Digital Preservation Activities in National Library of China (NLC)

Li Chunming
National library of China

Sep. 16, 2005  International Conference on Preservation of Digital Objects
Outline

➤ **strategic Aim of NLC**

➤ **Research in Progress**

➤ **Current Activities**

➤ **Others**
strategic Aim of NLC

✧ Maximize collection of Chinese digital heritage

✧ Selected collection of foreign academic resources

✧ Resolution under international standards

✧ Utilization of advanced IT technologies

✧ Enhancement of multilateral cooperation
National Digital Library of China Project

(National Library of China Phase II will be completed in 2007)
Research in Progress
OAIS and Software

SIP = Submission Information Package
AIP = Archival Information Package
DIP = Dissemination Information Package
Metadata schemes

MARC, DC, etc. Comprised in METS

Search systems

Compliant with OAI, Z39.50, X-server
Preservation strategies

- Secure storage
- Backing up
- Specification of naked data structure
- Normalization
Current Activities

- Obtain policy support and Raise public awareness
- DiNeR: Digital Newspaper Repository
- Progress on WI CP project
Obtain policy support and public recognition

Ask the government for policy support on permanent digital resource preservation

Let the public realize the significance of digital resource preservation with multiple methods
Digital Newspaper Repository

Newspapers are a major publishing source.

Accessibility to these collections is very limited.

The traditional search methods are too slow and difficult.

The digital collections of large daily newspapers made the access to newspapers more effective and efficient.
What kind of titles were selected?

2,119 newspapers in total

213 national newspapers

766 provincial newspapers

898 prefecture and city newspapers

242 county newspapers

Websites of Newspapers 1,800

Focused on 80 websites with PDF files
Intellectual property processing

Chinese law stipulates that the National Library is entitled to get donation of e-publication.

Contacting for authorization to obtain the right.
Capturing, collecting and service

- The capturing tool: Flashget
- Collecting
- Metadata Generation
- Object organization
- Access service
Future Development

- Content Enhancement
- Content Management
- Opening to the public
Progress on WICP project

Surface Web

Deep Web

WICP Project

Robot Collection
Processing → Cataloging

National Bibliography

Mirror Archive

Subject Archive

Legal Deposit
Processing → Cataloging

National Archive
Partnerships

- TianWang, Peking University
- Tsinghua Tongfang
- BeiDa Founder
- WanFang Data
中国Web信息博物馆

Web InfoMall

“中国Web信息博物馆”是在国家973和985项目支持下，北京大学网络实验室开发建设的中国网页历史信息服务与展示系统。目前已经维护有10亿中文网页，每月新增千万页的网页，为研究者提供了一个信息存储和研究的平台。

Web InfoMall简介

- 输入URL，浏览历史网页，回顾网页变迁
- 提供网页全文检索
- 汇集历史网页，再现大千变化，包含万千搜索信息

Web InfoMall服务

- Web InfoMall简介
- Web InfoMall服务
- 共有自2001年以来1,017,104,240网页在线（17.106GB）
Thank you very much for your attention.

My Email: licm@nlc.gov.cn

National Library of China
http://www.nlc.gov.cn