UNIVERSITI TEKNOLOGI MARA

A DATA MODEL FOR MULTIMEDIA DATABASE OF MALAYSIAN CULTURAL HERITAGE ARTEFACTS MANAGEMENT

MOHD NORASRI BIN ISMAIL

MSc IT

MAY 2007

UNIVERSITI TEKNOLOGI MARA

A DATA MODEL FOR MULTIMEDIA DATABASE OF MALAYSIAN CULTURAL HERITAGE ARTEFACTS MANAGEMENT

MOHD NORASRI BIN ISMAIL

Thesis submitted in partial fulfilment of the requirements
for the degree of
Master of Science in Information Technology

Faculty of Information Technology & Quantitative Sciences

May 2007

ABSTRACT

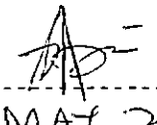
Multimedia is one of the much talked fenomena in the field of information technology. Currently, vast amount of multimedia content are created for multiple purposes particularly to convey information effectively. Cultural heritage domain is also benefited from multimedia technology where multimedia content is used for depicting tangible and intangible artefacts and for information dissemination. However, the vast amount of multimedia especially in the cultural heritage domain needed a special data model to cater the need of both collection management and digital multimedia content. There are finished and currently active research activities regarding multimedia database for cultural heritage application throughout the world especially in Europe. The problem is there is no research has been done in the local context especially in the field of multimedia database application in the cultural heritage domain. There are also no data model that support both information of cultural heritage and multimedia data in the local context. The intention of this research is to design a data model for multimedia database of cultural heritage artefact management as well as tries to fill the gap of cultural heritage dan multimedia database research in the context of Malaysia. A multiple case studies is conducted to gather information and requirements from the local museums governing organizations. Literature study also conducted to gain information regarding the guidelines, standards and reference models which is relevant to this research. From the requirements gathering as well as literature reviews, the logical and physical data for multimedia database of Malaysia cultural heritage artefacts management was designed and modelled. The data model employ a hybrid of museum metadata standard known as ISO 21127:2006 or CIDOC Conceptual Reference Model (CRM) and multimedia database standards known as MPEG-7. This research also employ a model developed by Jane Hunter (2002) where combination between CRM and MPEG-7 for describing multimedia in museums is made possible.

Candidate's Declaration

I declare that the work in this thesis was carried out in accordance with the regulations of Universiti Teknologi MARA. It is original and is the result of my own work, unless otherwise indicated or acknowledged as referenced work. This topic has not been submitted to any other academic institution or non-academic institution for any other degree or qualification.

In the event that my thesis be found to violate the conditions mentioned above, I voluntarily waive the right of conferment of my degree and agree be subjected to the disciplinary rules and regulations of Universiti Teknologi MARA.

Name of Candidate	Mohd Norasri Bin Ismail
Candidate's ID No.	760723-01-7351 / 2005102390
Programme	Master of Science in Information Technology
Faculty	Faculty of Information Technology & Quantitative Sciences
Thesis Title	A Data Model for Multimedia Database of Malaysian Cultural Heritage Artefacts Management

Signature of Candidate 

Date 31 MAY 2007

ACKNOWLEDGEMENTS

There are many people I would like to thank for the parts they played in making this research possible. Firstly, I would like to express my deepest gratitude and sincere appreciation to my supervisor, Puan Ariza Nordin, for her precious time, invaluable guidance, suggestions, comments, support and encouragement. Secondly, I wish to thank to all of my lecturers Associate Professor Dr. Nor Laila Md. Noor, Associate Professor Dr. Saadiah Yahya, Associate Professor Dr. Mohd. Isa Mohd. Samat, Encik Azlan Abdul Aziz, Puan Suriyati Razali, Encik Ali Seman, Encik Syamsulhairi Yaakop and Cik Ruhaila Maskat.

I would also like to thank my beloved wife, Masliana Bakar, for her patient, encouragement, love and support. To my loving sons Muhammad Akmal and Muhammad Luqman who always give me the inspiration and strength. Special thanks also to my parents, Ismail Nawawi and Norjanah Akhsan for their untiring support and unconditional love.

