

## A Study on Art Repository Buildings Based on Data Warehouses Technology

MA Geng<sup>1, a</sup>

1 Center for Liberal Arts Education  
North China University of Water Resources and Electric Power, Zhengzhou 450046, China.

<sup>a</sup> magengmg@126.com

**Abstract** — In this paper, we cover current research studies to realize technical standards for elaborate art resource databases. We use the principles of database construction, methods and significance analysis to clarify current issues used. We discuss common implementation methods and technical specifications of relevant issues. The database of art resources forms a platform for analysis and design, and gives a detailed description of the application and implementation of the development directions and key technologies. We provide analysis of: i) database storage technology development and implementation of the system, ii) analysis and design of the art resource database, iii) the important aspects of the data platform, iv) administrator and user functions, and v) the realization of some of the features of specific cases.

**Keywords-** *Art Repository, Data Warehouse, Data Element, Art Resources*

### I. INTRODUCTION

With the development of society, the era of progress, open field of vision than before, we should look at the development of scientific school in the future. Development of school libraries should also follow the trend, to promptly shift to digital library transformation from the traditional library, where the construction of various databases will become one of the main construction of digital resources. Professional institutions for small and medium libraries, special database construction in the direction of the topic should be practical, advanced and scientific principles to be followed in the construction of the fine and principles [1-2]. This paper from the professional art thematic databases produced art resource database of view, in an attempt for the library is concerned, more convenient and practical perspective to explore the topic Database. Art resource database has its special status and role in the automation, networking, and digital technology development direction of digital libraries. It makes up less than one million traditional paper books, abstract databases and full-text database of digital content and refined and specialized, can provide fine arts disciplines latest research results, the level of research and research trends and other information, allowing users to get first-hand information, thereby grasp the art industry research, to achieve new research results.

Present computer hardware and software technology, storage technology, information technology has been in a relatively developed stage, database construction has entered a rapid development period, it will become the core of Digital Library Construction. Modern libraries digital collection data accumulation and quality of library information service capabilities and strong academic research background, the increasing library research position teaching in colleges [3]. In the network environment, full development and utilization of information resources, must be stored in a database approach to network dissemination service. Development of databases at the same time promote

the development of related hardware and software supporting the technology, especially with the high-speed information network database combines information service opens up new areas to make information sharing possible, database services reached an unprecedented level, thus promote the development of scientific research. Art database to provide accurate, adequate and effective scientific research important information quickly, so researchers save time and avoid duplication of effort, promote innovation and development research. Art database in addition to setting the thematic subjects, the better the flexibility to set up interdisciplinary topics. Art thematic database of national economic construction, scientific research, culture and education development has provided important information security.

Art resource database content to meet user demand for urgent art resource data can be fully functional library information resources, centralized forefront of professional artistic disciplines or thematic latest results, in order to solve an unlimited number of Internet clutter of information and user needs contradiction between limited. As certain academic advantages library art institutions and professional advantages of digital resources is relatively backward, how in the new situation, breaking the traditional model, play their own unique advantages, to build a unique art professional digital repository, and ultimately art digital resource sharing between institutions is an important direction of development of today's art industry.

### II. OVER VIEW OF DATA WAREHOUSE

Data warehouse is based on body-oriented, highly integrated, feature has stability and timeliness. Data warehouse main function is to long years of accumulated wealth of information through the use of various analytical methods such as online analytical processing, data mining and further support such as decision support systems (DSS), decision support systems can help decision makers manage a lot of historical data from the analysis of the scientific

information useful to provide the necessary basis for decision-making decision-makers [4-6].

Data warehouse is essentially an information integration technology, data warehouse to obtain the raw data from multiple information sources, after finishing, stored in an internal database data warehouse, by providing access to tools to users, data warehouse can have a unified, coordinated and integrated information environment, so the data warehouse should have structure and functions described in Figure 1.

