

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

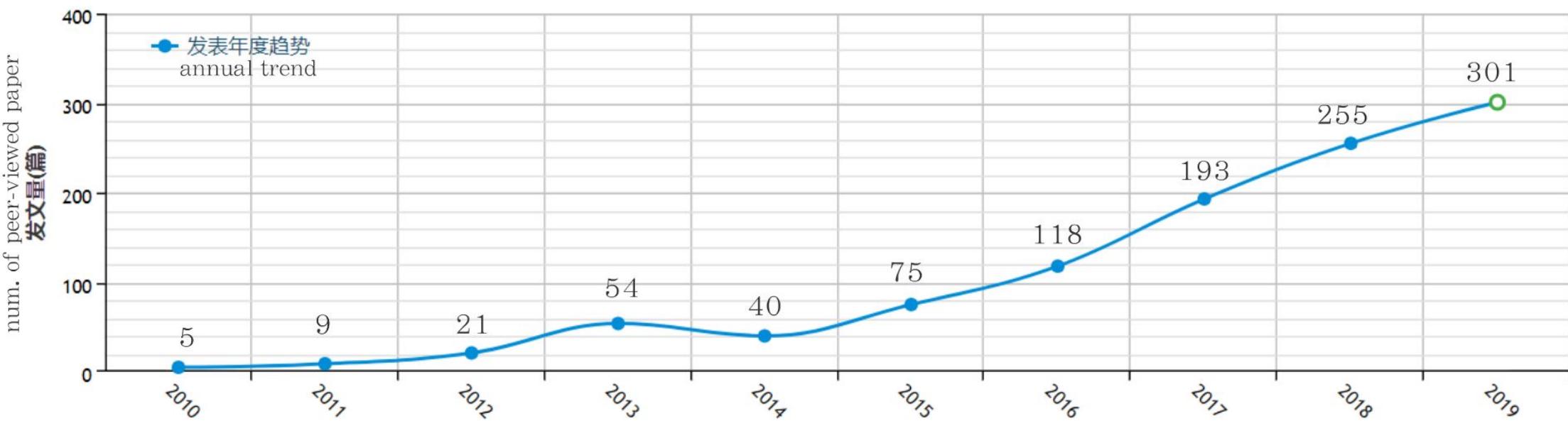
in 2010



align itself with western discourse

Post 2015

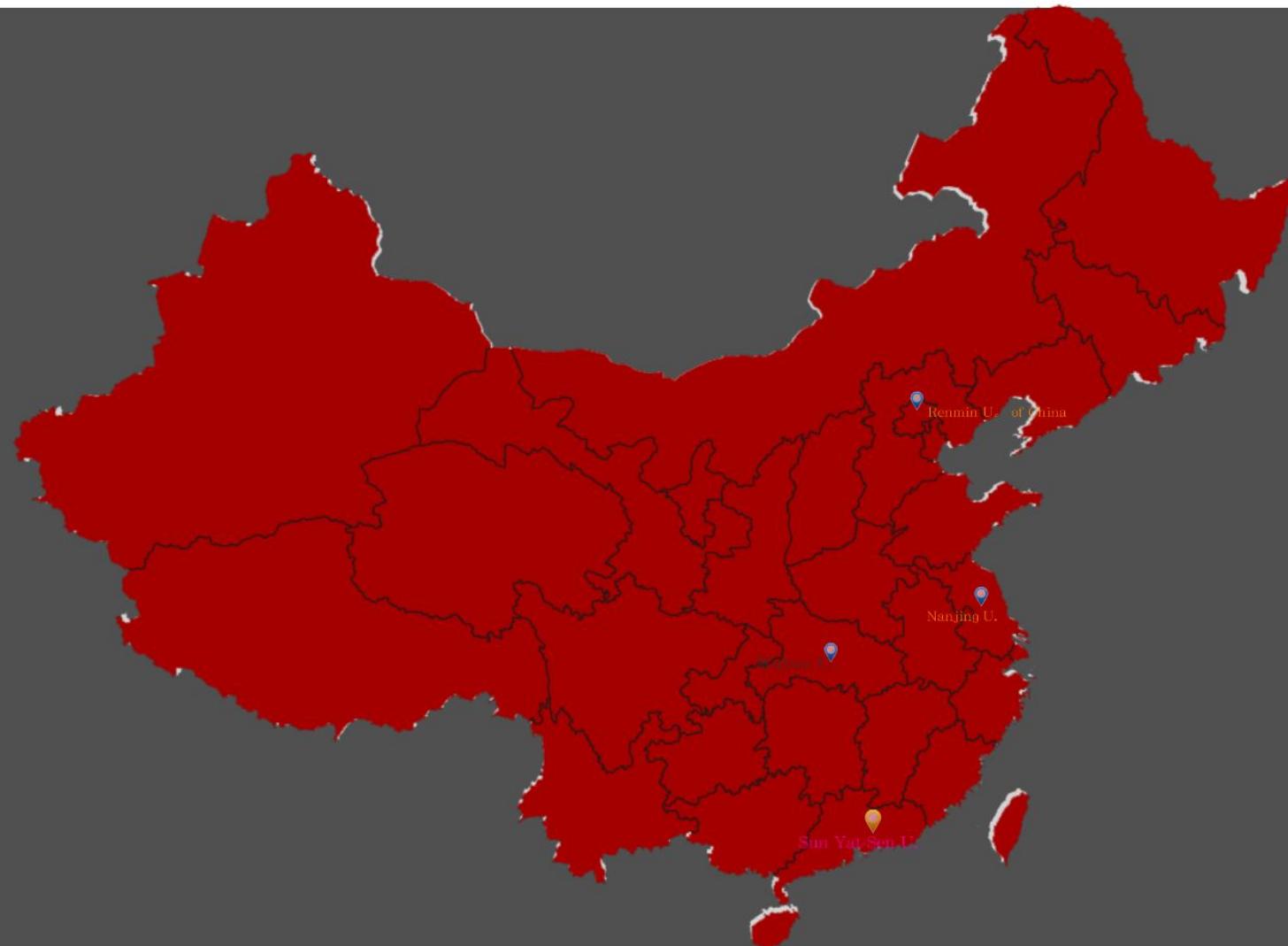
ranges from applying digital technology to humanities research (such as text mining, SNA, GIS), to collating humanities materials (such as digitization, database and dataset)

