Grid Services for Digital Archive

Wei-Long, Ueng
Academia Sinica Computing Centre
Content

- Introduction to Digital Archive
- Demands from Digital Archive
- DAGS Infrastructure and Architecture
- DORE Grid Service
- Geospatial Grid Service
- Summary
Definition of Digital Archive

- Digital Archive is a collection of digital objects.

- A digital object is defined as something (e.g., an image, an audio recording, a text document, a movie, a map) that has been *digitally encoded and integrated with metadata* to support discovery, use, and storage of those objects.
Goals for Digital Archive (functional point of view)

- Protection of the original
- Duplication for longevity
- Flexible Search and Retrieval
- Easy Access
- Resource Sharing
- Lower cost of maintenance and dissemination
- Max. flexibility for integration of heterogeneous/homogeneous information resources
- Providing abundant resources for knowledge discovery and knowledge construction
Important Issues of NDAP

- Intellectual property rights
- Time, Space and Language Coordination
- Multi-lingual issues
- Public information systems

- Meta-language and Documentation
  - *Metadata*
  - *Content Markup*
  - *References and Linking*

- Dissemination and Sharing
- Cooperation and collaboration
- Scalability, Adaptability and Durability
Demands of Digital Archive

- Persistent digital objects,
- Well-organized information structure for effective content management
- Efficient and accurate information retrieval mechanism
- Flexible services for variant users needs
- Consistency
- Integrate relationship between information management and data management
- High-performance remote data access
- Authentication and authorization
- Resource discovery and monitoring
What we shall supply

- Reliable and efficient storage system
  - Reliable replication system -> replica locating mechanism
  - Reduced query latency -> query routing scheme
  - Load sharing
  - Robust, high availability
  - Min. Access latency
  - Manageability
  - High Throughput
  - Adaptive
  - Transparency of location and protocol
Challenge

- Big Challenge of IT for cataloging, searching, retrieval, management, identification, knowledge discovery, and integration
- Integration and Retrieval of Information Resources
Approach

- Develop Grid Services that can integrate heterogeneous metadata systems, distributed database management systems and geospatial information systems.

- Provide a framework to exchange different metadata XML documents (EAD, DC, FGDC ...) in “National Digital Archives Program”.
Building Grid Service for DORE

- DORE (Document REtrieval) is:
  - A middleware
  - A library
  - A tool
  - for programmers to develop metadata database applications

- DORE is a tool in Open Digital Archive Environment (ODAE).

- Migrate DORE applications to Grid enabled, and also have backward compatibility to existing system.
The UML of DORE Grid Service

Client

Create

DoreService

getDocument

metadata (xml)

getObject

Object Location

destroy

FileService

logical name

physical location

RLS(Cmd)

edg-rm
Dore Grid Service
GUI Client
Next Steps of DORE Grid

- Rebuild DORE to become a Grid service
- Security Issue
  - Add CA authority in framework and achieve inter-organizational data sharing.
  - Security management.
- Deploy DORE Grid Service in organizations
- Other organization could build their own client application to use this framework
- Every organization has deployed full functional DORE Grid systems.
- Data sharing between organizations.
Geospatial Grid Service

- Three basic categories of GIS Grid Services:
  - Data Services
  - Processing Services
  - Catalog Services
Services Architecture

Approach

For Example...
Find a historical map?
Find place names in Qing Dynasty?

Users

Applications
e.g.
Historical research planning,
Administrative boundary Changes
Create map

Other Applications

Services
e.g., Metadata Service,
Gazetteer service,
Web Map Service

Data
e.g., topographic,
thematic, imagery,
toponomy, metadata

uses
Metadata Service,
Gazetteer Service,
Web map service

based on
Base historical maps,
Geographical Names,
Map features

For Example...
Find a historical map?
Find place names in Qing Dynasty?

Uses
Metadata Service,
Gazetteer Service,
Web map service

based on
Base historical maps,
Geographical Names,
Map features

For Example...
Find a historical map?
Find place names in Qing Dynasty?

Uses
Metadata Service,
Gazetteer Service,
Web map service

based on
Base historical maps,
Geographical Names,
Map features
**Metadata Service for Geospatial Data**

### 識別資訊:
- 資料來源: 臺灣考古遺址查詢系統
- 資料類型: Live Data and Maps (ArcIMS Image Service)
- 出版日期: 2003
- 相關連結: [http://webgis.sinica.edu.tw/twnarchaedsites/view](http://webgis.sinica.edu.tw/twnarchaedsites/view)

### 資料描述:
- 資料類型: 中原文化考古遺址分期資料系統
- 資料來源: 国立中央研究院歷史語言研究所

### 資料狀態:
- 編制者: 已完成系統開發，目前提供實驗使用
- 更新頻率: 疊年進行實驗性更新

### 關鍵字(例如: taiwan, town):
- 中原關鍵字: 臺灣考古, 臺灣, Taiwan

### 存放及使用資料:
- 存放限制: 限制
- 使用限制: 限制

---

**Legend**
- [cover]
- [圖]
DAGS is evolving to be an interoperable network of databases and information technology tools using **Web services** and **Grid technologies**.

- In the near term, DAP will provide a national metadata registry of the available data with open interfaces through Grid service.

- Building on the contents of this registry, DAGS will provide its own central portal that enables simultaneous queries against different databases held by distributed, even worldwide sources.

- In the long term, different level objects can be linked to the system.

- These will facilitate and enable data mining of unprecedented utility and e-Science.
Summary

- **Achievements**
  1. The Grid services cooperate with Geospatial Information system was developed and tested.
  2. The DORE Grid middleware was implemented and rebuilt.
  3. The metadata register of different provider and databases were completed.

- **Future Works**
  1. Keeping to Integrate heterogeneous databases.
  2. Refining the technologies of Data Grid
  3. Developing the knowledge and e-Science discovery
Thank you for your attend!!