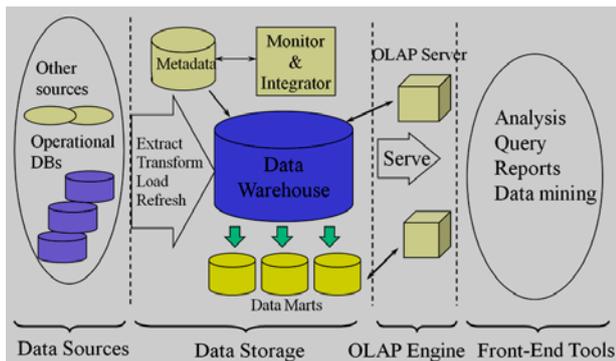


Figure 1. Function structure of data warehouse.

Data warehouse solutions focused on the first, that the data in a relational database on more practical common multidimensional analysis, multi-dimensional analysis can be a good business decision support analysis and processing, after processing as a necessary basis for analysis provide Managers use to decision-making [7].

Data warehouse data model. Data warehouse data model has based multidimensional characteristics, and the development of a data warehouse is generally driven by the data model, data model is a reflection of the reality of things and abstract, the model data as a data cube form [5]. Creating a data warehouse data modeling process is the most critical technical steps, therefore, to build a line with demand, the practicality of a data warehouse, select the data model is the key technical aspects. Select the data warehouse data model is generally in demand after analysis of business requirements analysis begins, correct complete data model can well reflect the business needs of users, data warehouse modeling this step is also the success of the data warehouse projects The most important technical factor. Data warehouse model can be divided into star model, snowflake model and hybrid model, details are as follows:

1) Star Model: star model is the most common model paradigm, it is made by a modeling paradigm a little external radiation, the object model of the star center is a center table, the center is a table does not contain redundant I data table, often called "fact sheet", around the center of the table is a set of connected thereto attached tables, often called "dimension tables", the fact that each dimension table has a corresponding dimension table associated with it. If the query when the query and query the fact table dimension table together, you can retrieve a lot of information. Look like star-shaped model of the stars, that is, by way of

dimension tables around the center of the table displayed on the ray in shape, showing the status of a star model.

2) Snowflake Model: snowflake model name suggests, is a star-shaped model of extended application. In terms of dimension tables, snowflake model is more scientific and standardized than the star model. Snowflake model is the star model in dimension data in the table to take further decomposition operation, will join the data decomposition of the attached tables, on the external features reflect the shape of the snowflake. So called "snowflake model." The biggest advantage snowflake model is decomposed by the data makes data-dimensional table reduces redundancy, makes the dimension tables easier to maintain, increase the flexibility of the application. The downside is that the snow-based data model during query operations, the need for a lot of table join operation will result in reduced query efficiency. Therefore, in the data warehouse design system, the current star model is more popular than snow models.

3) hybrid model: mixed mode between star model with snowflake model, star model in the fact table and dimension tables are the main components. Snowflake dimension tables model standardization and standardized settings. In the hybrid model, only the maximum standardized dimension tables.

### III. RELATED KEY TECHNOLOGIES

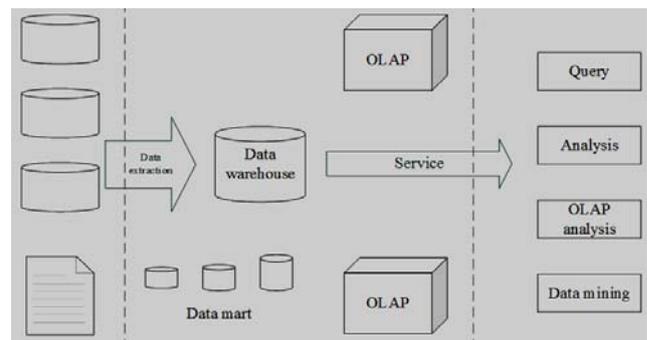


Figure 2. System structure of the data warehouse.

Data warehouse is subject-oriented, integrated, non-renewable, with the change of time and constantly changing, these characteristics determine the design of data warehouse system cannot be used with the development of traditional transaction-oriented database system the same design. The original demand data warehouse system is not clear, and the constantly changing and increasing developer originally not exactly aware of the user's explicit and detailed requirements, users can offer nothing more than a large part of the demand as well as the direction of demand, but not more accurate to anticipate future needs. And system development is a continuous cycle through, leaving the system feedback and improve the growing process. System structure of the data warehouse is shown in figure 2.