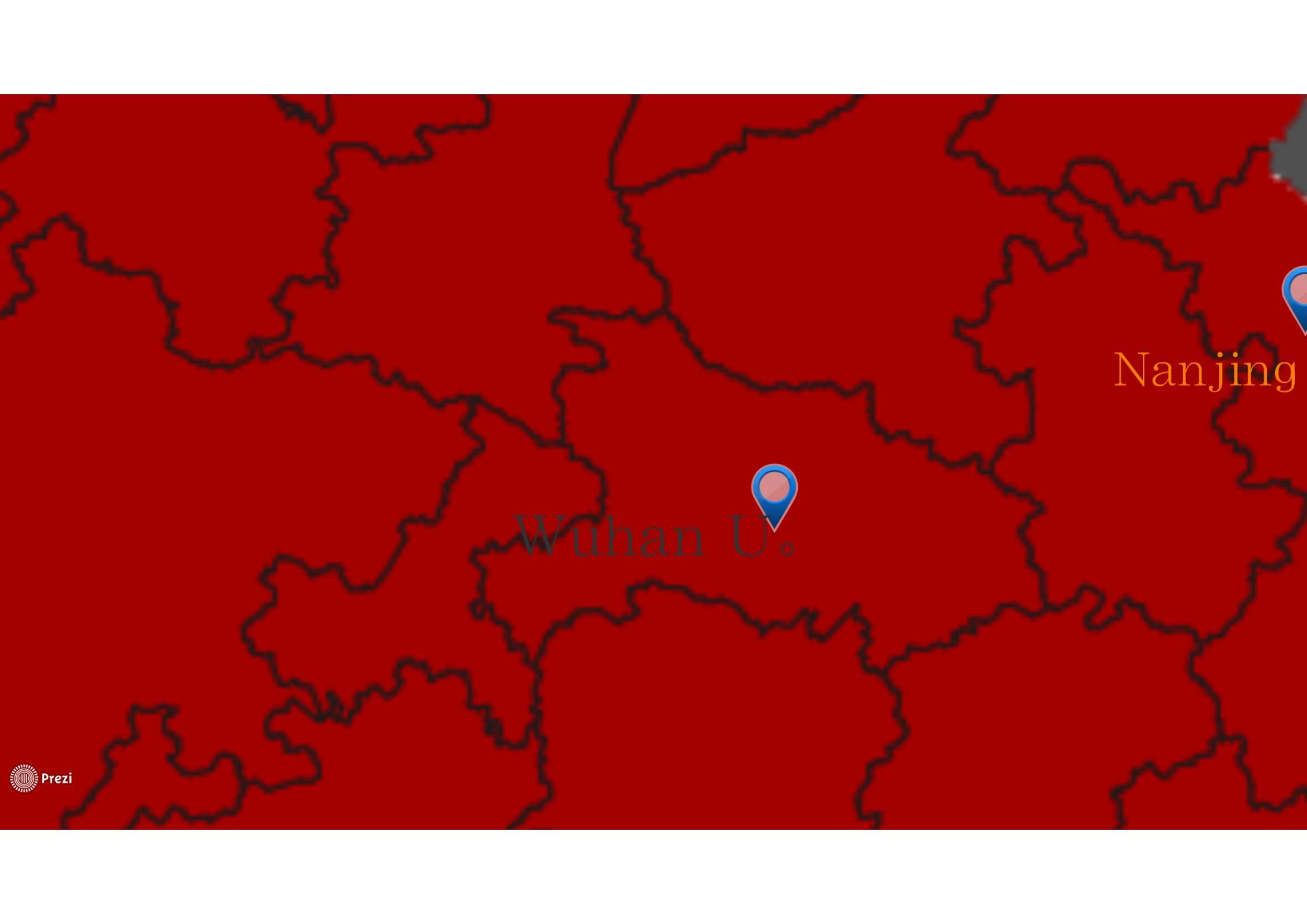


Introducing development cases and reviews on DH related conferences in western country

↑ results of retrieving "digital humanities" in title and keywords on cnki.net

2. Further Development

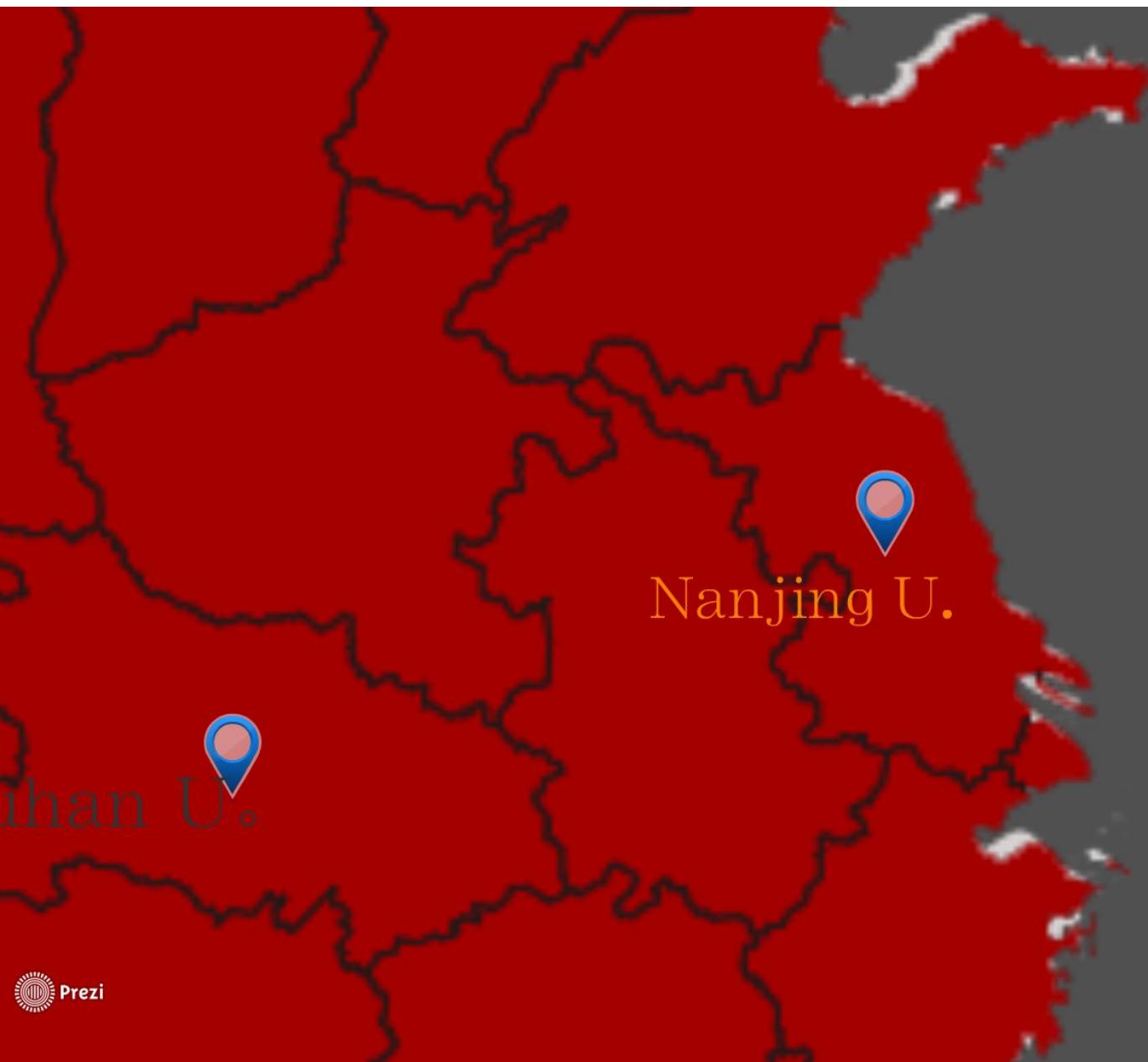




Nanjing

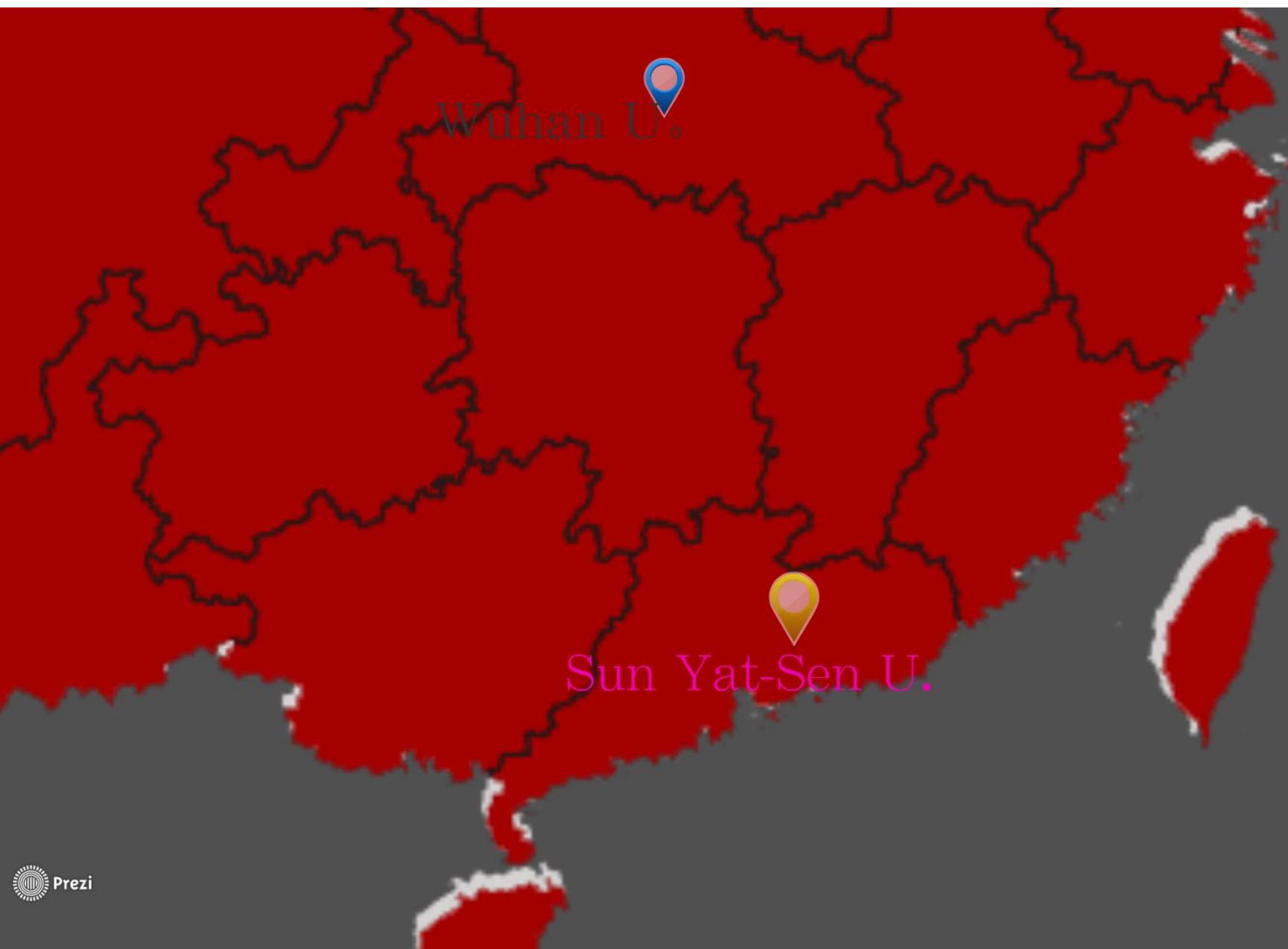
Wuhan U.