Finally, I would like to extend my gratitude to all my fellow graduate friends for the priceless support and contributions in making this thesis a success.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The aim of this research is to design a data model for multimedia database of cultural heritage artefacts management in Malaysia. The outcome of this study is the design of the database where it would be useful for the Government of Malaysia through the Ministry of Culture, Arts and Heritage in their effort and intention to manage and preserve our invaluable cultural heritage artefacts.

1.1 Background

Since the introduction of multimedia in personal computers, it has become more common every day to digitize part of the multimedia data around us. A major advantage of digitized data over traditional method is that digitized data can be shared easily with others. People now create their own homepages on the World Wide Web (WWW), partially as a tool to manage the information they collect. But, browsing the web makes clear that a computer with a web server is not the best tool to share your multimedia data. It is not easy for others to find your data, and, the information pointed at by search engines is often incorrect, or has been moved to another location.

The fact that the ever increasingly multimedia data which has been driven by the cheaper cost of computing technology as well as media capture devices is inevitable. As a consequence, multimedia database systems have emerged as an important research area for the storage, handling and retrieval of these vast multimedia data. A multimedia database system must provide the support for managing text, video, audio and image data, and it must also manage the retrieval of these data types.

Recent trend in the cultural heritage sector is creating digital collections of cultural heritage artefacts. Digital artefacts collection is known to be a new method of preserving cultural heritage. The idea of using digital artefacts collections for the cultural heritage artefacts is promoted by the fragility nature and space consuming characteristics of the physical artefacts. Digital artefacts collections may stores the artefacts permanently in the digital form (2D/3D images, graphics, audio, video or animation) for the purpose of preservation and archiving. Another advantage of digital cultural heritage artefacts collections is the dissemination for the purpose of research, education or tourism through portal and of course supported by multimedia database system for storage, retrieval and manipulation.

The leverage of multimedia data, advantages of having cultural heritage artefacts in digital form (i.e. multimedia) and the need of multimedia database system for digital cultural heritage artefacts collection storage, management and dissemination as discussed above motivate this research.

1.2 Problem Statement

There are finished and current active researches in cultural heritage information system especially in Europe where proposes a better handling of cultural heritage artefact data such as 3D MURALE (Grabczewski et. al., 2001), eCHASE (Sinclair et. al., 2005), BRICKS (Risse et. al., 2005), ACIS (Klamma et. al., 2005) and many more. On the other hand, very limited research and development are being done in Malaysia. Multimedia database system is necessary to support the ever increasing number of digital cultural heritage artefacts collections. The problem is, even though cultural heritage artefacts are currently being actively digitized, to date no multimedia database system has been implemented for the storage, management and preservation of these collections in Malaysia. Based on the stated problem, this research is done to study and design a data model for multimedia database of cultural heritage management.

1.3 Research Questions

The research questions of this thesis are as follows:

i. What are the models and standards for cultural heritage artefacts management system?

ii. What are the localized requirements of cultural heritage artefact management system as well as multimedia database implementation?

iii. How to design the multimedia database for cultural heritage artefacts management system?

1.4 Research Objectives

The objectives of this thesis are as follows:

i. To identify reference model and standards for cultural heritage information system.

ii. To identify requirements of cultural heritage artefacts management system and the database management system technology for museums in Malaysia.

iii. To design the logical and physical data model for multimedia database of cultural heritage artefacts management.

1.5 Contribution of the Study

The ever increasing of digital cultural heritage artefacts in various media types needs a suitable database system for data handling, manipulation and dissemination. This is needed to effectively manage Malaysia invaluable cultural heritage treasures. This research contributes in the data handling, management and dissemination of cultural heritage artefacts by designing a data model for multimedia database in the perspective of Malaysia's cultural heritage artefact management. This research will also support the Ministry of Culture, Arts and Heritage in achieving their missions "to highlight and popularize the arts and culture" and "to preserve national heritages in its tangible and intangible form to cultivate patriotism" (Ministry of Culture, Arts and Heritage, 2004) as one of major electronic government application. Moreover,

this research will also might be a base for the implementation of nationwide electronic cultural heritage online application for supporting cultural heritage management, education, cultural tourism and so on.

