

---

**DIGITAL LIBRARY MODELLING:  
SUPPORTING FOR KNOWLEDGE MANAGEMENT**

**Henderi<sup>1</sup>**  
**Ary Budi Warsito<sup>2</sup>**  
**Ahmad Sidik<sup>3</sup>**  
**Maimunah<sup>4</sup>**

[henderi@pribadiraharja.com](mailto:henderi@pribadiraharja.com), [aribudi@yahoo.com](mailto:aribudi@yahoo.com), [ahmadsidik@faculty.raharja.ac.id](mailto:ahmadsidik@faculty.raharja.ac.id),  
[maimunah@faculty.raharja.ac.id](mailto:maimunah@faculty.raharja.ac.id)

Diterima: 25 Maret 2010/Disetujui: 1 April 2010

**ABSTRACT**

*Increasing public demand for greater information to make the library's existence even more important. Besides being the source of information and reference various types of knowledge, the library also has a strategic role in the process of science. On the other hand, research results show that most libraries do not use information technology (IT) is optimal in knowledge processing. As a result of information and reference library possessed not easily obtained by the public to be developed. Therefore, this article discusses modeling digital library to support the implementation of knowledge management models, and facilitate the acquisition of information and development of library science. Digital libraries are made using modeling approaches unfield modeling language (UML), and described in the form: use case diagrams, class diagrams, state diagrams, collaboration diagrams, component diagrams, and screen prototype. The results of digital library implementations are also briefly discussed to provide a clearer picture.*

*Key words: knowledge management, digital library modeling, language modeling unfield*

**ABSTRAK**

*Peningkatan kebutuhan masyarakat terhadap informasi yang semakin tinggi membuat eksistensi perpustakaan semakin penting. Selain merupakan sumber informasi dan referensi berbagai jenis ilmu pengetahuan, perpustakaan juga memiliki peran strategis dalam proses*

- 
- 1. Dosen Jurusan Teknik Informatika, STMIK Raharja**  
Jl. Jend. Sudirman No. 40 Cikokol - Tangerang Telp. 5529692
  - 2. Mahasiswa Sistem Informasi, STMIK Raharja**  
Jl. Jend. Sudirman No. 40 Cikokol - Tangerang Telp. 5529692
  - 3. Dosen Jurusan Teknik Informatika, STMIK Raharja**  
Jl. Jend. Sudirman No. 40 Cikokol - Tangerang Telp. 5529692
  - 4. Dosen Manajemen Informatika, AMIK Raharja Informatika**  
Jl. Jend. Sudirman No. 40 Cikokol - Tangerang Telp. 5529692

*pengembangan ilmu pengetahuan. Pada sisi lain, hasil penelitian menunjukkan bahwa sebagian perpustakaan belum memanfaatkan information technology (IT) secara optimal dalam pengolahan pengetahuan. Akibatnya informasi dan referensi yang dimiliki perpustakaan tidak mudah diperoleh oleh publik untuk dikembangkan. Untuk itu, artikel ini membahas digital library modelling untuk mendukung penerapan model knowledge management, dan mempermudah perolehan informasi dan pengembangan ilmu pengetahuan dari perpustakaan. Digital library modelling dibuat menggunakan pendekatan unified modelling language (UML), dan digambarkan dalam bentuk: use case diagram, class diagram, state diagram, collaboration diagram, component diagram, dan screen prototype. Hasil implementasi digital library juga dibahas secara ringkas untuk memberikan gambaran lebih jelas.*

*Kata kunci: knowledge management, digital library modelling, unified modelling language*

## **PENDAHULUAN**

Perpustakaan merupakan sumber informasi dan referensi ilmu pengetahuan. Informasi tersebut berbentuk buku, jurnal, laporan penelitian, majalah, buletin, dan lainnya. Karenanya, pengembangan dan penemuan ilmu pengetahuan banyak berasal dari penggunaan informasi dan referensi dari perpustakaan. Perpustakaan juga mempunyai peran strategis dalam pengolahan ilmu pengetahuan dan mencerdaskan masyarakat dengan cara mempermudah publik untuk mendapatkan, menyaring dan menggunakan informasi dan referensi yang ada di perpustakaan. Sementara peningkatan jumlah koleksi informasi dan referensi yang disimpan di perpustakaan, dan tingginya kebutuhan terhadap informasi menimbulkan permasalahan baru. Informasi dan referensi yang tersimpan di perpustakaan tersebut tidak mudah ditemukan, disaring, diolah, dan digunakan oleh publik. Pada sisi lain, perkembangan *information technology* (IT) selama beberapa dekade terakhir telah mendorong perubahan paradigma sistem perpustakaan. Melalui penggunaan IT, informasi dan referensi ilmu pengetahuan tidak lagi terbatas pada buku atau sejenisnya yang tersimpan di rak-rak perpustakaan, tetapi disimpan bentuk *digital*, dan dikelola dalam *digital library*.

Dengan *digital library* berbagai informasi dan referensi ilmu pengetahuan akan lebih mudah dan cepat diakses, disaring, diolah, digunakan dan dikembangkan oleh publik tanpa dibatasi oleh ruang dan waktu. Kemampuan dan peran *digital library* ini sejalan dengan definisi tentang *knowledge management* yang disampaikan oleh Albert, S (1998), yaitu proses pengumpulan, pengelolaan, pengklasifikasian, dan penyebaran informasi kepada seluruh organisasi sehingga informasi tersebut dapat lebih dimanfaatkan oleh yang membutuhkannya.