Wuhan U.

Nanjing U.





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



第三届北京大学数字人文论坛·孵化与实践：需求驱动下的数字人文项目
The 3rd Peking University Digital Humanities Forum
Incubation and Best Practices: How Digital Humanities Projects Cater to Academic Needs



第二届北京大学数字人文论坛 · 互动与共生：数字人文与史学研究

Interaction and Coexistence: Digital Humanities and Historical Research



2017年5月，“互动与共生：数字人文与史学研究”
2nd Peking U. DH Forum: Interaction and Coexistence
Digital Humanities and Historical Research

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(中國歷代印鑑規範庫)

- 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缺的地步。

Reference: 王瑞来.2018.警惕数据库. 史学月刊 455(9):22-27



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

(2) Content is inaccurate

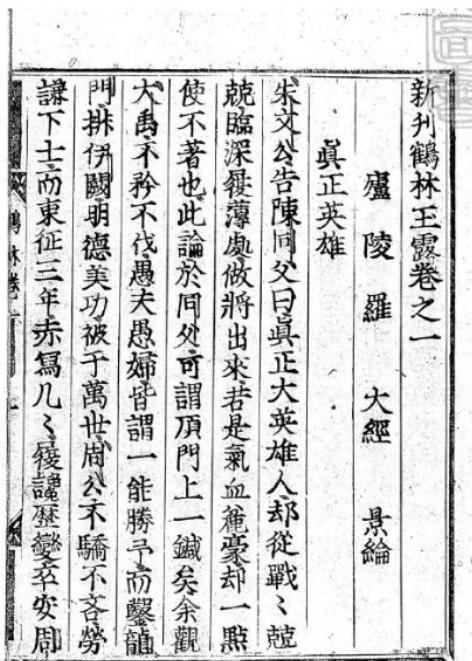


(3) Inconsistency of data

(4) Databases and datasets lack norms of knowledge system

(5) Semantic heterogeneity

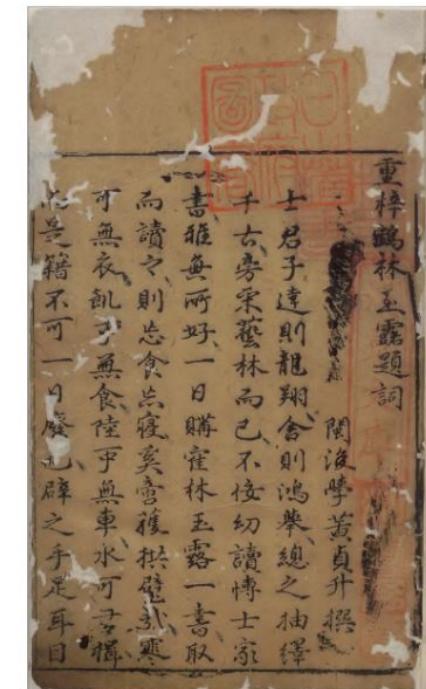
The book He Lin Yu Lu(鶴林玉露) has several versions, but only one was included in ZhongGuo Jiben Guji Ku (中國基本古籍庫)



新刊鶴林玉露：十八卷.庆长元和间.古活字印本



新刊鶴林玉露 著者羅大經（宋）慶長～元和刊本



鶴林玉露 著者羅大經（宋），明万曆12年，仁實堂刊本

(2) Content is inaccurate



Ruilai Wang

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

Ruilai Wang

(3) Databases or full texts lack authority



Cameron Campbell
HKUST

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



Zhan LIN
Renmin U. of China

(4)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"



Peter K.Bol

Harvard U.

(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."

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)

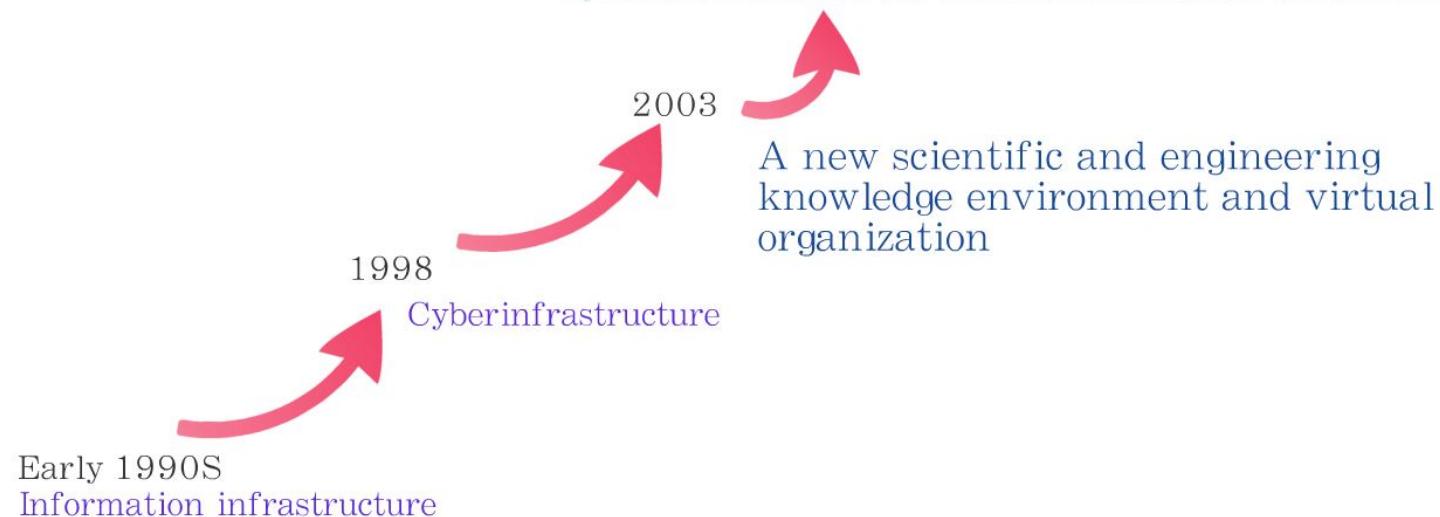
cyberinfrastructure for e-science (Hey and Trefethen 2005)

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)



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.

Yang, Chaowei, Robert Raskin, Michael Goodchild, and Mark Gahegan. 2010.
“Geospatial cyberinfrastructure: Past, present and future.” Computers, Environment and Urban Systems34 (4): 264-77.