1.6 Research Scope

This study investigates the implementation of multimedia database in the cultural heritage artefacts management in Malaysia. The setting of the study is museum institutions in Malaysia. The requirements from the cultural heritage institutions will be collected for designing logical data model of multimedia database. In this study, only tangible cultural heritage artefact in the form of images will be taken into consideration.

1.7 Thesis Organization

The content of this thesis is presented in 7 chapters. The remainder of this thesis is organized as follows:

- **Chapter 2: Multimedia Database System**
  This chapter gives an overview of database technology and particularly discussed the multimedia database system.

- **Chapter 3: Cultural Heritage Information System**
  This chapter give background information of the cultural heritage domain especially in cultural heritage information system and standards.

- **Chapter 4: Research Methodology**
  To answer the research questions, this chapter outlines the research design.

- **Chapter 5: Analysis and Findings**
  This chapter presents findings from data analysis.

- **Chapter 6: Discussion**

  This chapter discuss the achievement, contribution, significance and limitations of this research. The future related work is also suggested.

- **Chapter 7: Conclusion And Recommendations**

  This chapter present overall conclusion of this thesis.

CHAPTER 2

MULTIMEDIA DATABASE SYSTEM

The proliferation of multimedia data in this era causes the critical need of effective handling and management of these kinds of data. Database management system (DBMS) has proven to be efficient and effective in data handling and management. However, traditional DBMS is not capable of handling and managing multimedia data efficiently because of the nature of multimedia data which is different compared to the traditional textual data. Therefore, multimedia database system is needed for handling and managing multimedia data.

To design a multimedia database system, an understanding of database technology as well as multimedia database system is necessary. Initially, brief overview of database includes database definition, database management system, database system, and data model will be given. Later, multimedia database system will be discussed. Included in the discussion of multimedia database system are; definition, multimedia data types, characteristics, storage, indexing, retrieval, querying, issues and current research in multimedia database system.

2.1 Overview of Database Technology

Since the birth of computers, data handling, manipulating and storing received great notice. Database technology is one of the major field in computing which concentrate on issues in data handling, manipulating and storing. There are numerous explanations and definitions of what a database actually is. The definition of database is a collection of related data (Elmasri & Navathe 2000; Connolly & Begg, 2005) and a description of this data, designed to meet the information needs of an organization

(Connolly & Begg, 2005). Elmasri and Navathe (2000) list three implicit properties for a database:

- A database represents some aspect of the real world, i.e. some aspect that is of interest for users.
- A database consists of a collection of data which is logically coherent and has some inherent meaning
- A database is designed, built, and populated for a specific purpose.

Usually, when "database" is used, it is mean something that is stored electronically (on disk, on tape, on in a computer's main memory) for handling, manipulation and retrieval, and is managed by a specialized software, a database management system (DBMS). The importance of database system is the information gathered from the data which can be used in analysis and decision making. Databases are structured to facilitate the storage, retrieval, modification, and deletion of data for various data-processing operations.

2.1.1 Database Management System (DBMS)

A DBMS is a generalized set of software package for accessing, implementing and maintaining one or more computerized databases (Elmasri & Navathe, 2000; Beynon-Davies, 2004). Connolly and Begg (2005) give clearer definition by defining DBMS as a software system that enables users to define, create, maintain, and control access to the database. The DBMS defines data types, structure and constraints before constructing a database by handling incoming data for storage and organizes it. The DBMS should also capable of manipulating data by providing users or applications ways to extract and modify the stored data. Another DBMS capability is allowing database sharing and concurrent access to multiple users or applications. Last but not least, the DBMS protects and maintains a database.

2.1.2 Database System

A database system can be defined as the combination of a database, a DBMS, and application programs (Elmasri & Navathe, 2000; Connolly & Begg, 2005). In a database system, database is functioning as data repositories where all data and data definition are stored. Database management system (DBMS) basically functioning as a mediator between application programs or queries and database. Users may retrieve the desired data or information using application programs or queries.

An important aspect of a database system is that a description of the structure of the database is stored in the database itself. Such information is called metadata ("data about data") and is used by the DBMS in its work to access the data files. In a relational database the names of tables and columns are examples of such metadata.

