



# Digital Preservation Activities in National Library of China(NLC)

Li Chunming  
National library of China

1

Sep. 16, 2005

International Conference on Preservation of Digital Objects

GÖTTINGEN



# Outline

- **strategic Aim of NLC**
- **Research in Progress**
- **Current Activities**
- **Others**

# National Library of China(NLC)



3

Sep. 16, 2005

International Conference on Preservation of Digital Objects

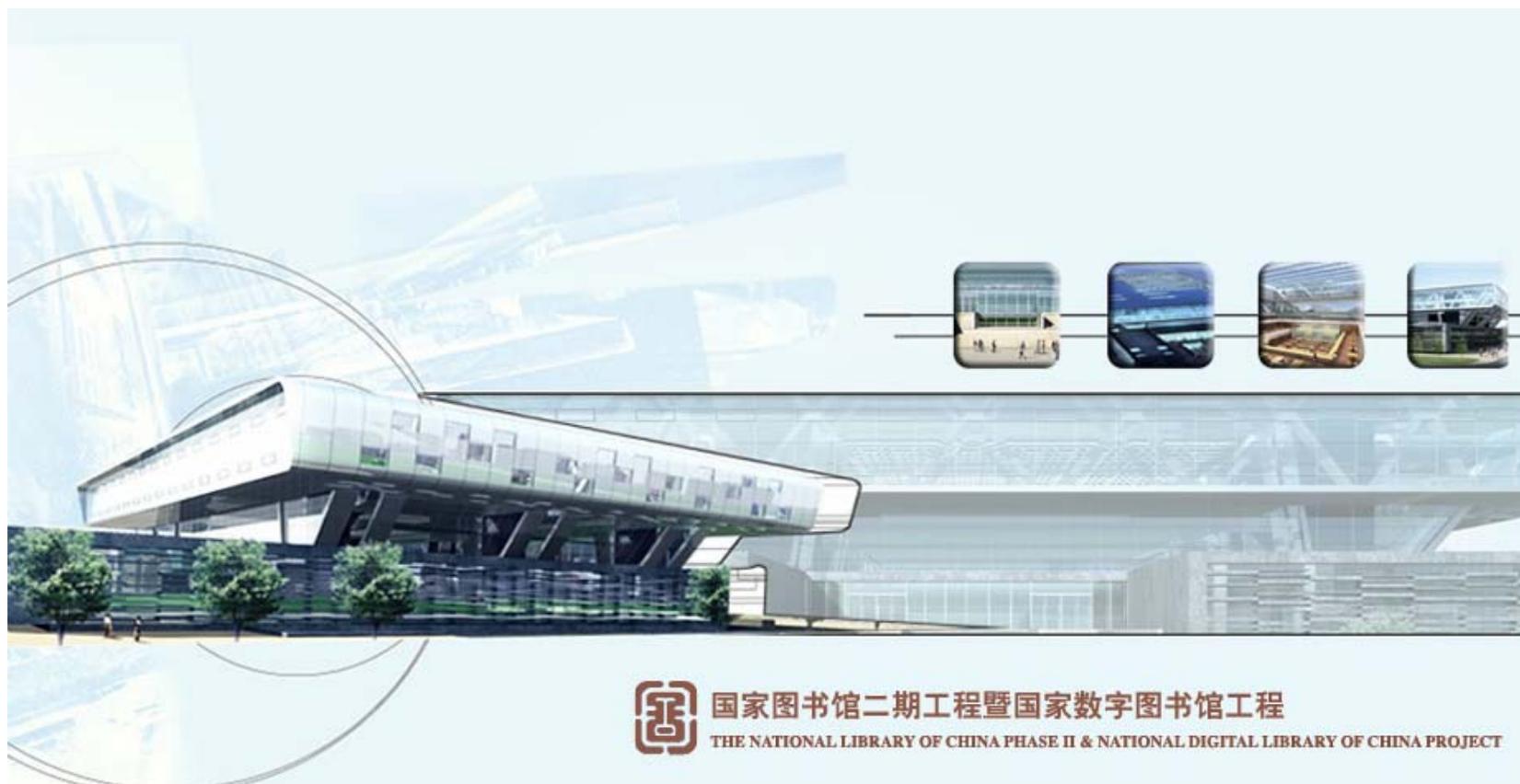
GÖTTINGEN



## 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)

5

Sep. 16, 2005

International Conference on Preservation of Digital Objects

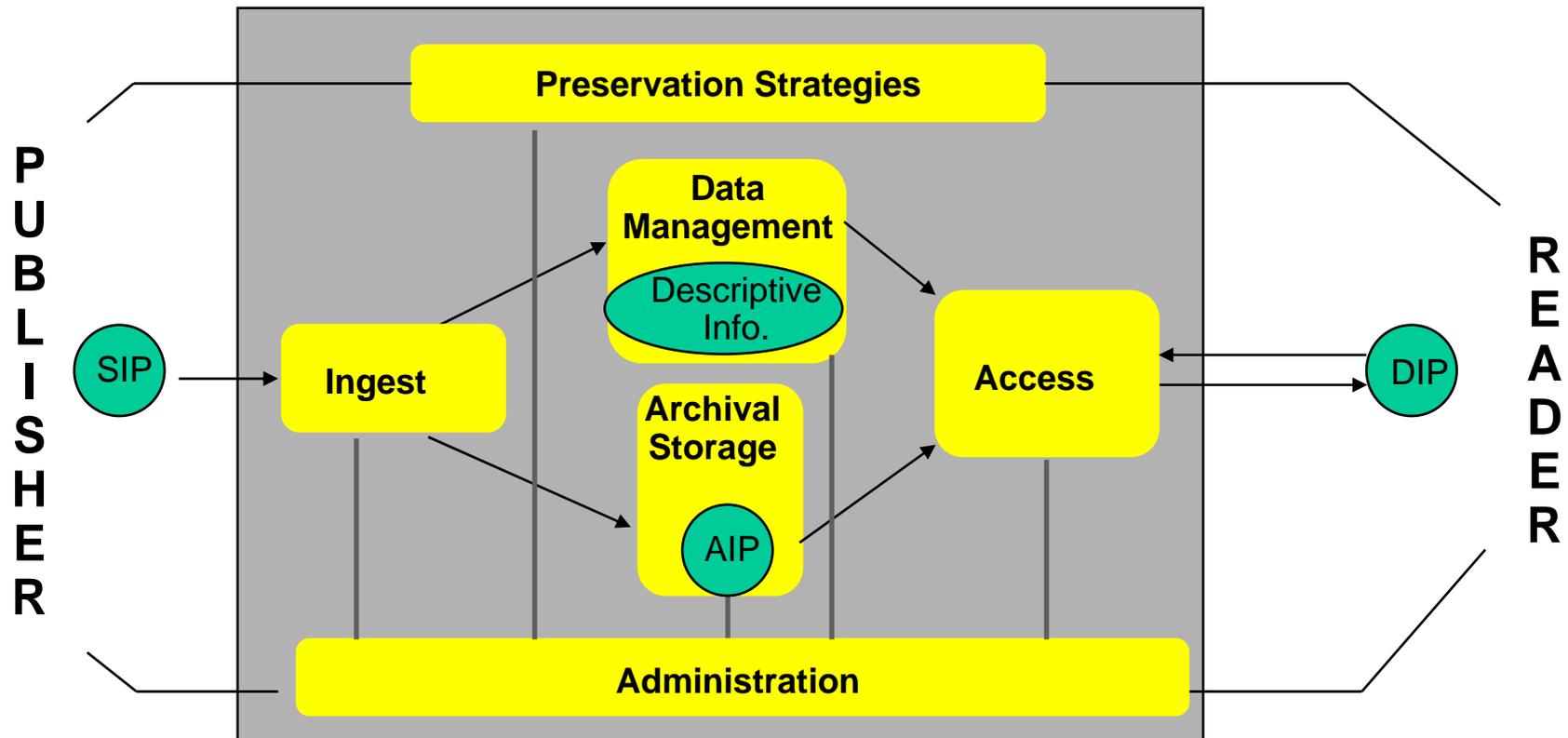
GÖTTINGEN



# Research in Progress



# OAIS and Software



SIP = Submission Information Package

AIP = Archival Information Package