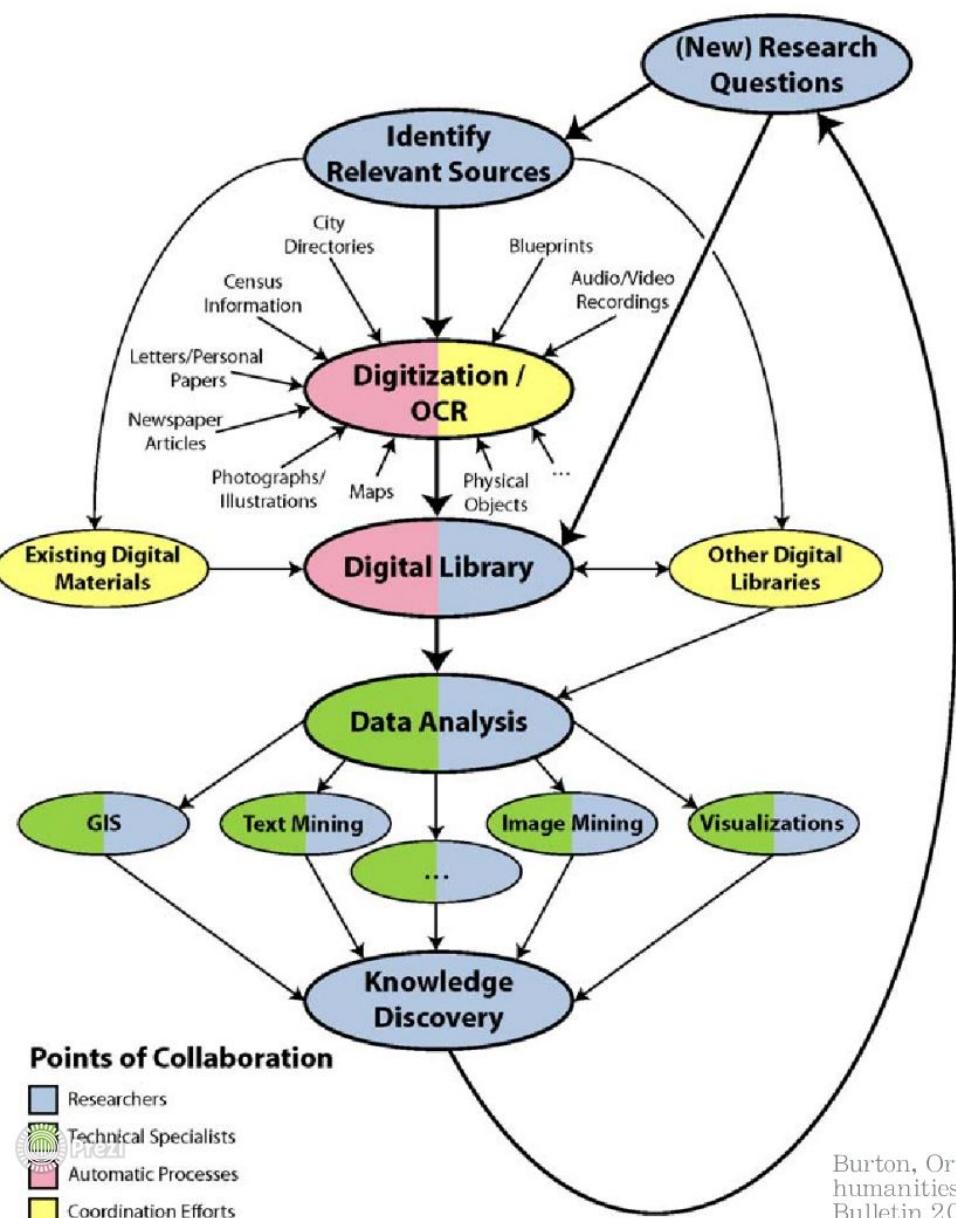
2. Cyberinfrastructure in Humanities

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.)

The screenshot shows the homepage of the American Council of Learned Societies (ACLS). The header features the ACLS logo, which is a red stylized flower or star shape. The navigation bar includes links for News, OFA Portals, Board/CEO Net Log-in, Search, and a red DONATE button. Below the header, there are several program sections: FELLOWSHIP & GRANT PROGRAMS, AFRICAN HUMANITIES PROGRAM, THE ROBERT H. N. HO FAMILY FOUNDATION PROGRAM IN BUDDHIST STUDIES, ACLS PROGRAMS IN CHINA STUDIES, and PAST PROGRAMS. A sidebar on the right is titled "REPORT" and contains a link to "Our_Cultural_Commonwealth". The main content area discusses the "ACLS Commission on Cyberinfrastructure", mentioning its support by the Andrew W. Mellon Foundation and its role in gathering information and developing perspective for the calendar year 2004.

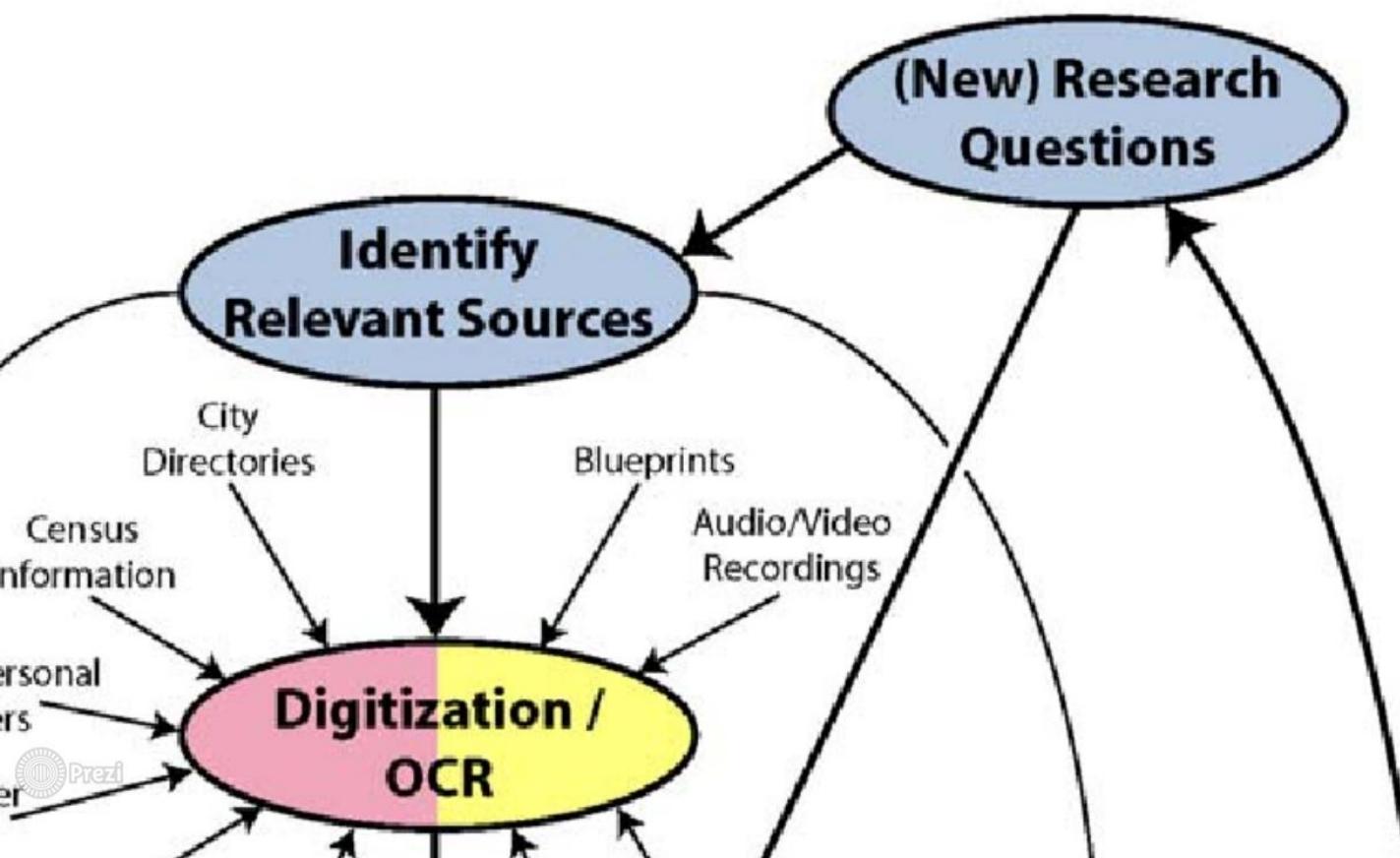
<https://www.acls.org/cyberinfrastructure>