Namun demikian, sebagian perpustakaan (di Indonesia) belum menerapkan dan memanfaatkan *digital library* dengan baik. Sementara jumlah informasi dan referensi yang ada di perpustakaan semakin meningkat dari waktu ke waktu. Informasi

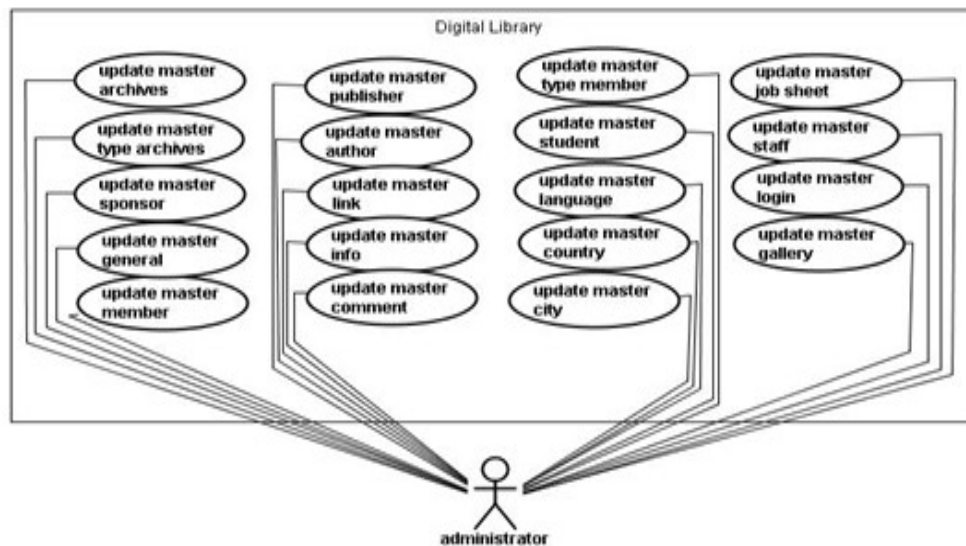
dan referensi tersebut tidak mudah diakses, disaring, diolah, digunakan, dan dikembangkan secara luas oleh masyarakat yang membutuhkannya. Untuk itu, artikel ini membahas *digital library modelling* untuk mendukung penerapan prinsip dan cara kerja *knowledge management* melalui pemanfaatan IT.

## PEMBAHASAN

### 1. Digital Library Modelling

#### A. Use Case Diagram

Diagram ini mendeskripsikan siapa yang akan menggunakan sistem, dan dalam cara apa *user* mengharapkan interaksi dengan sistem itu. Hal ini sejalan dengan pendapat Henderi (2008), yang menyatakan bahwa *use case diagram* dibuat untuk memodelkan interaksi antara sistem, sistem eksternal dan pengguna. Pada bagian lain, Miller Randy (2008), menyatakan bahwa *use case diagram* juga digunakan untuk menggambarkan sekuensial langkah-langkah dari setiap interaksi. Sejalan dengan definisi dan pendapat tersebut, maka Ary Budi Warsito et al (2010), membuat model *digital library* dalam bentuk *use case diagram* seperti pada gambar 1.

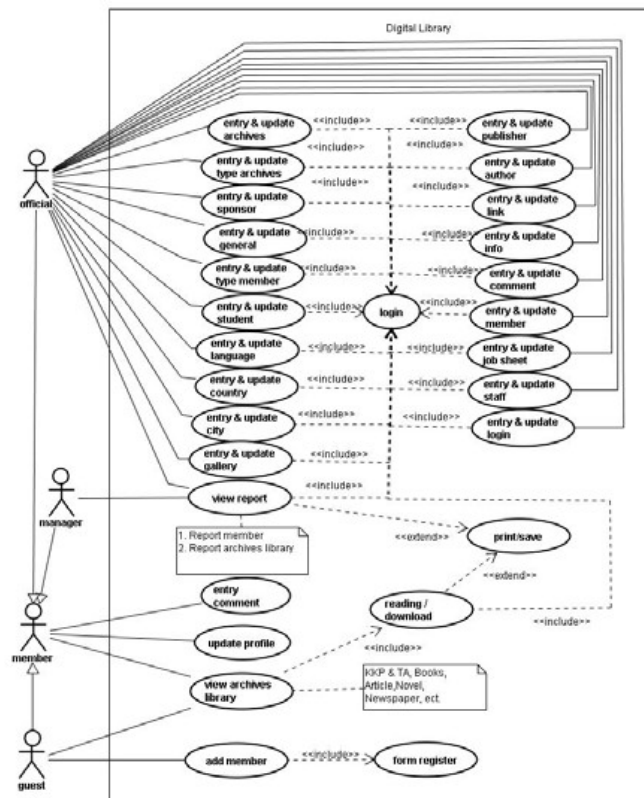


Gambar 1. Use case diagram digital library (domain administrator)

Berdasarkan gambar satu di atas, tampak bahwa salah satu aktor *digital library* adalah administrator. Aktor *administrator* bertugas melakukan 19 aksi terhadap objek. Hanya aktor administrator yang dapat melakukan aksi terhadap 19 aksi tersebut. Kegiatan aksi yang paling sering dilakukan oleh

administrator adalah merubah artikel dan katalog buku seiring dengan bertambahnya koleksi *digital library*.

Selain administrator, terdapat beberapa aktor lain yang berinteraksi dengan *digital library*. Interaksi dan kegiatan yang dilakukan oleh aktor tersebut tampak pada gambar 2 *use case diagram* berikut.



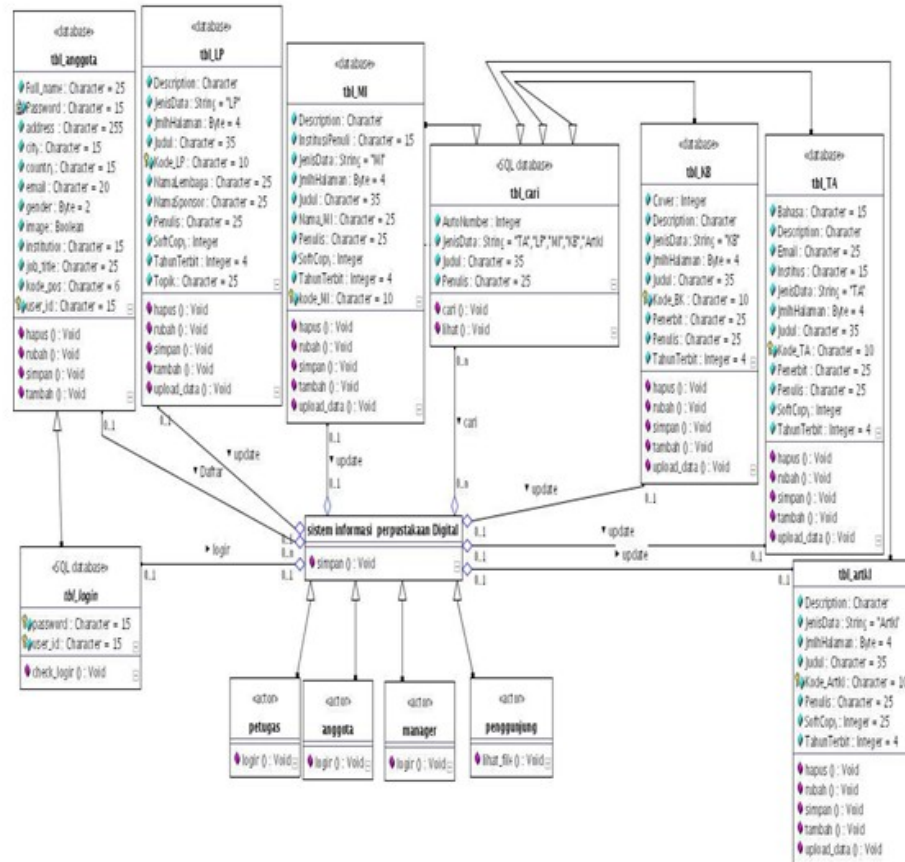
Gambar 2. Use case diagram digital library (domain user)