Extracted data. Extracting data is the data into the warehouse entrance. Since the data warehouse is a separate data environment, it needs through the extraction process to import data from online transaction processing systems, external data source, offline data storage medium into the

data warehouse. Data extraction technology in several areas mainly related to interconnection, copy, incremental, conversion, scheduling and monitoring. Data warehouse is not required to maintain real-time synchronization and online transaction processing systems, so data can be extracted timing, but a plurality of extraction time operation performed, mutual order, the success of the validity of the information in the data warehouse is essential. Since the data extraction process involves the conversion of data, it is closely related to the practical application of a portion of its complexity makes embedded user-programmable non-extraction tools often cannot meet the requirements. Therefore, the actual data warehouse implementation process often do not necessarily use the extraction tool. The entire extraction process can incorporate the use of tools and effective management, scheduling, and maintenance is even more important [8-10]. When a row of data is copied from the source to the destination, the basic process of data extraction as follows:

- 1) extracting a line of source data.
- 2) is selectively applied to convert the line.
- 3) Try the row data into the destination buffer.
- 4) exception error handling.
- 5) If you specify a batch, the result is stored to the specified batch.
- 6) Repeat the above steps until the specified batch size, and then submit the data stored in the buffer, followed by a group of data or data go back to the beginning with the next batch.
- 7) When the data extraction After processing the last row of data, then the end of the task, while data extraction operation is terminated.

Storage and management of data. The first problem encountered in the data warehouse is to store and manage large amounts of data. What is involved here is much larger data than traditional transaction processing, and accumulate over time. Relational database system has been supported by data partitioning technology, the ability to disperse a large database tables across multiple physical storage devices, further enhancing the system manage large amounts of data expansion capability.

The second problem to be solved is the data warehouse parallel processing. In the traditional online transaction processing applications, features user access to the system is short and dense; for a multiprocessor system, it is possible to balance the user's request to share is the key, this is the concurrent operation. In a data warehouse system, features a user access to the system is large and sparse, each query and statistics are complex, but the frequency is not very accessible. The system needs to have the ability to mobilize all processors for this one complex query service, the requests processed in parallel. Therefore, parallel processing technology in a data warehouse is more important than ever. Currently, the relational database system in parallel processing has been able to do for query decomposition parallel, parallel-based data segmentation, and cross-platform support multiprocessor cluster environment and MPP environment that can support up to hundreds of

processors Hardware system and maintaining performance scalability.

Performance data. Data is actually equivalent performance data warehouse facade, its performance is mainly concentrated in the multi-dimensional analysis, mathematical statistics and data mining. The multi-dimensional analysis is an important manifestation of the data warehouse, in recent years due to the development of the Internet, makes tools and products in the field of multi-dimensional analysis is more focused on providing front-end Web-based online analytical interface, rather than just publishing data on the Internet.

#### IV. THE CONTENT CONSTRUCTION OF THE RESOURCE DATABASE

In the art of digital library resources database construction automation development direction of digital, networked, has its special status and role. Which make up the traditional paper book to millions of abstracts database and digital full-text databases, content and nature and design, the latest findings may provide art classes, and study the level of dynamic information, such as user get first-hand information, master the art of industry research, new research results [11]. The main task of art resources database system platform is based on the needs of users to gather information, to provide users with accurate content, secondary processing, convenient and flexible project information. The main construction process is as follows: to establish the principle, determine the scope of the art resource database to collect information. Strict adherence to standardized digital standard system processes the information collected to form a standard digital file. + Manually using automatic indexing and indexing method to generate metadata for a digital document. It will provide in the form of a network or a single project database to the user, and provides database retrieval systems, including full-text search. Additional data maintenance and new data. Arts resources database resource construction flow chart shown in Figure 3.