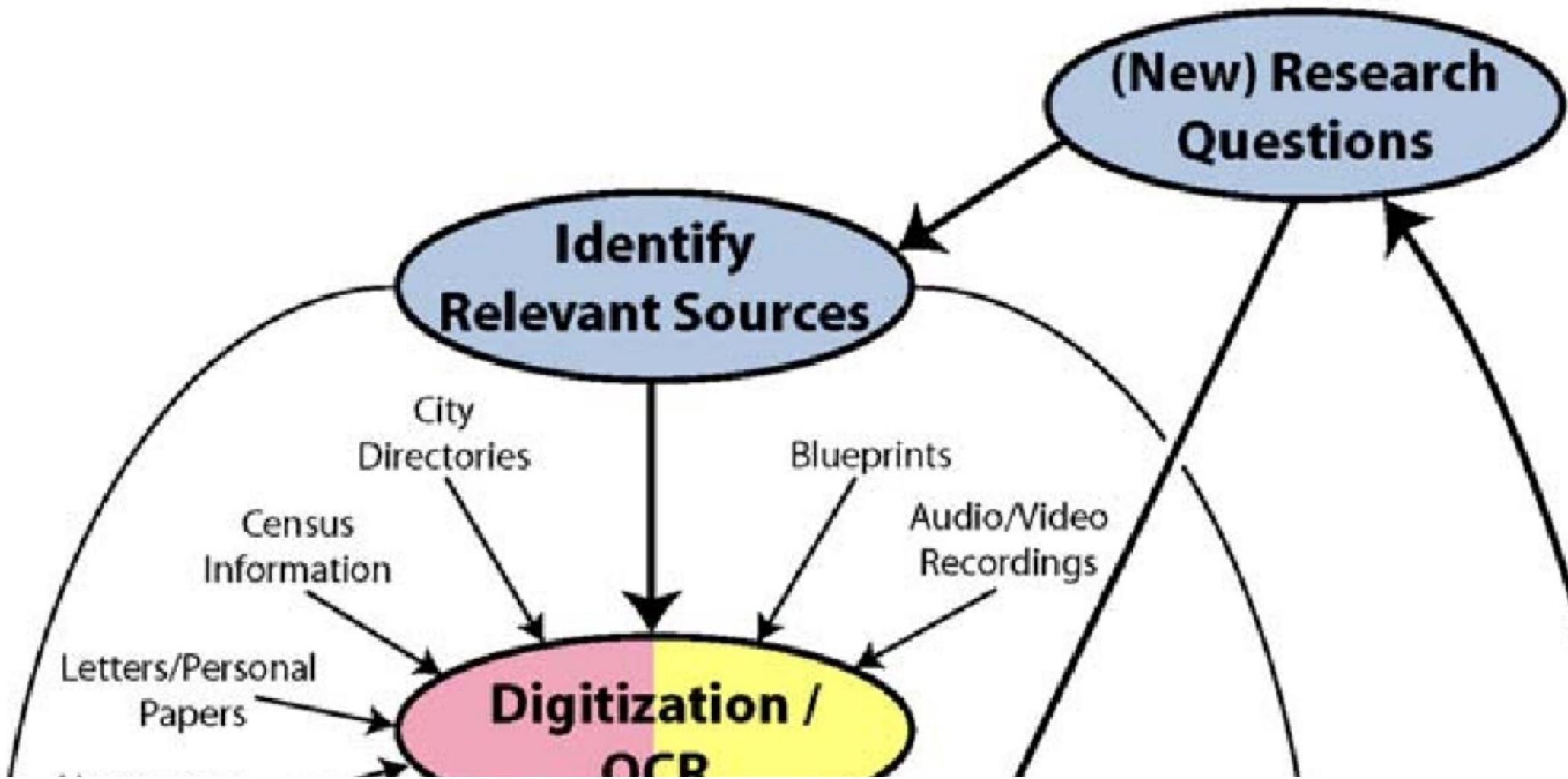


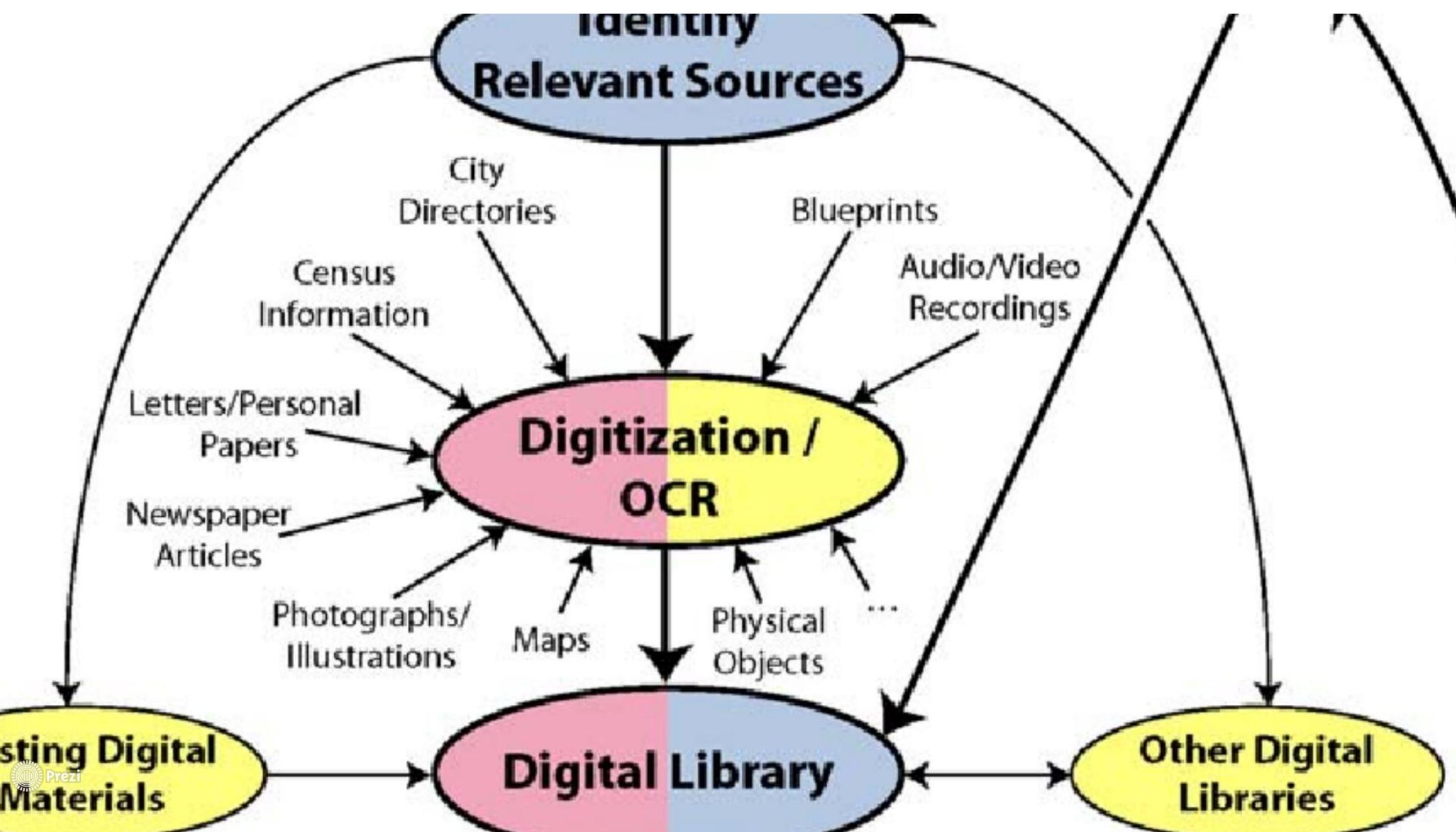
Cyberinfrastructure for the Humanities, Arts, and Social Sciences, 2009