Berdasarkan gambar dua di atas diketahui bahwa selain aktor administrator aktor digital library juga terdiri dari *official*, *member*, *manager* dan *guest*. Sementara objek aktifitas mengikuti kegiatan, dan objek bisa diakses jika aktor sudah melakukan login. Aktor pada gambar satu dan dua dapat melakukan berbagai aktivitas (yang ditetapkan) tanpa dibatasi oleh ruang dan waktu. Dengan demikian digital library menjadi pemungkin dan pendukung penyebaran dan pengolahan ilmu pengetahuan (*knowledge management*) secara lebih baik.

## B. Class Diagram

Struktur statis dan *class-class* yang ada pada *digital library* digambarkan dalam bentuk *class diagram*. Struktur ini dibuat dengan memperhatikan pendapat Whitten L Jeffery et al (2004), yang menyatakan bahwa *class diagram* adalah gambar statis mengenai struktur objek statis dari suatu sistem, menunjukkan

*class-class* objek yang menyusun sebuah sistem, dan juga hubungan antara *class* objek tersebut. *Class diagram digital library* untuk mendukung penerapan *knowledge management* tampak pada gambar 3.



Gambar 3. *Class diagram digital library*

Berdasarkan gambar tiga, tampak bahwa struktur tabel pada *database* pada *digital library* digambarkan secara detail, mulai dari aktor yang mengakses data dan respon dari *object (class)* sebagai reaksi terhadap yang memberi aksi/sinyal. Berdasarkan *class diagram digital library* pada gambar tiga dapat juga dilihat adanya perubahan dengan *object* ketika terjadi suatu aksi.

### C. State Diagram

State diagram dibuat untuk menggambarkan perubahan antara objek pada sistem *digital library* sebagai akibat dari stimulasi yang diterima oleh sistem. Penggambaran state diagram digunakan setelah memperhatikan pendapat Miller Randy (2008), dan Henderi (2007) yang menyatakan bahwa *state diagram* juga menggambarkan *behavior* (perilaku) dinamis sistem dalam merespon stimulasi atau aksi yang berasal dari luar dan dibuat untuk menggambarkan



*modelling* dapat menjadi pemungkin dan mempercepat proses penyebaran ilmu pengetahuan (*knowledge management*) secara lebih luas.

#### D. Collaboration Diagram

Sementara itu, interaksi objek di dalam dan di sekitar pengguna *digital library* dalam bentuk *collaboration digram* tampak pada gambar 5 berikut.

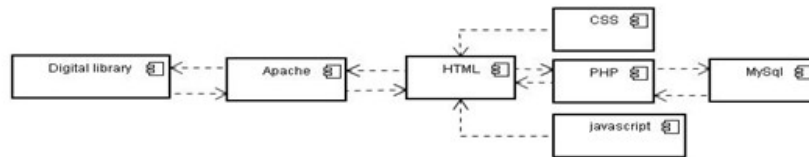


Gambar 5. Collaboration diagram digital library

Berdasarkan gambar lima, aktor yang berhak atau *object* berinteraksi dengan *digital library* dengan memberikan aksi/stimulus kepada sistem, dan selanjutnya sistem *digital library* memberikan respon kepada user yang memberikan stimulus. Aktor dapat memberikan aksi/stimulus dari mana saja dan kapan saja melalui koneksi internet, dan sistem akan memberikan respon secara real time kepada aktor yang memberikan stimulus. Karenanya, aktivitas *digital library* akan semakin meningkat seiring dengan semakin meningkatnya kebutuhan masyarakat terhadap informasi dan referensi, yang pada akhirnya juga dapat meningkatkan fungsi penerapan *knowledge management* untuk masyarakat secara luas.

## E. Componen Diagram

Teknologi (khususnya *software*) yang dibutuhkan dalam mengimplementasikan *digital library modelling* dalam bentuk *componen diagram* tampak pada gambar 6.



Gambar 6. *Component diagram digital library*

Gambar enam menjelaskan bahwa pengembangan *digital library* membutuhkan teknologi *software* (juga komponen *hardware* lainnya) untuk mendukung koneksi sistem *digital library* ke jaringan global (*internet*). Dalam implementasi *digital library modeling* ini, maka sistem mutlak harus terkoneksi dengan *internet* agar informasi dan referensi yang ada di *digital library* dapat diakses dengan mudah, cepat, dan secara luar oleh masyarakat. Dengan demikian, *digital library* sebagai salah satu penyedia dan penyimpan informasi dan pengetahuan (*information provider* dan *knowledge repository*) dapat mengantisipasi dan memenuhi tingginya kebutuhan masyarakat terhadap informasi dan referensi, serta menerapkan prinsip dan cara kerja *knowledge management*.