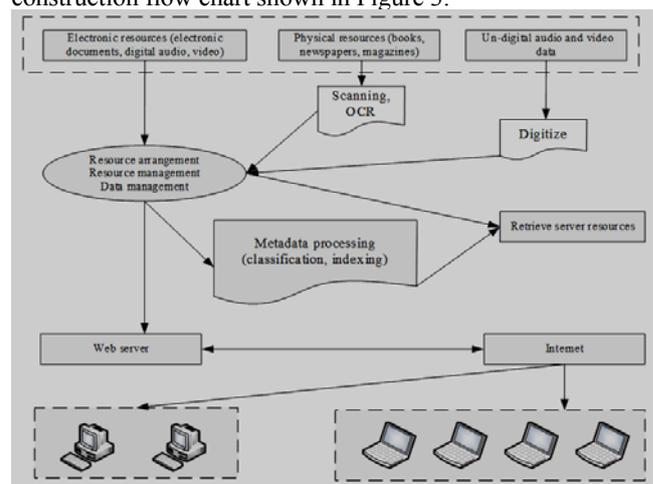


Figure 3. Flow chart of art resources database resource construction.

Content management is the core of art resources database system platform server resources retrieval, it implements the information resource organization of production, storage and retrieval. The traditional library into digital collection; Provide fast retrieval service; Content management platform can make a variety of network database can meet the needs of different users. On the same level management used by different permissions for database management or operation.

V. DATA ANALYSIS OF ART RESOURCES DATABASE

Art content resource database construction content resource database construction directly affects the quality and the value of the database resources, resources to build database of special topics on art, invest a lot of energy to the construction of digital content, work mainly as shown below.

The picture database. Using library rich collection resources, including books, periodicals, electronic publications in the form of material library, and the large amount of material on the Internet, and so on. Shows the latest art research and development trends, scientific research achievements at home and abroad, to help understand the arts teaching, scientific research personnel system frontier development level, grasping the direction of scientific research.

The teacher students work of library. Teachers college's works is the outstanding achievements in teaching, reflect the academy teaching scientific research and information. Digital works by scanning the teachers and students, build a library of digital works, readers can refer to.

The dissertation library. The writing of the graduation thesis is students to reflect their own talent and learning as well as a way of its own value, comprehensive utilization of their theoretical knowledge learned in the university life to analyze guide practice, this is a form of student ability, and a comprehensive evaluation of teaching effect and information feedback. Students leave for the school graduation thesis is the inevitable outcome of the school teaching, not only is the teacher teaching content reform of a promotion in the future.

Academy of fine arts gallery works library. Academy of fine arts gallery is an academy of fine arts at home and abroad for academic exchange of the window, as well as professional, academic and research of arts institutions. There are many different kinds of art museum's collection, 4000 about collection.

CD retrieval. It includes CD attached with books, and CD-ROM database information.

VI. ARCHITECTURE OF THE ART RESOURCES DATABASE

In the data warehouse design concept model, we identified several basic subject areas, here is a conceptual model design

Several basic subject areas identified in the analysis step, and choose the subject field to be implemented first. Granularity Division level determines the data warehouse, the appropriate level of granularity division not directly affect the amount of data warehouse and the appropriate query type. To select the appropriate data segmentation criteria, generally consider the following factors: the amount of data, data analysis and processing of the actual situation,

simple and granularity division strategy. The amount of data size is a major factor to decide whether and how to split the data were split; require the data analysis process is to select a data segmentation criteria based primarily on data segmentation is because with object data analysis process closely; we have to consider to the selected data segmentation criteria should be natural, easy to implement: taking into account also the standard and level of data granularity division is divided adaptation. Each theme data warehouse is implemented by a number of tables, relying on the theme of public code key link between these tables together to form a complete theme. In the design of the conceptual model, we determined the basic theme of the data warehouse, and the public code keys for each topic, the basic content was described here on the theme of the selected mode to the current enforced division, forming a plurality of table and determine the relationship between mode of each table.

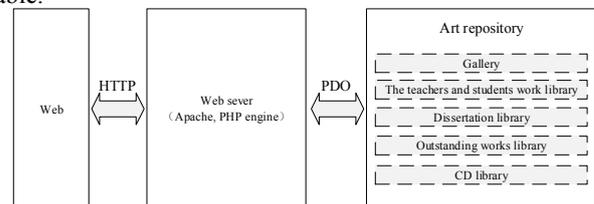


Figure 4. Architecture of the art resources database.