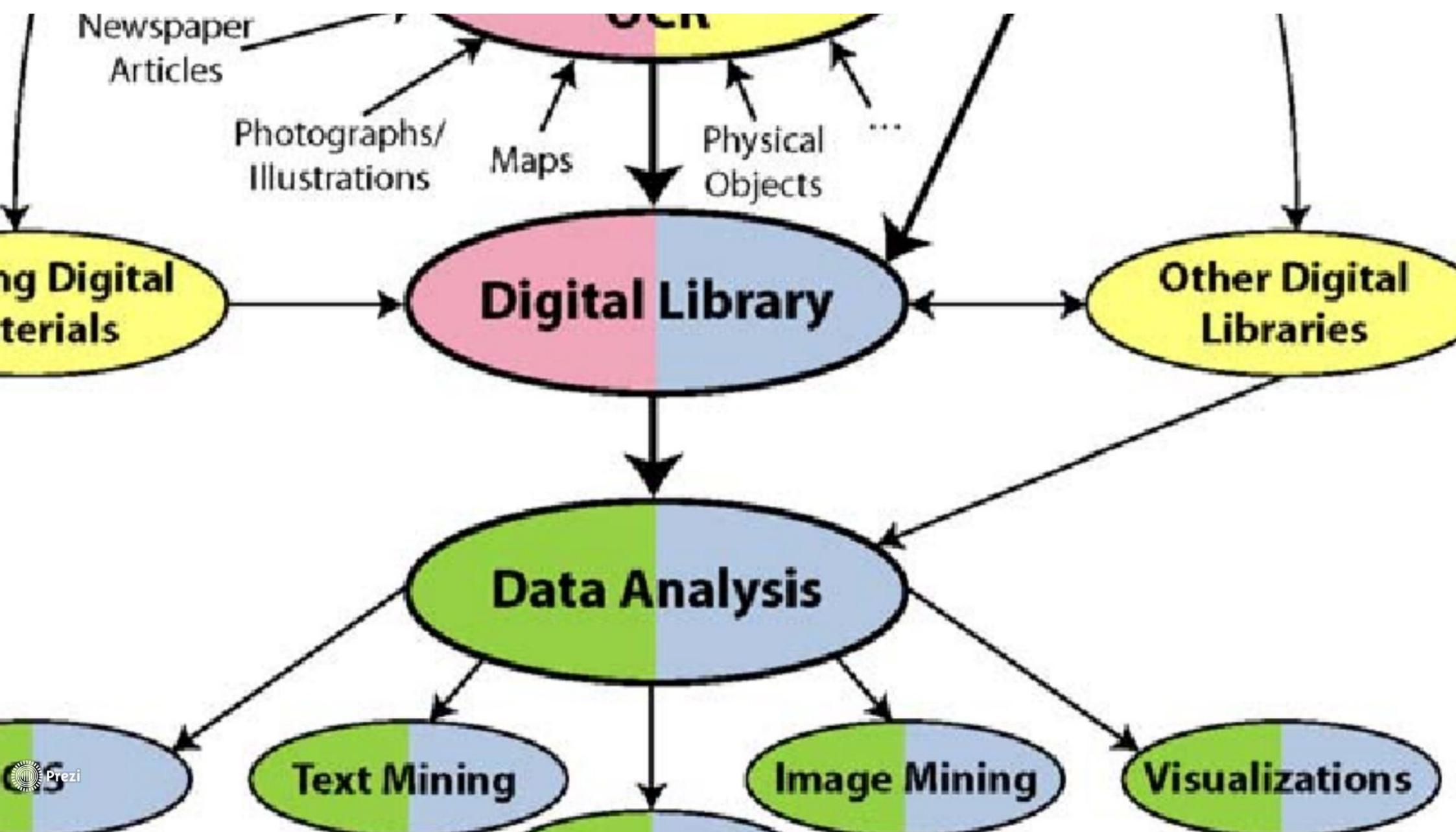
Burton, Orville Vernon, and Simon Appleford. 2009. "Cyberinfrastructure for the humanities, arts, and social sciences." EDUCAUSE Center for Applied Research-Research Bulletin 2009, no. 1.