### 2. Data, Informasi dan Services yang dibutuhkan

Pada implementasinya, *digital librabry* membutuhkan dan menyimpan data, informasi, dan berbagai referensi ilmu pengetahuan dalam format digital sehingga mudah dan cepat diakses. Untuk itu, komponen dan *services* penting, dibutuhkan, dan direkomendasikan ada pada *digital librabry* adalah sebagai berikut :

- a. Fasilitas pencarian (*search*) fasilitas ini dapat di tingkatkan (menjad *advanced search*) agar pencarian data, informasi dan referensi dapat dilakukan dan didapatkan sesuai dengan permintaan aktor.
- b. Koleksi informasi dan referensi di digital library antara lain berbentuk *bibliographic*, karya ilmiah, katalog buku, *e-book*, *e-learning*, *material*, *course material*, *clipping*, multimedia, dan sejumlah artikel umum lainnya.
- c. Semua data, informasi, dan referensi ilmu pengetahuan yang dimiliki oleh digital library bisa ditampilkan dalam berbagai versi dan bentuk, misalnya dalam bentuk abstraksi saja, full teks, dan secara umum dapat ditampilkan secara detail sesuai dengan kebutuhan dan pengaturan hak akses.

### 3. Implementasi Digital Library

Secara umum halaman website *digital library* terdiri dari dua *state* fungsional, yaitu *state* untuk administrator, dan *state* untuk publik. Berikut penjelasan hasil uji coba implementasi *digital library* pada penelitian ini.

#### a. Halaman Website Digital Library untuk Administrator

Melalui halaman ini administrator *digital library* dapat melakukan berbagai kegiatan pengelolaan. Diantaranya: menambah, merubah, dan menghapus data, informasi dan referensi yang ada di *digital library*. *Screen* utama halaman *web* untuk administrator ada pada gambar 7.



Gambar 7. Homepage digital library atau administrator

Berdasarkan gambar tujuh, *homepage digital library* menyediakan tujuh *services* utama untuk administrator yaitu: artikel, buku, informasi, sponsor, anggota, gallery, kumpulan website, komentar, dan pengaturan umum. Homepage *digital library* juga menampilkan informasi tentang jumlah artikel, buku, informasi, anggota, gallery, link, dan komentar yang disimpan oleh sistem. Untuk memudahkan tugas administrator dalam melakukan *update*, sistem *digital library* juga dilengkapi dengan *features update*. *Features* ini berfungsi untuk melakukan *update* terhadap data, informasi dan referensi yang ada dalam *digital library*. Bentuk *screen dialog services update digital library* yang dimaksud tampak pada gambar 8.



Gambar 8. Screen dialog update data/informasi pada digital library

Berdasarkan gambar delapan, *digital library* (yang memiliki berbagai fasilitas tersebut) dapat memungkinkan administrator memberikan dukungan terhadap penerapan *knowledge management* secara lebih efektif, efisien, luas, dan *bordless*.

#### b. Halaman Website Digital Library untuk Layanan Publik

Melalui halaman yang ditunjukkan pada gambar 9, pengunjung dapat melihat semua informasi dan referensi yang ada di *digital library*, bahkan bagi pengunjung yang sudah menjadi *member* disediakan layanan untuk melakukan *download*.



Gambar 9. Homepage website digital library untuk layanan public

Berdasarkan gambar sembilan diketahui bahwa implementasi *digital library modeling* dapat mendukung penerapan *knowledge management*, karena sistem *digital library* dapat dijadikan sebagai media penyebaran data, informasi, referensi dan ilmu pengetahuan kepada publik yang membutuhkannya tanpa dibatasi oleh ruang dan waktu. Selain dapat menemukan dan mengakses data, informasi, dan referensi yang ada di *digital library*, publik yang menjadi anggota *digital library* juga dapat melakukan *download* data, informasi dan referensi sesuai dengan kebutuhannya secara cepat dan *bordless*.

## SIMPULAN

*Digital library modelling* pada penelitian ini dibuat untuk menjawab tantangan peningkatan kebutuhan publik terhadap informasi yang semakin tinggi, cepat, dan bermutu. *Digital library modelling* ini juga dibuat untuk memberikan dukungan terhadap penerapan prinsip dan cara kerja *knowledge management* secara lebih luas, dan mengoptimalkan peran strategis perpustakaan dalam proses pengembangan ilmu pengetahuan. *Digital library modelling* ini yang disusun dalam bentuk *component diagram*, *use case diagram*, *class diagram*, *state diagram*, dan

*collaboration diagram* menggunakan pendekatan *unifield modelling language* (UML). Penerapan *digital library* dapat memperluas dan mempermudah penyebaran data, informasi dan referensi berbagai ilmu pengetahuan kepada publik tanpa dibatasi oleh ruang dan waktu. Dengan demikian *digital library* dapat mendukung penerapan *knowledge management* secara luas kepada publik secara *bordless*.

#### PUSTAKA

Albert, S. 1998. Knowledge Management: Living Up To The Hype? "Midrange System," 11 (13), Sep 7, page 52.

Ary Budi Warsito, Henderi, Ahmad Sidik, 2010. Desain dan Implementasi Sistem Perpustakaan Digital Menggunakan UML untuk Meningkatkan Pelayanan Perpustakaan, Laporan Penelitian Skripsi dibawah Supervisi Henderi dan Ahmad Sidik, STMIK Raharja, Tangerang

Henderi, 2008. Unified Modeling Language (UML): Konsep dan Implementasinya Pada Pemodelan Sistem Berorientasi Objek dan Visual (Buku I), STMIK Raharja – Tangerang.

\_\_\_\_\_, 2007, Analysis and Design System with Unified Modeling Language (UML), STMIK Raharja, Tangerang

Miller Randy, 2008. Practical UML: A Hands-On Introduction for Developers. [www.Sequence-Diagram-Hands-On-Introduction-for-Developers.mht](http://www.Sequence-Diagram-Hands-On-Introduction-for-Developers.mht), diakses pada tanggal 14 Mei 2008.