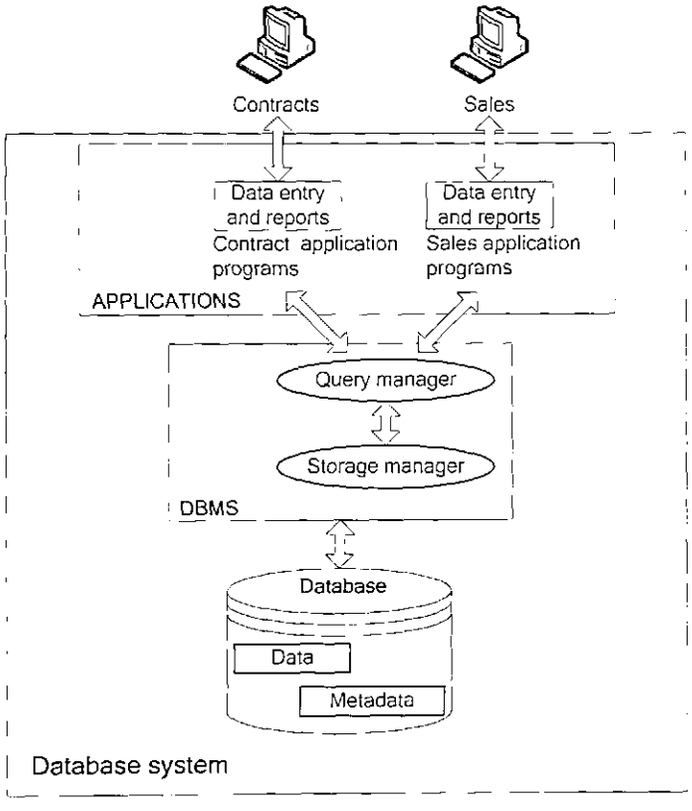


Figure 2.1: Database System Environment

2.2 Data Model

A data model can be loosely used to describe an organized and ordered set of data stored on a computer (Powell, 2005). Data model provide a solution for structuring these ordered set of data in order to more efficiently retrieve and manipulate that data. The purpose of data model is to represent data and to make the data understandable in order to be easily used for database design (Connolly & Begg, 2005). Several data model was introduced since the advent of database technology including network model, hierarchical model, relational model, object-oriented model, object-relational model and etc.

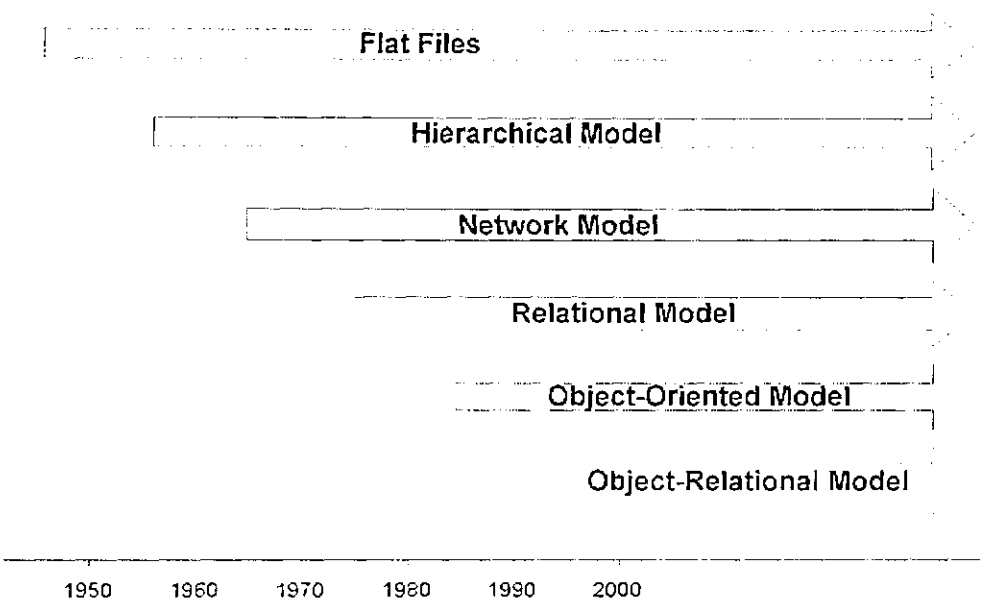


Figure 2.2: The Evolution of Data Modelling (Powell, 2005)