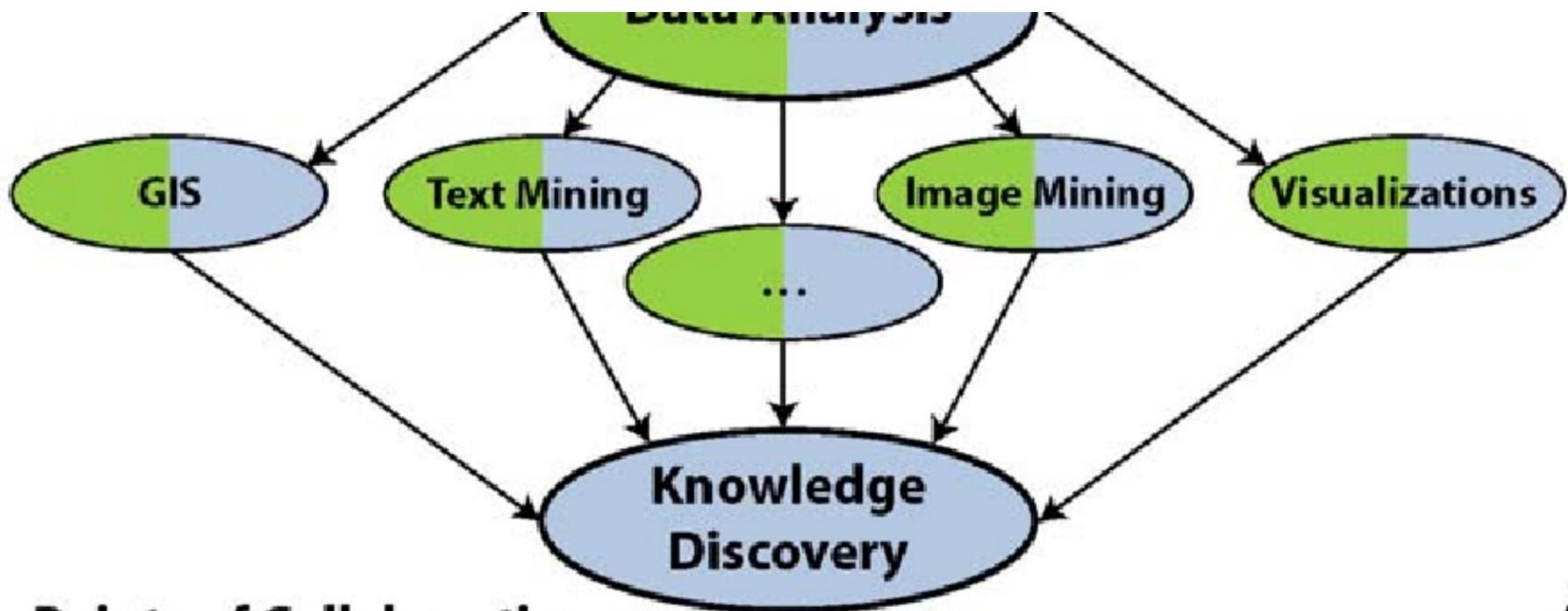
Cyberinfra Humaniti Sciences, 2









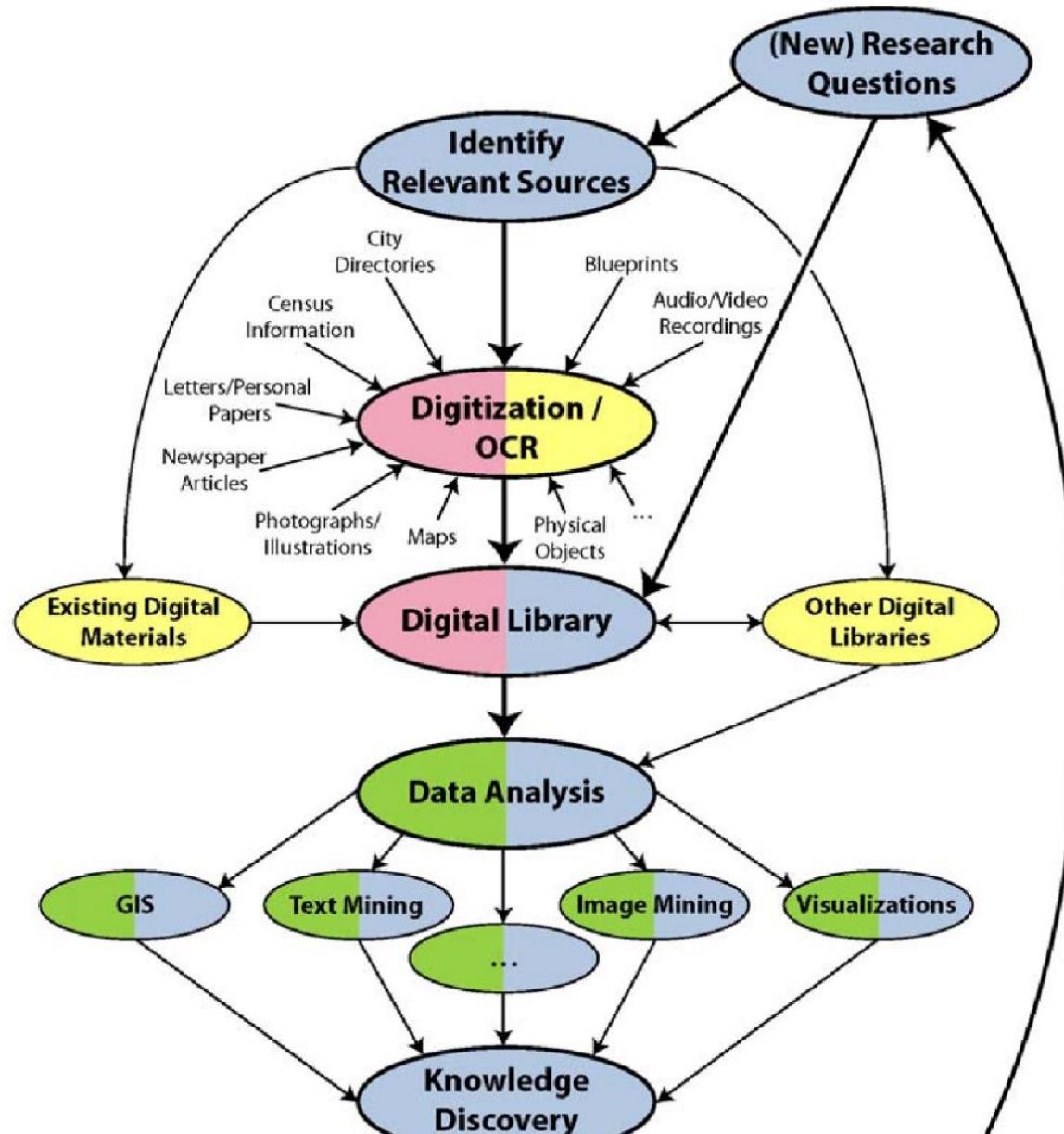


Points of Collaboration

- [Blue square] Researchers
- [Green square] Technical Specialists
- [Pink square] Automatic Processes
- [Yellow square] Coordination Efforts

Burton, Orv
humanities,
Bulletin 20

Cyberinfrast Humanities, Sciences,200





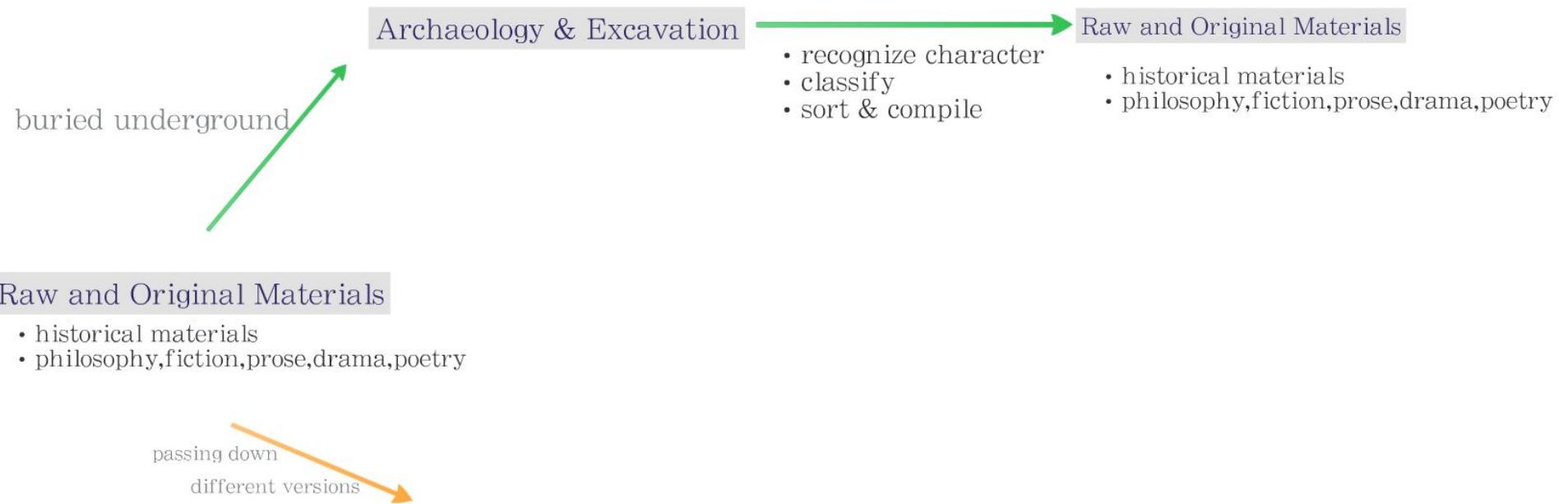
Peking University

Traditionally strong in humanities

- Philosophy
- Chinese language and literature
- Archaeology
- History
- Language and art

PKU has begun research on DH Cyberinfrastructure for China Studies

Procedures and Activities of Humanists?



ry

INTERPRET (釋義)

PROOFREAD (校勘)

COMMENT (評論)

ANNOTATE (標註)

MAKE NOTES (眉批)

REWRITE(改寫)

Correct and Give Exact Implication(正義)

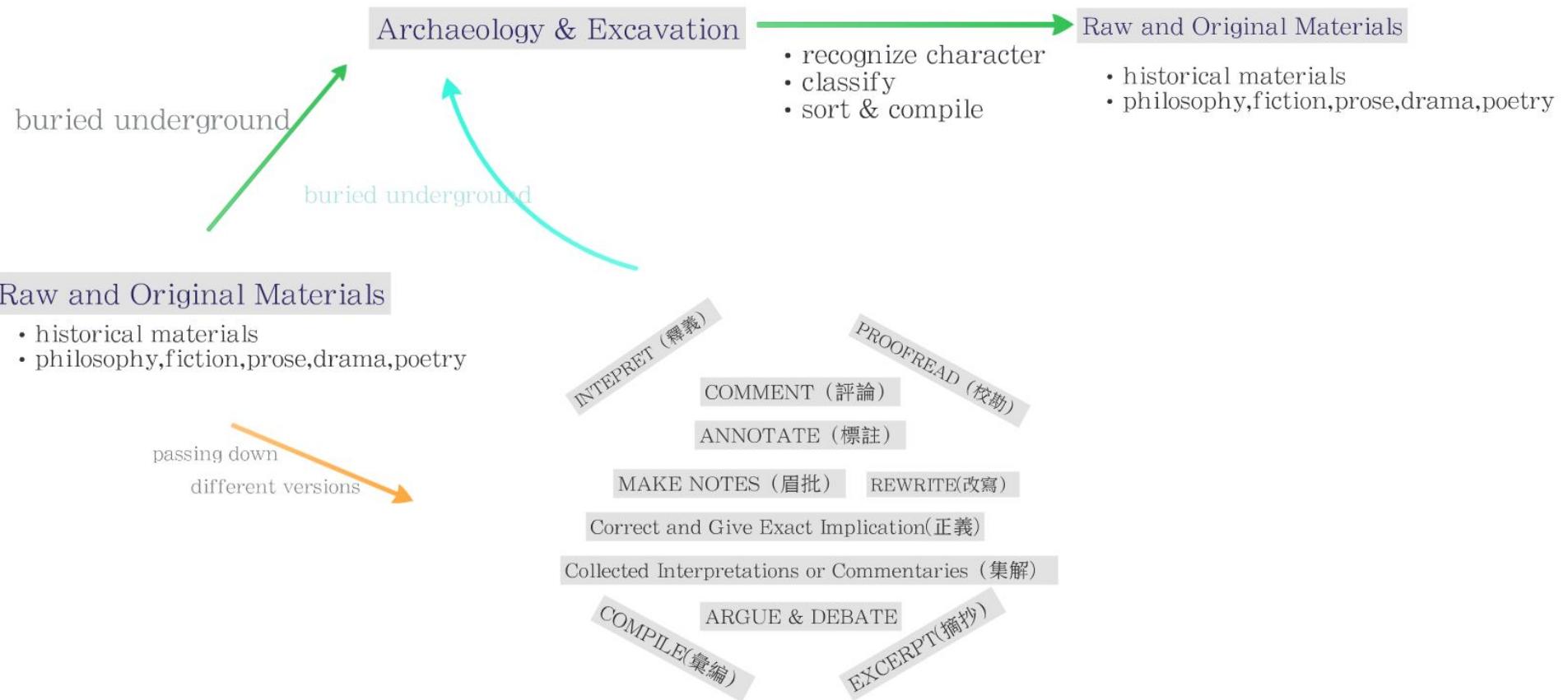
Collected Interpretations or Commentaries (集解)

COMPILE(彙編)

ARGUE & DEBATE

EXCERPT(摘抄)

Procedures and Activities of Humanists?



What are the Practices of Humanists?