Whitten L Jeffery, Bentley D. Lonnie, Dittman C. Kevin, 2004. Metode Desain dan Analisis Sistem, Edisi 6., Edisi Internasional Mc.Graw Hill Education dan Penerbit Andi Jogjakarta

## EDUCATIONAL RESOURCE SHARING IN THE HETEROGENEOUS ENVIRONMENTS USING DATA GRID

Aan Kurniawan<sup>1</sup>  
Zainal A. Hasibuan<sup>2</sup>

aan.kurniawan@ui.ac.id<sup>1</sup>, zhasibua@cs.ui.ac.id<sup>2</sup>

Diterima: 3 Maret 2010/Disetujui: 10 Maret 2010

### ABSTRACT

*Educational resources usually reside in the digital library, e-learning and e-laboratory systems. Many of the systems have been developed using different technologies, platforms, protocols and architectures. These systems maintain a large number of digital objects that are stored in many different storage systems and data formats with differences in: schema, access rights, metadata attributes, and ontologies. This study proposes a generic architecture for sharing educational resources in the heterogeneous environments using data grid. The architecture is designed based on the two common types of data: structured and unstructured data. This architecture will improve the accessibility, integration and management of those educational resources.*

Keywords: resource sharing, data grid, digital library

### INTRODUCTION

Currently, the increasing social demands on high quality educational resources of higher education cannot be fulfilled only by the available educators and conventional libraries. With the advances of information technology, many learning materials and academic journals created by universities have been converted into digital objects. The rapid growth of Internet infrastructure accelerates the transformation of conventional libraries and learning to the digital libraries and e-learning. This transformation greatly affects the way of people to get information and learn. Accessing information and learning now can be done from anywhere at any time.

Since many digital library, e-learning and e-laboratory systems have been developed using different technologies, platforms, protocols and architectures, they will potentially introduce the problem of information islands. In order to address this problem, some

1. Student Of Doctor Faculty Of Computer Science, University Of Indonesia  
Faculty Of Computer Science UI, Depok 16424 Indonesia
2. Lecture Of Doctor Faculty Of Computer Science, University Of Indonesia  
Faculty Of Computer Science UI, Depok 16424 Indonesia

previous works [1][2][3][4] proposed the use of grid technology that has the capability of integrating the heterogeneous platforms. However, most of them considered that the shared resources are only files or unstructured data. Educational resources consist of not only unstructured data, but also structured data. Much information such as the metadata describing the shared digital objects and the XML formatted documents is stored in a database. This information also needs to be shared with other systems. In this study, we propose a generic architecture for sharing educational resources in heterogeneous environment using data grid. We also show how this architecture applies in the digital libraries using Indonesian Higher Education Network (INHERENT) [5].

## DISCUSSION

### INHERENT

INHERENT (Indonesian Higher Education Network) [5] is a network backbone that is developed by Indonesian government to facilitate the interconnection among the higher education institutions (HEIs) in Indonesia. The project was proposed by the directorate of higher education. Started on July 2006, currently it connects 82 state HEIs, 12 regional offices of the coordination of private HEIs, and 150 private HEIs (see Figure 1).

All state HEIs in Java are connected by STM-1 National Backbone with the bandwidth of 155 Mbps. Other cities in the other islands use 8-Mbps leased line and 2 Mbps VSAT connections.

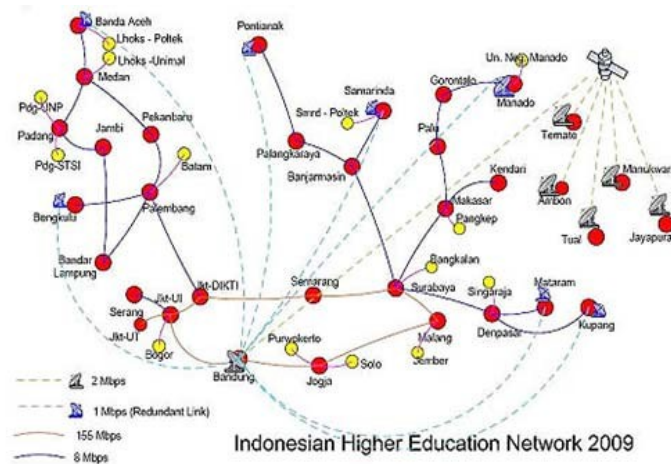


Figure 1. Indonesian Higher Education Network in 2009 [5]

This network has been used for various educational activities including video-conferencing and distance learning. Every university can build their own digital libraries and learning management systems (LMSs) and then publish their educational resources through the network. Although the resources can be shared to each another via FTP

or web servers, the systems (digital libraries and LMSs) cannot provide an integrated view to users. Users still have to access every digital library systems in order to find the resources required by them. This network has a potential of sharing various educational resources using data grid.

## **DATA GRID**

Data grid is one of the types of grid technologies. The other types are computational and access grid. Originally, the emphasis of grid technology lay in the sharing of computational resources [6]. Technological and scientific advances have led to an ongoing data explosion in many fields. Data are stored in many different storage systems and data formats with different schema, access rights, metadata attributes, and ontologies. These data also need to be shared and managed. The need then introduces a new grid technology, namely data grid. There are some existing data grids. In the following, we will overview two of them (iRODS and OGSA-DAI) and highlight their features to address the need.

### **– iRODS**

iRODS (Integrated Rule-Oriented Data System) [7] is a second generation data grid system providing a unified view and seamless access to distributed digital objects across a wide area network. It is extended from the Storage Resource Broker (SRB) that is considered as the first generation data grid system. Both SRB and iRODS are developed by the San Diego Supercomputing Center (SDSC).

Classified as the first generation of the data grid, SRB is mainly focused on providing a unified view over distributed storages based on logical naming concepts using the client-server architecture. The concepts facilitated the naming and location transparency where users, resources, data objects and virtual directories were abstracted by logical names and mapped onto physical entities. The mapping is done at run time by the Virtualization sub-system. The information of the mapping from the logical name to physical name is maintained persistently in a database system called the Metadata Catalog. The database also maintains the metadata of the data objects that use the schema of attribute-value pair and the states of data and operations. Built upon this logical abstraction, iRODS takes one level higher by abstracting the data management process itself called policy abstraction.

Whilst the policies used for managing the data at the server level in SRB are hard-coded, iRODS uses another approach, Rule-oriented Programming (ROP), to make the customization of data management functionalities much easier. Rules are explicitly declared to control the operations performed when a rule is invoked by a particular task. In iRODS, these operations are called micro services and implemented as functions in C programming language.

