Digital Humanities & Future of Academic Library in China

Benjun Zhu

Peking University Library
bjzhu@pku.edu.cn
OUTLINE

• A Brief Introduction on DH in mainland China
• Our Group's Research and Practices in DH
• My Opinion on Trends of Academic Libraries
I

A Brief Introduction on DH in mainland China
1. Mainland China Has Experienced A Process from A Rudimentary Stage to An Active Stage

before 2010

humanistic care in the digital age

in 2010

align itself with western discourse
Post 2015

ranges from applying digital technology to humanities research (such as text mining, SNA, GIS), to collating humanities materials (such as digitization, database and dataset)

Introducing development cases and reviews on DH related conferences in western country

↑ results of retrieving "digital humanities" in title and keywords on cnki.net
2. Further Development
Simon Mahony  
Director, UCL Centre for Digital Humanities

Peter K. Bol  
Harvard University

J. Stephen Downie  
Professor of Illinois University

Susan Schreibman, Professor of Digital Humanities and Director of An Foras Feasa, Maynooth University, Ireland
3. Purposes and functions of DH

- to help humanists to save their time and energy against a low efficiency in literature
- to introduce new digital tools and methods to answer traditional humanistic questions
- to raise new questions and answering them using new digital thoughts
- disrupt the norms and introduce creativity in the field of traditional humanities
4. Many DH Projects Done by Libraries and Humanists

- **Digitization**
  - China Epitaph Database (中國歷代墓誌)
  - Oracle Bone Database (甲骨文)
  - China Memory Project (中國記憶)

- **Dataset**
  - Seals Authority Datasets (中國歷代印鑒規範庫)

- **DH tools and applications**
  - Poet life Map of Tang and Song Dynasty (唐宋詩人地圖)
  - Soyun (搜韻)
  - ShuTongWen (書同文)
5. Reflections and Debates on DH in China

To study literature and history, databases are almost indispensable.

研文治史，數據庫幾乎是到了須臾不可或離的地步。


Wang Ruilai (王瑞来), a professor good at Chinese literature and history at Gakushuin University of Japan
5. Reflections and Debates on DH in China

- Just an alternative approach to retrieve information.
- Just to explain and present our existing research results in a technical way repeatedly but without any innovation.
- Just to solve problems in a narrow perspective not knowing broad contexts of problems.
- Digital tools developed originally for social sciences research, but they don’t play a decisive role in promoting humanities research.
6. Problems lie:

(1) Full text is fragmented and incomplete
The book He Lin Yu Lu (鶴林玉露) has several versions, but only one was included in ZhongGuo Jiben Guji Ku (中國基本古籍庫).
(2) Content is inaccurate

"Wrong characters produced by OCR, which affects scholars’ analysis on text."

Ruilai Wang
(3) Databases or full texts lack authority

- Authority datasets are rare in field of humanities
- Large-scale, high-quality, acknowledged and structured humanistic databases are important

Cameron Campbell  
HKUST

Zhan LIN  
Renmin U. of China
Databases and datasets lack norms of knowledge system

The core challenge nowadays the libraries facing is how to process massive unstructured full text.

Chen HUANG
Secretary-General of CADAL Project
(5) Systems lack interoperability

"there are too many different databases independent of each other without interoperation"

(6) Data, full text, and databases are always repeated

"librarians and scholars repeat the same work, digitize the same resources time and again, and cause huge waste of energy."

Peter K. Bol
Harvard U.
Not the DH,

but the DH Cyberinfrastructure
II
DH Cyberinfrastructure
1. Past and Present of Cyberinfrastructure

- Cyberinfrastructure for science and engineering (Freeman et al. 2005)
- Cyberinfrastructure for e-science (Hey and Trefethen 2005)
- Cyberinfrastructure for biomedical research (Buetow 2005)
- Cyberinfrastructure for the solid earth sciences (Keller and Baru 2011)
- Spatial cyberinfrastructure (Wright and Wang 2011)
- Cyberinfrastructure for seafloor observations (Li et al. 2016)

A new scientific and engineering knowledge environment and virtual organization

Early 1990s Information infrastructure

Cyberinfrastructure

1998

2003
Cyberinfrastructure

A combination of data resources, network protocols, computing platforms, and computational services that brings people, information, and computational tools together to perform science or other data-rich applications in this information-driven world.