Art resources database management system uses the mature technology of three layer architecture technology, logically divided into system and user interface layer, application service layer and database layer, the system architecture is shown in figure 4. In the three layer structure, the main functions of the system and business process mainly concentrated in the application service layer, namely the Web server, database is mainly used to store agricultural resources data and information system management, user interface layer is mainly used to display the system running results and accept user operation.

User interface layer. Which is through the browser user interface, to realize the system the interaction with the user.

Application service layer. The Web server's level, the level is the key to the system level, the system is mainly the business processing and related interface implementation realized in this layer. The layer at the same time is also a bridge that crosses from the user interface and the database layer.

The database layer. It includes agricultural resources basic attribute database, database system management, production management and transaction management database, to be responsible for the system of data storage, organization and optimization.

VII. THE OVERALL DESIGN FOR ART RESOURCES DATABASE

Information Resources Construction Process art resource database includes information resources, collection, organization, where the organization of information resources is the most critical part, directly affect the

availability of information, with the ultimate goal is to establish information resource collection and retrieval tools conducive to the development of information resources and recycling, the relationship of information resources and information resources and organizational building is shown in Figure 5.

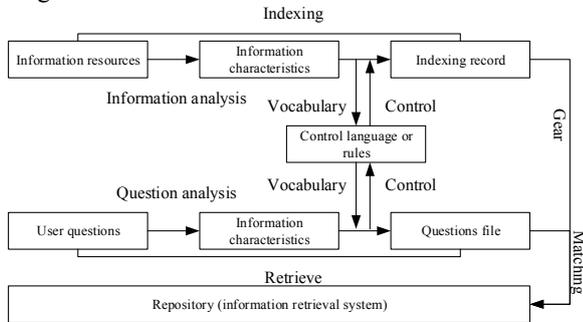


Figure 5. The relationship between the organization and utilization of information resources.

Collection of information resources requires fast, effective, orderly and lawful manner the various media resources collected digitized and stored in digital repositories, which involves many aspects of technology, management and law, which automatically Chinese books and materials entry is the biggest difficulty, the most arduous task. External characteristics of information resources of the information refers to a substance carrier directly reflect the characteristics of the external and forms, such as physical presence information, title, owner, and type of information, generation time, sources, distribution channels and so on. Content characteristic information refers to information contained in the specific content and the carrier, that carrier transfer and exchange of information specific content. And in accordance with the basic object it is external features and content features information organized information resources of these two aspects. Information Resources Organization and can be called information organization, information technology or information collation sequence is the use of certain scientific rules and methods of external features and content feature information to describe and ordering, so that information into disorder ordered information, user-friendly retrieval of information, access, and effective use. Organization of information resources is the basis for the use of information retrieval, information retrieval and use of information resources is the ultimate purpose of the organization, the two complement each other. Information is organized on the basis of the use of information resources retrieval, usually based on the retrieval and use as needed to a certain unit of information resources for the unit, record storage resource characteristics required by the application of information resources necessary correlation processing, the processed information resource, record or entry by the method specified input system. Information Resources Organization Procedure To define e-books can cover a variety of media types, including metadata standards, as well as content indexing and classification method based on this specification. Due to the current digital resources typically TB meter, manual indexing and classification clearly

unrealistic, we need to study the automatic processing technology-based multimedia content.

Arts Resources Database mainly collect art theory, skill books, picture albums, art journals and generations of painting Treasures function. Data on the number, quality, and class has been formed its own characteristics and systems. On the basis of these resources, integration of resources Subject Librarian database must Academy of Fine Arts Library Literature and Internet resources of art, the art form. User database system access process design library system is divided into two categories, one for the ordinary user, the other is the administrator user after the user logs in these two libraries, the system will enter a different user different operating space. Administrators space including category management, resource management, user management database system management functions, common user space limited resource search, browse, upload, download, and so on. The art resource database is shown in figure 6.