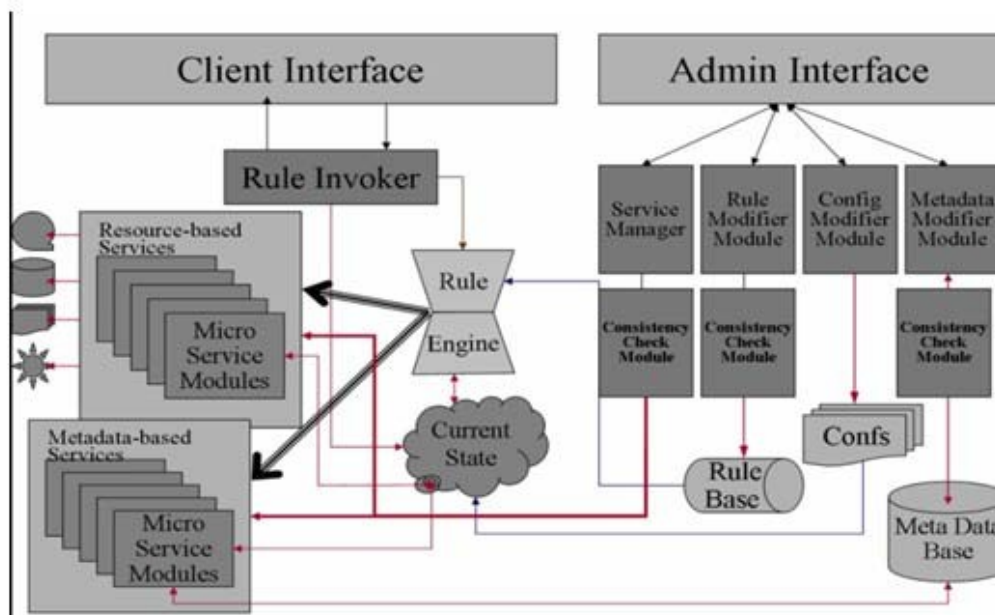


Figure 2. iRODS Architecture [7]

Figure 2 displays the iRODS architecture with its main modules. The architecture differentiates between the administrative commands needed to manage the rules, and the rules that invoke data management modules. When a user invokes a service, it fires a rule that uses the information from the rule base, status, and metadata catalog to invoke micro-services [8]. The micro-services either change the metadata catalog or change the resource (read/write/create/etc).

Figure 3 illustrates a scenario when a client sends a query asking for a file from an iRODS zone. Firstly, he connects to one of iRODS servers (for example server A) using a client application and sends the criteria of the file needed (e.g. based on the metadata, filename, size, etc). The request is directed to server A that will find the file using information available in Metadata catalog. The query result is sent back to the client. If he/she wants to get the file, server A asks the catalog server which iRODS server that stores the file (for example in the server B). Server A then communicates with server B to request the file. Server B applies the rules related with the request. The rules can be the process of authorization (whether the client has a privilege to read the file) and sending the file to the client using iRODS native protocol. The client is not aware of the location of the file. This location transparency is handled by the grid.

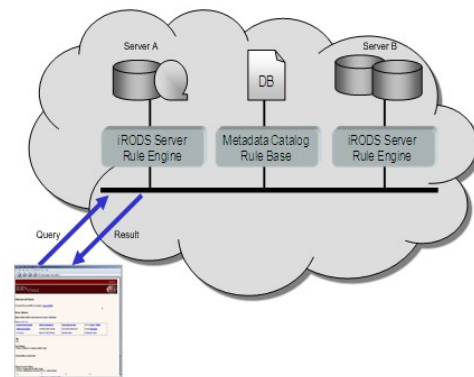


Figure 3. A client asks for a file from an iRODS data grid [7]

Based on the explanation above, we conclude that iRODS focuses on managing unstructured data objects such as files. Although it can also access structured data resources, its orientation is mainly on distributed file management. However, it also uses structure data (relational database) to manage the metadata of the data objects, the states of data and the states of operations. The metadata can potentially be integrated using the OGSA-DAI data grid.

#### – OGSA-DAI

OGSA-DAI (Open Grid Services Architecture – Data Access and Integration) [9] is a middleware software that allows structured data resources, such as relational or XML databases, from multiple, distributed, heterogeneous and autonomously managed data sources to be easily accessed via web services. It focuses on cases where the assembly of all the data into a single data warehouse is inappropriate [9].

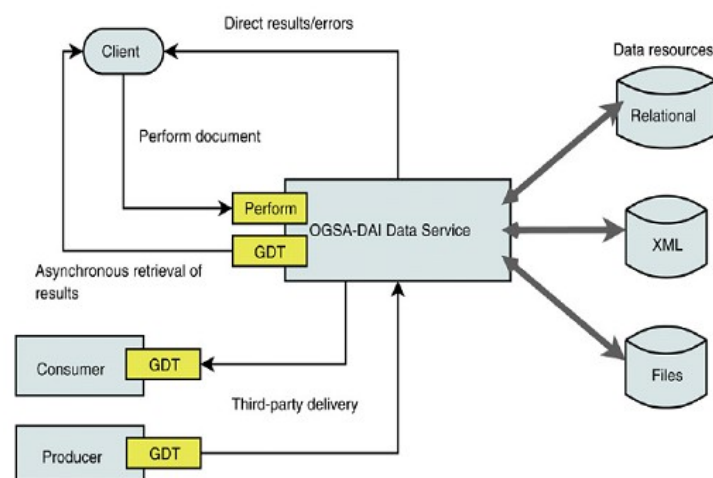


Figure 4. An overview of OGSA-DAI components [10]

OGSA-DAI is designed to enable sharing of data resources to make collaboration that supports:

- a. Data access service, which allows to access structured data in distributed heterogeneous data resources
- b. Data transformation service, which allows to expose data in schema X to users as data in schema Y
- c. Data integration service, which allows to expose multiple databases to users as a single virtual database
- d. Data delivery service, which allows to deliver data to where it's needed by the most appropriate means, such as Web service, email, HTTP, FTP and GridFTP

