, and then the next step followed by adding static and operational constraints to complete it to the Z specification. We need this step to design a class diagram of this database integration by using rational rose environment. As soon as a UML class Diagram model is designed, we can follow by transforming it into a Z specification.

D. Verification and validation of Z specifications

Verification and validation must be done on Z specifications and theorems that been developed before a theorem proving process is implemented. A verification processes are done to make sure the correctness of those specifications and theorem. Using z/eves, the verification process will be running and checking the syntax errors as a whole. The result of this process can prove either all the schemas or theorems are free from any syntax errors.

V. RESULT: PROVEN THEOREM USING Z SPECIFICATION

In this paper, we will perform a formal specification for two databases in SIDIM model. FS is only part of a software development process. In order to develop a Z specification, we will use the mathematical theory to formalize the notions that the potential users of a system can rely on to illuminate their dealings with it. Table 1 show a *given sets* for each table in SIDIM database system.

TABLE I. DECLARATION OF Z SPEC.

Database	Given Sets Declaration
Artificial Reefs Distribution information	[*KODTUKUN, KwsnTangkapan, KddknLatitude, KddknLongitude, JnsTukun, BilTukun, ThnBina, JumPeruntukan, Biaya*]
Daily Fish landing Data Collection	[*NoBot, JnsPkakasan, TarikhDarat, MasaDarat, KodIkan, JnsIkan, JumHslTgkapan, KwsnTgkapan, PstDarat*]

From the table 1, we can integrate the data from both databases using mathematical expression described below:

Integration: DB1 φ DB2
dom *Integration == {x:DB1 | Ey:DB2∞x ÷ y ε Integration}*

The first item that needs to describe is the *state schema* for each database in SIDIM database system. The database can now be used to start the operation in the database system. An operation schema represents some operation that the system can perform. Fig. 3 illustrates the integration operation schema for all the information with different type of database in a single server.

TukunPendaratanRelRel
PendaratanikanExt
KedudukantukunExt
Pendaratan: KEDUDUKANTUKUN → PENDARATANIKAN
Tukun: PENDARATANIKAN → F KEDUDUKANTUKUN

dom *Pendaratan* = *Kedudukantukun*
ran *Pendaratan* ⊆ *Pendaratanikan*
Tukun
= { *x*: ran *Pendaratan* • *x* ↦ { *y*: dom *Pendaratan* | *Pendaratan y* = *x* • *y* } }

Figure 3. Z Schema for Integration of Multiple Types of Databases

PendaratanikanExt
∧ *KedudukantukunExt*
∧ *Pendaratan* ∈ *KEDUDUKANTUKUN* ⇸ *PENDARATANIKAN*
∧ *Tukun* ∈ *PENDARATANIKAN* ⇸ F *KEDUDUKANTUKUN*
∧ (dom *Pendaratan* = *Kedudukantukun* ∧ ran *Pendaratan* ⊆ *Pendaratanikan*)
∧ *x* ∈ ran *Pendaratan*
∧ *y* ∈ dom *Pendaratan*
⇒ *y* ∈ dom *Pendaratan*

The expressions above illustrate a special kind of relation that the element in KEDUDUKANTUKUN is related to the element in PENDARATANIKAN. After the design specification process completed, verification of the system were done using theorem proven to make sure that the schema has been performed correctly. Schemas' verification is needed to ensure that the specification meets the user requirement.

However, verification of the other design steps requires further automation to keep design development cycles under control. After we complete the formal specification, we need to verify it to check either all schemas fit the objectives. As mention above, the theorem proven use Z/EVES support tool. This tool will responds 'Done' if there are zero defects and verification can be completely done. Z/EVES extend the methodology of traditional checking techniques, such as the safety property of predicate checking.

Theorem proving involves verifying the truth of mathematical theorems that are postulated about the design using a formal specification language. It remains largely an academic area today and is unlikely to be viable for widespread commercial use in the near future. This section describes regarding the theorems that may represents the properties of operations in SIDIM. Pre-condition theorem proving based on operation schema that has been specified.

Theorem proving will only be perform when all the theorems that have been developed are free from any syntax errors. In this paper, pre-condition theorems were developed based on the operation schema in Z specification that had been discussed earlier. Pre-condition theorem proving will enable to prove that each operation does not applied outside domain where the results of an operation are unidentified.