2.2.1 Flat Files

Before the advent of database technology, the approach for data storing and manipulation was file-based system. File-based system is a collection of application programs that perform services for the end-users such as the production of reports where each program defines and manages its own data (Connolly & Begg, 2005). This means only certain programs can access certain data files because of the specifications and parameters were only defined in that particular program. As for

example, the payroll department could only access files used by the payroll department programs, the accounts department could only access files used by the accounts department programs and the sales department could only access files used by the sales department. Using a file-based method mean that no modelling techniques are applied and that the data is stored in flat files in a file system, utilizing the structure of the operating system alone (Powell, 2005). The term flat file means a simple text file where data is simply dumped without any structure.

The file-based system had several limitations including separation and isolation of data, duplication of data, data-application dependence, incompatible file formats and fixed queries/proliferation of application programs (Connolly & Begg, 2005).

- **Separation and isolation of data**

  File-based systems have the tendency to separate and isolate logically-related data. With important data being spread out across different application, the retrieval of crucial data can be time consuming.

- **Duplication of data**

  Redundant data in different files may occur especially when different application uses the same data but in decentralized manner. Any modifications of data in one application is not affect the same data on the other application, thus create inconsistencies.

- **Data-application dependence**

  Since the data definition and structure was controlled and may only be defined within a particular application, a change in these parameters required a change in the application making it difficult and time consuming to change or add in new data.

- **Incompatible file formats**

  Applications may be incompatible with each other, preventing data from easily being transferred from one application to the next.

- **Fixed queries/proliferation of application programs**

  File-based systems are very dependent to the fixed query where required during the development of the application programs. However, from time to time new or unplanned queries may be required where it is not supported by the application programs. Other limitations of file-based systems were including no provision for security and integrity; limited or non-existent of recovery process; and access to the application programs was restricted to only one user at one time.

All of above limitations may increase program maintenance and application development time, hence labour intensive and cost consuming. Therefore, the database management system was introduced to address these shortcomings.

2.2.2 Hierarchical Model

The development of Information Management System (IMS) by a joint venture of IBM and North American Aviation (NAA, now Rockwell International) is one of the important milestones in database technology. The IMS was developed based on Generalized Update Access Method (GUAM) developed by NAA for NASA's Apollo moon-landing project in the 1960s where huge amount of data needed to be handled. GUAM was based on hierarchical structure which conforms to an upside-down tree. The hierarchical model is derived from the IMS and was adopted by many banks and insurance companies as well as inventory and accounting systems used by government departments and hospitals.

The Hierarchical Model stores data in a series of records where it is equivalent of tables in the relational model. The structure of this model is much like an inverted tree where the relations of tables are using parent-child relationship. Each parent table can have multiple child tables while each child table can only have one parent table. Therefore, child tables are completely dependent on parent tables where a child table can only exist when corresponding parent table exist.

- **Fixed queries/proliferation of application programs**
  File-based systems are very dependent to the fixed query where required during the development of the application programs. However, from time to time new or unplanned queries may be required where it is not supported by the application programs. Other limitations of file-based systems were including no provision for security and integrity; limited or non-existent of recovery process; and access to the application programs was restricted to only one user at one time.

All of above limitations may increase program maintenance and application development time, hence labour intensive and cost consuming. Therefore, the database management system was introduced to address these shortcomings.

2.2.2 Hierarchical Model

The development of Information Management System (IMS) by a joint venture of IBM and North American Aviation (NAA, now Rockwell International) is one of the important milestones in database technology. The IMS was developed based on Generalized Update Access Method (GUAM) developed by NAA for NASA's Apollo moon-landing project in the 1960s where huge amount of data needed to be handled. GUAM was based on hierarchical structure which conforms to an upside-down tree. The hierarchical model is derived from the IMS and was adopted by many banks and insurance companies as well as inventory and accounting systems used by government departments and hospitals.

The Hierarchical Model stores data in a series of records where it is equivalent of tables in the relational model. The structure of this model is much like an inverted tree where the relations of tables are using parent-child relationship. Each parent table can have multiple child tables while each child table can only have one parent table. Therefore, child tables are completely dependent on parent tables where a child table can only exist when corresponding parent table exist.