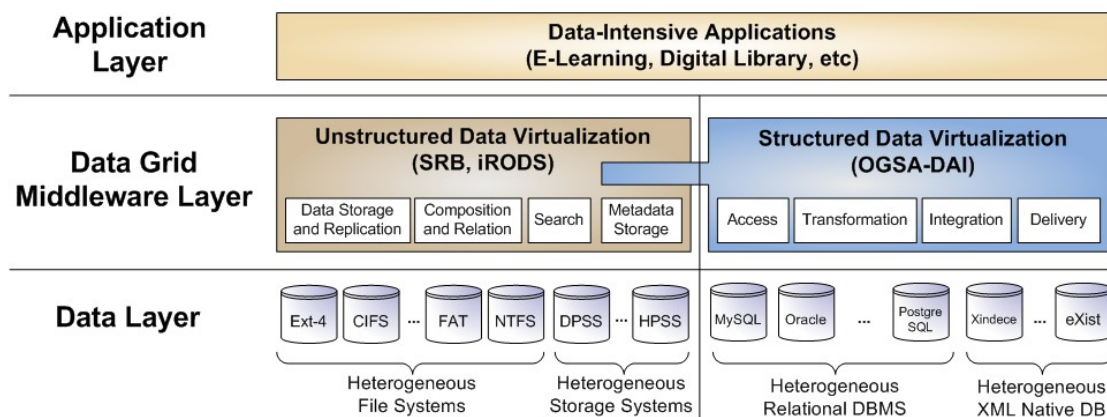


Figure 5. The proposed generic architecture for resource sharing

OGSA-DAI has adopted a service oriented architecture (SOA) solution for integrating data and grids through the use of web services. The role of OGSA-DAI in a service-based Grid, illustrated in Figure 4, involves interactions between several following components [10]:

- a. OGSA-DAI data service: a web service that implements various port types allowing the submission of requests and data transport operations
- b. Client: an entity that submits a request to the OGSA-DAI data service. A request is in the form of a *perform* document that describes one or more activities to be carried out by the service.
- c. Consumer: a process, other than the client, to which an OGSA-DAI service delivers data.
- d. Producer: a process, other than the client, that sends data to an OGSA-DAI data service.

When a client wants to make a request to an OGSA-DAI data service, it invokes a web service operation on the data service using a perform document. A perform document is an XML document describing the request that the client wants to be executed, defined by linking together a sequence of activities. An activity is an OGSA-DAI construct corresponding to a specific task that should be performed. The output of one activity can be linked to the input of another to perform a number of tasks in sequence. A range of activities is supported by OGSA-DAI, falling into the broad categories of relational activities, XML activities, delivery activities, transformation activities and file activities. Furthermore, the activity is an OGSA-DAI extensibility point, allowing third parties to define new activities and add them to the ones supported by an OGSA-DAI data service.

OGSA-DAI focuses on managing heterogeneous structured data resources. Although it can also access the unstructured data using file transfer, its orientation is mainly on structured data (such as relational and XML database) management.

According to the two existing data grid middlewares, there are two main orientations in handling data based on the nature of data: structured and unstructured data. In the following section, we propose a system architecture that accommodates these two kinds of data.

## **THE PROPOSED ARCHITECTURE**

In this study, a generic architecture for sharing educational resources in heterogeneous environment using data grid middleware is proposed based on the two types of data: structured and unstructured data.

From the perspective of computer processing, the digital objects are merely data. Generally, data can be classified into two categories: unstructured and structured data. Unstructured data consists of any data stored in an unstructured format at an atomic level. There is no conceptual definition and no data type definition in the unstructured content. Furthermore, unstructured data can be divided into two basic categories: bitmap objects (such as video, image, and audio files) and textual objects (such as spreadsheets, presentations, documents, and email). Both of them can be treated as a string of bits. The unstructured data is usually managed by operating system. Structured data has schema information that describes its structure. The schema can be separated from the data (such as in relational database) or it can be mixed with the data (e.g. XML format). The structured data is usually managed by a database management system. The system facilitates the processes of defining, constructing, manipulating, and sharing the data among various users and applications [11]. This differentiation of the nature of data brings into different treatment when the various formats of data and storage systems are handled in the middleware layer.

Figure 5 shows the proposed architecture that utilizes the similar hierarchy used in [12] but it is applied in managing heterogeneous data resources. The architecture consists of three layers: data layer, data grid middleware layer and application layer. At the data layer, the various data resources in the heterogeneous file systems and storage systems can be joined into one large data collection. We distinguish between structured and unstructured data because of their different inherent characteristics. At the data grid middleware layer, the data virtualizations for each data type are separated. The unstructured data are virtualized by file-oriented data grid middleware, such as SRB and iRODS, while the structured data virtualization is handled by database-oriented data grid middleware, such as OGSA-DAI.

Based on the analysis of the file-orientated data grids (such as SRB and iRODS), the unstructured data virtualization provides the following basic services [13]:

- a. Data storage and replication service, which allows to store any type of digital object content and to replicate it into several other resources. The service is independent of the content type because only the clients need to be aware of the content internal format and structure.
- b. Composition and relation service, which allows to define various relations between digital objects and to define multiple groups of related objects. Those relations may be used to create complex digital objects, to show parent/child relationships between objects or to create collections of digital objects.
- c. Search service, which allows to search in previously defined sets of digital objects. The search can be based on the query matching with the metadata catalog.
- d. Metadata storage service, which allows to store metadata describing digital objects. One object can be described in many metadata records. The metadata also records information associated with replication. These records can be utilized for the search service. Usually, database systems are used to store and manage the metadata. Furthermore, some database systems containing metadata can be integrated using structured data virtualization components of the data grid middleware.

The structured data virtualization provides the four basic services as described in the section of OGSA-DAI.

At the application layer, data-intensive applications, such as e-learning management systems and digital libraries, can utilize the two data virtualizations in order to publish and share their digital content objects.

Figure 6 shows a typical implementation of the proposed architecture for digital library. Every digital library sites that are registered in the Integrated Higher Education Digital Library Portal manage their own data resources consisting of the collection of digital

objects and structured data (relational database and XML). The digital objects are stored in various storage systems that are managed by iRODS storage servers. Since all of iRODS servers are registered in one zone, namely Zone IDL (Indonesian Digital Libraries), the digital objects can be replicated among the servers. Some files of site A can be replicated to the servers of site B, and vice versa. The metadata catalog servers in both sites will contain the same information of all collected educational resources.