2. Cyberinfrastructure in Humanities

As early as 2004, the American Council of Learned Societies (ACLS) noted that:

- scientists and engineers see digital technologies as a force creating environments" (CNI 2004)

- cyberinfrastructure is the most appropriate new environment (ACLS n.d.)
Cyberinfrastructure for the Humanities, Arts, and Social Sciences, 2009

Cyberinfrastructure for Humanities and Sciences, 2000
Points of Collaboration:
- Researchers
- Technical Specialists
- Automatic Processes
- Coordination Efforts

Data Analysis:
- GIS
- Text Mining
- Image Mining
- Visualizations

Knowledge Discovery
Traditionally strong in humanities
- Philosophy
- Chinese language and literature
- Archaeology
- History
- Language and art

Peking University

PKU has begun research on DH Cyberinfrastructure for China Studies
Procedures and Activities of Humanists?

Archaeology & Excavation
- recognize character
- classify
- sort & compile

Raw and Original Materials
- historical materials
- philosophy, fiction, prose, drama, poetry

buried underground

passing down
different versions
Procedures and Activities of Humanists?

Archaeology & Excavation
- recognize character
- classify
- sort & compile

Raw and Original Materials
- historical materials
- philosophy, fiction, prose, drama, poetry

Raw and Original Materials
- historical materials
- philosophy, fiction, prose, drama, poetry

passing down
different versions

buried underground

buried underground
Writings of Humanists?

**Raw and Original Materials**
- historical materials
- philosophy, fiction, prose, drama, poetry

**Preparation & Excavation**
- recognize character
- classify
- sort & compile
Collect & Compile (彙編)
INTERPRET (釋義與解讀)
Conclusions
1) Not all materials hold the same importance
2) CI should be designed around academic activities.

III
Conceptual DH Cyberinfrastructure Model for China Studies
1. Conceptual Model in An Ecosystem Perspective

\[ \text{DHCI} = \text{Digitization} + \text{Datafication} + \text{Digital Ecosystem} \]
1. Conceptual Model in An Ecosystem Perspective

$$\text{DHCI} = \text{Digitization} + \text{Datafication} + \text{Digital Ecosystem}$$
山海經
Paintings (繪畫)
南華真經序

河南郭象子玄撰

夫莊子者，可謂知本矣。故未始藏其狂言，言雖無會而獨應者也。夫應而非會則雖當無用言事則雖高不行與。夫寂然不動不得已而後起者固然有間矣。斯可謂知無心者也。夫心無為則雖應應隨應隨其時言唯謹爾故與化為體流萬代而冥物豈曾設對獨業而遊談乎？方外哉！其所所以不經而為百家之冠也。然莊生，雖未體之言則至矣。通天地之統序萬物之性。

Make sure:

- Full text is proofread
- correct
- authoritative
Vernacular title: 南華真經: 10卷
Title: Nan hua zhen jing: 10 juan
Attribution: 郭象注; 陸德明箋義.
Guo Xiang zhu; Lu Deming yin yi.
Published: [China: 頤徳堂. 明嘉靖12 i.e. 1553]
[China: Gu Chun Shi de tang, Ming Jiajing 12 i.e. 1553]
Description: 12 v.
Series / collection: Harvard-Yenching Library Chinese Rare Books Digitization Project-Philosophy
Language: Chinese
Notes: 明嘉靖12年[1553]頤徳堂刊《六子書》本．
八行十七字, 四周雙邊．白口, 單魚尾, 書口上刻「世德堂刊」, 槁高 19.8 x13.5.
Hart collection, in rare books room.
Subjects: 莊子 -- Commentaries.  
Zhuangzi -- Commentaries.
Author / Creator: 郭象, -312.  
陸德明, 556-627.  
顧春.  
莊子. 南華經.  
Guo, Xiang, -312.  
Lu, Deming, 556-627.  
Gu, Chun.  
Zhuangzi. --Nanhua jing. --
Other title: 六子書.
Liu zi shu.
Creation Date: Ming Jiajing 12 i.e. 1553]
HOLLIS number: 99007550710203941
Permalink: http://id.lib.harvard.edu/alma/99007550710203941/catalog
Source: HVD - Core
南華真經序