DIP = Dissemination Information Package

7



## *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 WICP 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.



Keep close to people  
Li Chuan

每日新  
A叠 共24  
今日正报 32版

胡锦涛访美重在表明  
温家宝回信鼓

胡  
要大力  
国

胡  
要大力  
国

胡  
要大力  
国



Wang Yun, 41, is a...  
Katrina may have killed 50 in Mississippi

胡  
要大力  
国

# 东方都市报



CITY LADY  
有奖新闻热线 0531-82905008

# 都市女报

——和都市女性一起成长

总编辑信箱 dsnbzb@e23.com.cn  
国内统一刊号 CN37-0109 邮发代号 23-163 网址: http://www.e23.cn

2005年8月  
星期三  
农历乙酉年七月廿七  
1004期 今日32版

- 足彩三人行专家
- 01 富勒姆 vs 切尔西
  - 02 维拉 vs 阿森纳
  - 03 曼城 vs 利物浦
  - 04 米德尔 vs 曼联
  - 05 纽卡斯 vs 热刺
  - 06 热刺 vs 西汉姆
  - 07 西汉姆 vs 维冈
  - 08 维冈 vs 杜伊斯
  - 09 杜伊斯 vs 沙尔克
  - 10 沙尔克 vs 不来梅
  - 11 沃尔夫 vs 不来梅
  - 12 不来梅 vs 拜仁