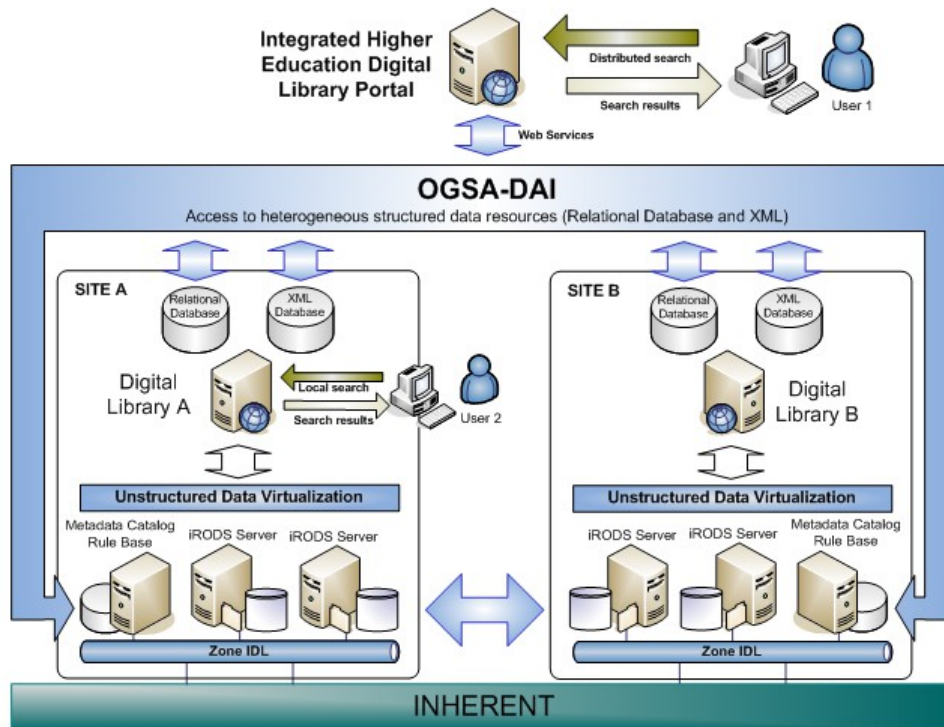


Figure 6. A typical implementation of the proposed architecture for digital library

Some digital library systems store index files for the use of searching in relational databases. The systems can also manage some kinds of educational resources formatted in XML (e.g. semi-structured documents) using native XML databases. All of this information can be accessed and integrated by the Integrated Higher Education Digital Library Portal using OGSA-DAI. Therefore, a user can do distributed searching for files stored in all resources of both sites. The integrated system also enables a user to get all resources closer to him if the resources are already replicated to some locations.

Since all sites are connected in INHERENT with high-speed bandwidth, there is no need for the Integrated Higher Education Digital Library Portal to harvest the metadata from all member sites such as proposed in [4]. No central metadata repository is

required. This ensures that the query results of distributed searching will always be up to date because they come from the local query processing of each member sites.

## CONCLUSION

In this study, we propose a generic architecture for sharing educational resources in the heterogeneous environment. The architecture distinguishes the managed data into two categories, namely structured and unstructured data. The data grid middleware used for virtualization is separated based on the two categories of data. In our design, we use iRODS and OGSA-DAI as the data grid middleware. iRODS is mainly file-oriented data grid and uses its own protocols. OGSA-DAI is mainly database-oriented and can only be accessed through web service mechanism. The combination of the two data grids completely handles all kinds of data types. Hence, this architecture can improve the accessibility, integration and management of those educational resources.

## REFERENCES

- [1] Yang, C.T., Hsin-Chuan Ho. *Using Data Grid Technologies to Construct a Digital Library Environment*. Proceedings of the 3rd International Conference on Information Technology: Research and Education (ITRE 05), pp. 388-392, NTHU, Hsinchu, Taiwan, June 27-30, 2005. (EI)
- [2] Candela, L., Donatella Castelli, Pasquale Pagano, Manuele Simi. *Moving Digital Library Service Systems to the Grid*. Springer-Verlag. 2005.
- [3] Sebestyen-Pal, G., Doina Banciu, Tunde Balint, Bogdan Moscaiuc, Agnes Sebeten-Pal. *Towards a GRID-based Digital Library Management System*. In *Distributed and Parallel Systems* p77-90. Springer-Verlag. 2008.
- [4] Pan, H. *Research on the Interoperability Architecture of the Digital Library Grid*. 2007. in IFIP International Federation for Information Processing, Volume 251, Integration and Innovation Orient to E-Society Volume I, Wang, W. (Eds), (Boston: Springer), pp. 147-154.
- [5] Indonesian Higher Education Network (INHERENT). <http://www.inherent-dikti.net>
- [6] Foster, I., Carl Kesselman. *The Grid: Blueprint for a New Computing Infrastructure*. 2<sup>nd</sup> Edition. Morgan Kaufmann. 2006.

- 
- [7] iRODS (Integrated Rule-Oriented Data System. <https://www.irods.org>)
- [8] Weise, A., Mike Wan, Wayne Schroeder, Adil Hasan. *Managing Groups of Files in a Rule Oriented Data Management System (iRODS)*. Proceedings of the 8<sup>th</sup> International Conference on Computational Science, Section: Workshop on Software Engineering for Large-Scale Computing, Krakow, Poland. 2008.
- [9] OGSA-DAI (Open Grid Services Architecture - Data Access and Integration). <http://www.ogsadai.org.uk/index.php>
- [10] Chue Hong, N.P., Antonioletti, M., Karasavvas, K.A. and Atkinson, M. *Accessing Data in Grids Using OGSA-DAI*, in Knowledge and Data Management in GRIDS, p3-18, D. Talia, A. Bilas, M.D. Dikaiakos (Eds.), 2007, ISBN: 978-0-387-37830-5
- [11] Elmashri, R., Shamkant B. Navathe. *Fundamental of Database Systems*. 5<sup>th</sup> Edition. Addison Wesley. 2006.
- [12] Coulouris, G., Jean Dollimore, Tim Kindberg. *Distributed Systems : Concepts and Design*. 4<sup>th</sup> edition. Addison Wesley. 2005.
- [13] Kosiedowski, M., Mazurek, C., Stroinski, M., Werla, M. and Wolski, M. *Federating Digital Library Services for Advanced Applications in Science and Education*, Computational Methods in Science and Technology 13(2), pp. 101-112. December, 2007