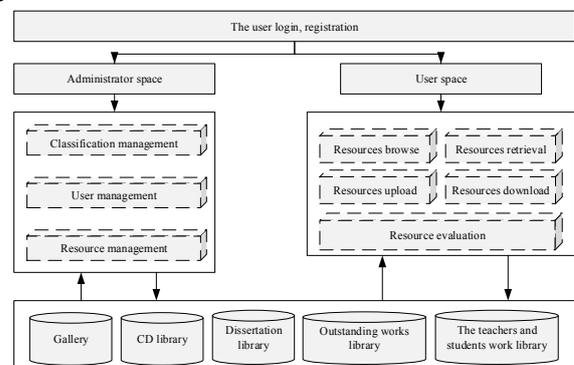


Figure 6. The art resource database for academy of fine arts.

The database contains various functional modules, including nine modules user registration module, the user login module, resource retrieval module, including system management module is divided into category management module, user management module and resource management module three sub-modules. The user's browser to connect through an intermediary server to the database server, the middle layer is the result of a request by the back-end database server, and database operations corresponding request after processing feedback to the user's browser through the middle tier server.

VIII. CONCLUSION

With the era of knowledge economy, people pay more attention to the integration and use of information. How to effectively collect, collate, process and organization of thematic information resources for researchers specializing in the field of learning and communication, with a more far-reaching significance. Through the establishment of a database of art resources, facilitate the user access to information resources, allowing users to unrestricted time and space and ability levels, saving time and effort. Paper completed the overall planning and design art resource database. Analyze and classify artistic resources. According

to the characteristics of the art courses and the content of the repository design specific circumstances, combined with professional art teacher suggested that the resources in a reasonable classification. We completed the function of each module implementation discussed, and discussed the implementation of value in the process.

This paper discusses the design and implementation process focuses on the entire system, focusing on practical in every aspect of the discussion, while the relevant theoretical knowledge discussed in depth and breadth are inadequate. Network, mirror sites database indexing automatically collect information of the thematic issue for collection of paper documents, audio and video files of this thematic information, this paper gives a solution, and for thematic information network, the subject uses Download is still artificial, artificial indexing approach, which obviously cannot meet the needs of Special Database Construction.

#### REFERENCES

- [1] Kimball R, Ross M. "The data warehouse toolkit: the complete guide to dimensional modeling". John Wiley & Sons, 2011.
- [2] March S T, Hevner A R. "Integrated decision support systems: A data warehousing perspective". *Decision Support Systems*, vol.43, No.3, pp.1031-1043, 2007.
- [3] Mandinach E B, Honey M. "Data-Driven School Improvement: Linking Data and Learning. Technology, Education--Connections (TEC) Series". Teachers College Press. 1234 Amsterdam Avenue, New York, NY 10027, 2008.
- [4] Goble C, Stevens R. "State of the nation in data integration for bioinformatics". *Journal of biomedical informatics*, vol.41, No.5, pp.687-693, 2008.
- [5] Gupta C, Mehta A, Wang S, et al. "Fair, effective, efficient and differentiated scheduling in an enterprise data warehouse". *Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology*. ACM, pp. 696-707, 2009.
- [6] Williams G. "Data mining with Rattle and R: the art of excavating data for knowledge discovery". Springer Science & Business Media, 2011.
- [7] Bury S. "Developing NYARC: the New York Art Resources Consortium". *Art Libraries Journal*, 2011.
- [8] Wei Y, Wang X. "Talking about the Construction of Tianjin's Culture and Art Resources Database". *Sci-Tech Information Development & Economy*, 2014.
- [9] Zhao M. "The Building of the Database of Art Resources Research for Academy of Fine Arts". *Applied Mechanics & Materials*, pp. 644-650:3057-3060, 2014.
- [10] Simor S. "Visual Art Resources Online: Issues, Trends and Challenges". *Art Documentation Bulletin of the Art Libraries Society of Nort*, vol.22, No.1, pp.33-40, 2003.
- [11] Tobar C. "Between The Lines: Arte Cubano: A Reference Guide to Cuban Art Resources at the New York Public Library". *Art Documentation Journal of the Art Libraries Society of North America*, vol.28, No.2, pp.67-71, 2009.