泉城挂职干部资助湘妹子

5版

麦当劳肯德基回应“致癌说”

6版

生活周刊

直面“癌症之王”肝癌

12版

新婚买家电 怎样更划算



## 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**



Default - FlashGet

文件(F) 类别(C) 编辑(E) 查看(V) 任务(T) 工具

FlashGet interface showing playback controls (play, pause, stop, next, previous, full screen) and a file list.

- 正在下载
- 已下载
  - 软件
  - 游戏
  - 驱动程序
  - Mp3
- 已删除

文件名  
s010831.pdf

详细信息  
链接

| 名称     | 值                 |
|--------|-------------------|
| * URL  |                   |
| * 注释   |                   |
| * 引用页  |                   |
| * 续传   |                   |
| * 本地文件 |                   |
| * 日期   |                   |
| * 大小   |                   |
| * 完成   |                   |
| * 总耗时  |                   |
| * 平均速度 |                   |
| * 创建时间 | 08/31/05 21:21:48 |
| * 完成时间 | 08/31/05 21:22:03 |

Adobe Reader - [s010831.pdf]

文件(F) 编辑(E) 视图(V) 文档(O) 工具(T) 窗口(W) 帮助(H)

96% 打开 保存副本 打印 发送 搜索 文本选择工具 组织您的数码照片。

深圳日报  
**Shenzhen Daily**

Wang Yusi, a 15-year-old SZ girl, wins modeling award Page 3

Katrina may have killed 80 in Mississippi Page 6

HK filmmaker Tsui Hark planning sequel to "Seven Swords" Page 14

Wednesday August 31, 2005 Vol. 9 No. 1489 Price: 1 yuan A publication of the Shenzhen Press Group 国内统一刊号: CN44-0178 邮发代号: 45-113

**Keep close to people: mayor**

Rita Liu

THE municipal government would continue its efforts to communicate with citizens, Mayor Xu Zongheng said Tuesday at a phone-in radio program.

Xu said a mature government should attach a great deal of importance to communicating with its citizens, and closer ties between the government and citizens were important.

The municipal government has set up a public relations office recently to strengthen the communication with citizens.