夫莊子者，可謂知本矣。故未始藏其狂言，雖無會而獨應者，也夫應而非會則雖當無用言，非物事則雖高不行，與夫寂然不動，而已而後起者固有聞矣。斯可謂知無心者也。夫心無則隨感而應應隨其時言，唯謹爾故與化為體流萬代而實物豈曾設對獨遂而游趨乎方哉？若所以不經而為百家之冠也。然莊生雖未體之言則至矣。通天地之統序萬物之性。
| 复制
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>character (印面文字)</td>
</tr>
<tr>
<td>image (印文圖片)</td>
</tr>
<tr>
<td>owner (印章主人)</td>
</tr>
<tr>
<td>sculptor (治印者)</td>
</tr>
<tr>
<td>source (來源)</td>
</tr>
<tr>
<td>dynasty (時間)</td>
</tr>
</tbody>
</table>
Digitization + Datafication + Digital Ecosystem
DNA = Digitization + Datafication + Digital Ecosystem
2. Imaginary Scenarios

- Libraries, museums, archives, and galleries
- Humanists
- Make traditional research more easier
- Research paradigm can be changed
此印為「乾隆五簑」之一，所謂「乾隆五簑」，是指乾隆皇帝鑄鈐在內府藏品上皇帝收藏印，按通常的說法，這五簑的印文分別是「乾隆御覽之寶」、「平西鑄箴」、「三希堂精鑑璽」、「乾隆寶璽」、「宜子孫」。

三希堂精鑑璽

此印為「乾隆五簑」之一，所謂「乾隆五簑」，是指乾隆皇帝鑄鈐在內府藏品上皇帝收藏印，按通常的說法，這五簑的印文分別是「乾隆御覽之寶」、「平西鑄箴」、「三希堂精鑑璽」、「乾隆寶璽」、「宜子孫」。

乾隆御覽之寶

此印為「乾隆五簑」之一，所謂「乾隆五簑」，是指乾隆皇帝鑄鈐在內府藏品上皇帝收藏印，按通常的說法，這五簑的印文分別是「乾隆御覽之寶」、「平西鑄箴」、「三希堂精鑑璽」、「乾隆寶璽」、「宜子孫」。

乾隆鑑賞

此印為「乾隆五簑」之一，所謂「乾隆五簑」，是指乾隆皇帝鑄鈐在內府藏品上皇帝收藏印，按通常的說法，這五簑的印文分別是「乾隆御覽之寶」、「平西鑄箴」、「三希堂精鑑璽」、「乾隆寶璽」、「宜子孫」。
□ 朝代 dynasty
- 唐 Tang
- 北宋 North Song
- 南宋 South Song

□ 治印者 sculptor
- 陈巨来 Chen Julai
- 张志鱼 Zhang Zhiyu
- 齐白石 Qi Baishi

□ 地域 area
- 浙江 Zhejiang
- 湖北 Hubei
- 山东 Shandong

□ 寓意 moral
- 师承纪学
- 世家传承
- 濮源来历
- 經歷紀事
- 寓言譬喻

□ 姓名字號 names
- 姓
- 名
- 字
- 號

EXPORT
To make drawing historical atlas more easier

- [ ] Which government
  - [√] 清 (Qing)
  - [ ] 军政府 (Military government)
  - [ ] 南京临时政府 (Nanjing provisional governm)
  - [ ] 苏维埃共和国 (chinese soviet republic)
  - [ ] 伪满 (The "Manchukuo" Puppet Regime)

- [ ] Geography
  - [ ] 沿革
  - [ ] 疆域 (spheres of influence)
  - [√] 交通 (Traffic routes)

- [ ] 人物
  - [ ] 中央
  - [ ] 地方
  - [ ] 文献
2. Imaginary Scenarios

- Libraries, museums, archives, and galleries
- Humanists
- Make traditional research more easier
- Research paradigm can be changed
3. Current Work and Future Strategies

- Design an experimental platform based on open source software
  - Digitization module, we developed it based on MediaWiki along for some personal modifications;
  - Datafication module, we developed based on WordPress with different plugins and themes;
  - Ecosystem module, we developed it based on UCenter also along for some personal modifications.
  - All systems interoperates with APIs.

- Redefined metadata scheme in a DH perspective
  - for existed catalogue of common literature
  - for content annotation

- Timeline associated with geography significance

histchina.cn
Future of Academic Library
2. Library Trends: Data Library

1990s VS. 2020s

Libraries should play an important role in DHCI
Change as soon as possible!
Thank You!