a) Safety property of predicate *AddKedudukantukun*

theorem *AddKedudukantukun_Pre*

$$\forall KedudukantukunExt; kedudukantukun?: KEDUDUKANTUKUN \bullet \text{pre } AddKedudukantukun$$

$$\begin{aligned}&KedudukantukunExt \land kedudukantukun? \in KEDUDUKANTUKUN \\ &\Rightarrow (\exists Kedudukantukun': \mathbb{P} \langle\!| PembiayaanProjek: Biaya; \\ &\quad bilanganTukun: BilTukun; \\ &\quad jenisTukun: JnsTukun; \\ &\quad jumlahPeruntukan: JumPeruntukan; \\ &\quad kawasanTangkapan: KwsnTangkapan; \\ &\quad kedudukanLatitude: KddknLatitude; \\ &\quad kedudukanLongitude: KddknLongitude; \\ &\quad kodTukun: KODTUKUN; \\ &\quad tahunDibina: ThnBina |\!\rangle \bullet AddKedudukantukun)\end{aligned}$$

theorem *RemoveKedudukantukun_Pre*

$$\forall KedudukantukunExt; kedudukantukun?: KEDUDUKANTUKUN \bullet \text{pre } RemoveKedudukantukun$$

b) Safety property of predicate *RemoveKedudukantukun*

The above theorems proven have a pre-condition of operation scheme. The theorem consists of two state, that is, before operation and after operation been proved. A state

$$\begin{aligned}&KedudukantukunExt \land kedudukantukun? \in KEDUDUKANTUKUN \\ &\Rightarrow (\exists Kedudukantukun': \mathbb{P} \langle\!| PembiayaanProjek: Biaya; \\ &\quad bilanganTukun: BilTukun; \\ &\quad jenisTukun: JnsTukun; \\ &\quad jumlahPeruntukan: JumPeruntukan; \\ &\quad kawasanTangkapan: KwsnTangkapan; \\ &\quad kedudukanLatitude: KddknLatitude; \\ &\quad kedudukanLongitude: KddknLongitude; \\ &\quad kodTukun: KODTUKUN; \\ &\quad tahunDibina: ThnBina |\!\rangle \bullet RemoveKedudukantukun)\end{aligned}$$

before or after operation should meets a relationship that has been specified earlier. If that operation schema have input and output, then the statement above meant that a state after operation and issued output confirm by a state before operation with given input. If a result of this theorem proving is true, it meant that a pre-condition operation scheme accurately relates to a state after an operation and have complied relationship that has been specified.

Otherwise if theorem proving results is not true, in this case, it will not allow a right and clear pre-condition which can be accepted as a true results. This meant that a pre-condition of that operation schema is inconsistent. Inconsistency of the system meant that the operation schema or pre-condition theorem in that specification needs to be corrected because there is unidentified operation in domain.

Both theorems above have been proved using Z/EVES theorem proven through *"prove by reduce"* command. If the result of both theorems proving is *"true", i*t meant that both operations that have been specified are applied in the right domain and accurately relates to a state after an operation and have complied relationship that has been specified. So, we can conclude that the Z specifications used in the frameworks that have been discussed above are reliable. Once a formal specification, verification and validation is successful, we can proceed create a coding to develop the system.

VI. CONCLUSION

We hope that with the use of a formal specification approach in the design process will help in early removal of many errors that would not otherwise be detected during the end of stage. Also, the software prototype that will be developed will be robust, save the cost of development and reduce in time. Lastly, by using the formal specification and verification of a case study in fishery industries databases using Z/EVES support tool, could improve the quality of database integration.

ACKNOWLEDGMENT

We conveyed our appreciation to Malaysian Fisheries Development Authority (LKIM), Institut Penyelidikan Perikanan Malaysia, Fisheries Department and Ministry of Science, Technology and Innovation Malaysia (MOSTI) for contribution to the data used and ARPOS project which were funded under e-Sciences Fund research grant.

REFERENCES

[1] Sommerville I., *Software Engineering*, 6th Ed. UK: Addison-Wesley, 2001.

[2] Babich F. and Deotto L., 2002. Formal Methods for Specification and Analysis of Communication Protocol. IEEE Communication Surveys & Tutorials., 4(1): 2-15.

[3] Bowen J. , Formal Specification and Documentation using Z: A Case Study Approach: Thomson Publishing, 2003.

[4] Bowen J.P and Hinchey M.G., 1995. Ten Commandments of Formal Methods. Computer, 28(4): 56-63.

[5] Saaltink M., "The Z/EVES System," presented at ZUM, LNCS, 1997.

[6] Mustafa M, Yazid M.S.M, Khalid S. and Aezwani W.A.B.W (2008a). GIS spatial data visualization tools for artificial reefs

distribution. The 3rd International Conference on Mathematics and Statistics (ICoMS-3) Proceeding 2008. pp 25-38.

[7] Mustafa Man, Mohd Shafry Mohd Rahim, Muhammad Zaidi Mohd Rahim (2012). SIDIF: Spatial Information Databases Integration Model for Fishery Activities. Archive Sciences Journal – ISI Index.

[8] Parent, C. and Spaccapietra, S. 1998. Issues and approaches of database integration. CACM, 41(5): 166–178. 1998.

[9] Parent, C., Spaccapietra, S., Dupont, Y. 1992. Model independent assertions for integration of heterogeneous schemas. Very Large Database Journal 1. 81–126. 1992.

[10] W. Ismail and K.S. Joseph (2010). Future IT trends for GIS/Spatial Information Management. Scientific Research and Essay Vol. 5(10), pp. 1025-1032.