During the one-hour broadcast, Xu discussed

**Li, Chuang vow closer ties**

10.47 x 14.6 英寸

第 1 / 1 页

Windows taskbar: 开始, 我的文档, 21:24

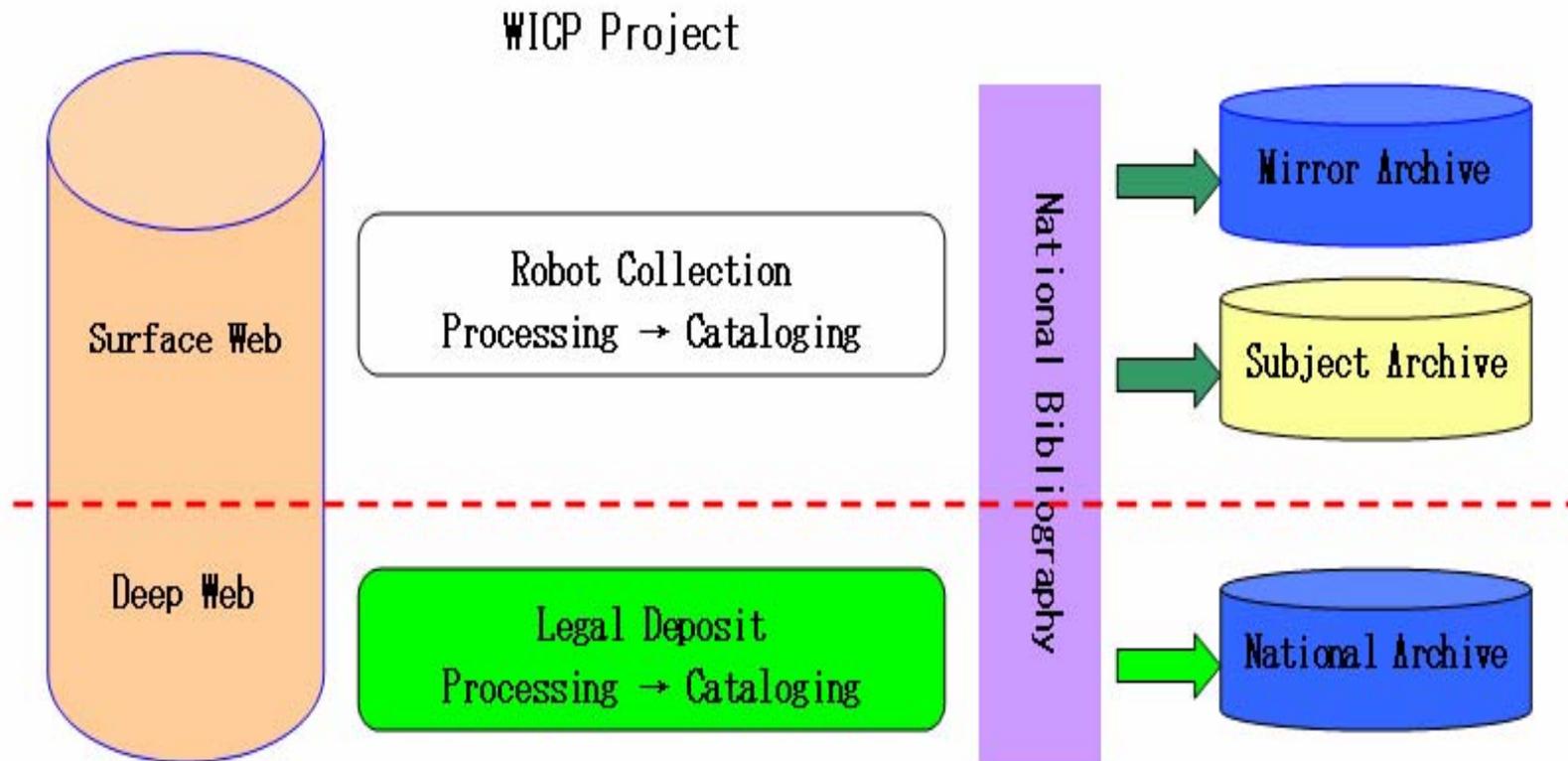


# Future Development

- ⊕ Content Enhancement
- ⊕ Content Management
- ⊕ Opening to the public



# Progress on WICP project





# Partnerships

- ▶ TianWang, Peking University
- ▶ Tsinghua Tongfang
- ▶ BeiDa Founder
- ▶ WanFang Data



# 中国Web信息博物馆

## Web InfoMall

[English](#)

### Web InfoMall 简介

“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存储与展示系统。目前已经维护有10亿以中文为主的网页, 并以平均每月一千万网页的速度扩大规模。

### 历史网页回放

http://

### Web InfoMall 信息服务

- ▶ 输入URL, 浏览永久保存的历史网页, 欣赏旧时网页的风采
- ▶ 畅游昔日网站, 随意纵横比照, 品味网络世界的兴衰变迁
- ▶ 关注重大历史事件, 将发展进程历历览尽, 感受时代的进步
- ▶ 申请网页数据, 研究深层联系, 挖掘信息世界的潜在秘密
- ◉ Web InfoMall是一项服务社会的公益事业
- ◉ 汇集历史网页, 再现大千变化, 包容万千信息探索无穷智慧

### Web InfoMall 数据量

▶ 共有自2001年以来 1,017,104,240 网页在线 (17,106 GB) [详细 >>>](#)

**历史事件专题回放**

- [党的十六大专题](#)
- [聚焦伊拉克战争](#)
- [2003年4月底广东网上信息的一次快照与分析](#)
- [抗击非典的日子](#)
- [中国首次载人航天](#)
- NEW** [其他..](#)
- [建立自己的主题..](#)

**信息产品与服务**

- [中国Web网页全文](#)
- [中国Web网页链接URL](#)
- [天网搜索引擎用户日志](#)
- [中国IP地址信息大全](#)
- [中文网页分类训练集](#)
- [理念与规则](#)
- [公共许可证](#)
- [申请使用数据](#)

**相关链接**

- [天网: Web InfoMall工作的基础](#)
- [燕窝: 在Web InfoMall基础上的科学研究和增值信息服务体系](#)
- [CNDS: Web InfoMall的摇篮](#)

**典型历史网页展示**

- [1996年中国网页拾零](#)
- [1997年中国网页拾零](#)

**关于我们**

- [使用帮助](#)
- [问题解答](#)
- [Web InfoMall论坛](#)
- [联系我们](#)

Thank you very much for your attention.

My Email: [licm@nlc.gov.cn](mailto:licm@nlc.gov.cn)

National Library of China

<http://www.nlc.gov.cn>