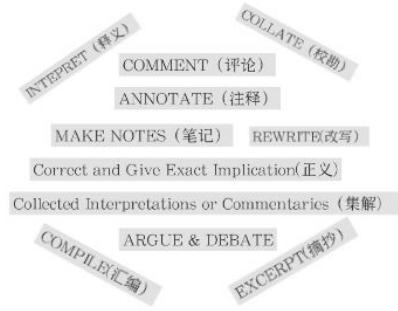
Archaeology & Excavation

- recognize character
- classify
- sort & compile

Raw and Original Materials

- historical materials
- philosophy, fiction, prose, drama, poetry

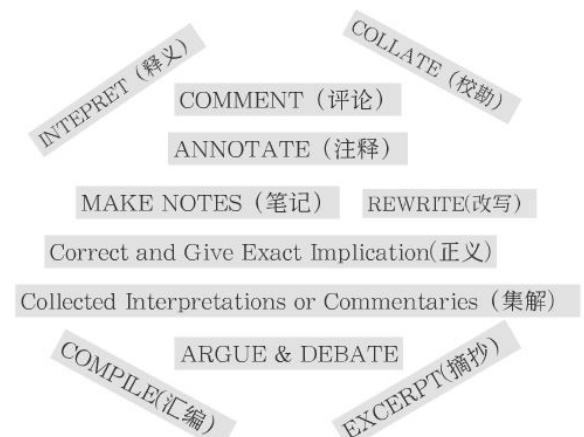
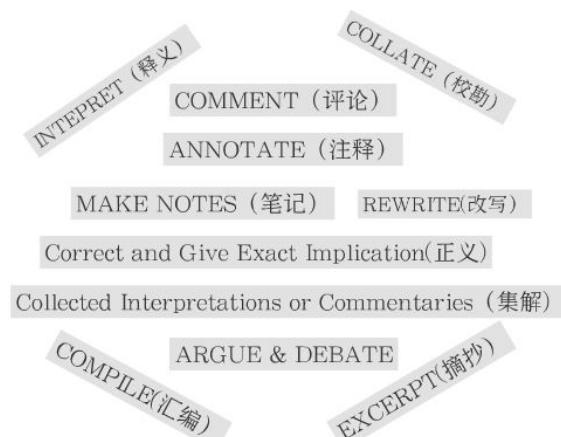
Timeline

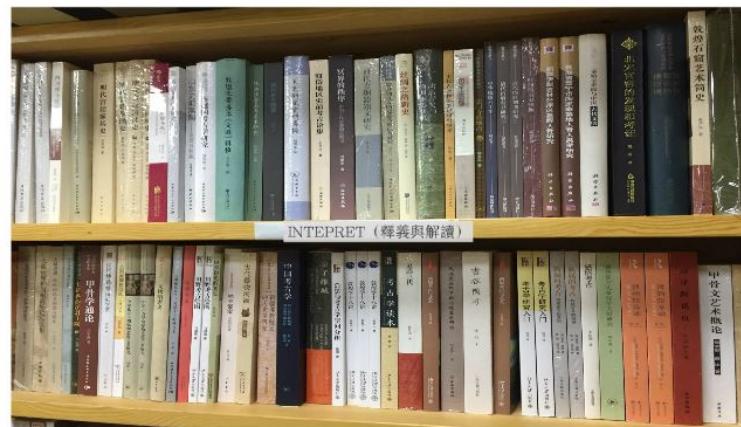


Timeline

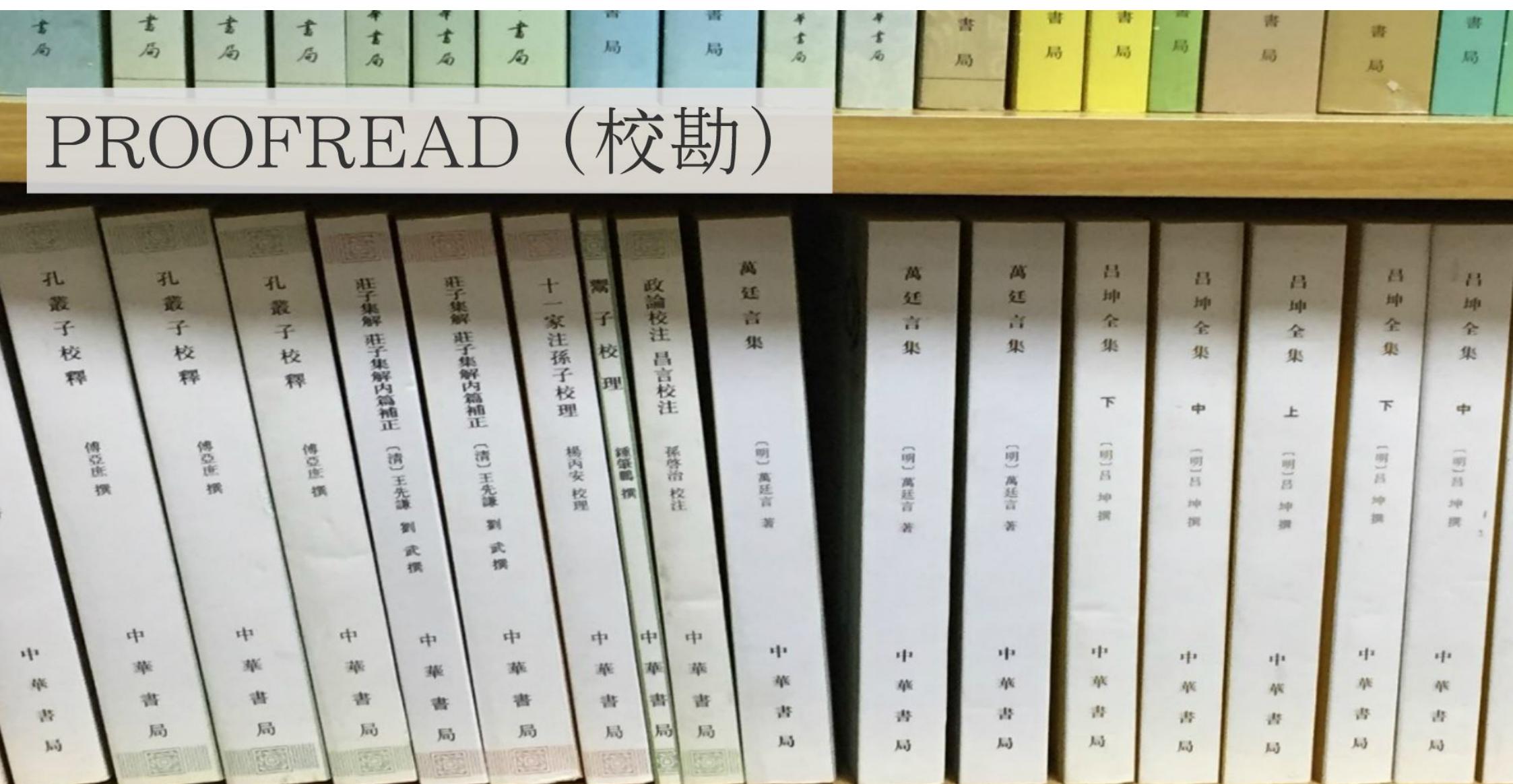


Timeline





PROOFREAD (校勘)



楊培均 陳長士著

浙江古籍出版社

Collect & Compile(彙編)



INTEPRET (釋義與解讀)

中国考古学

旧石器时代晚期
到早期青铜时代

刘莉
陈星灿著

宗子维城

从考古学看商周分封制度
与周朝 500 年的中国社会

GIS 与考古学空间分析

张勇著

北京大学出版社

敦煌学十八讲

樊锦诗著

北京大学出版社

敦煌学十八讲

樊锦诗著

北京大学出版社

考古学读本

杨桂军著

北京大学出版社

追迹三代

孙华伟著

追迹三代

西域考古与艺术

林树村著

北京大学出版社

民族史视野下的北魏墓志研究

刘连香著

北京大学出版社

书谷偶考

沈乃文著

上海古籍出版社

西域考古与艺术

林树村著

北京大学出版社

考古学研究入门

陈淳著

北京大学出版社

考古学研究入门

陈淳著

北京大学出版社

族属的考古
构建古今的身份

李伯谦著

上海古籍出版社

族属的考古
构建古今的身份

陈淳著

上海古籍出版社

感悟考古

李伯谦著

上海古籍出版社

清华简与先秦经学文献研究

姚小鸥主编

上海古籍出版社

R 博物馆基础

MUSEUM BASICS

北京大学出版社

R 博物馆基础

MUSEUM BASICS

北京大学出版社

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

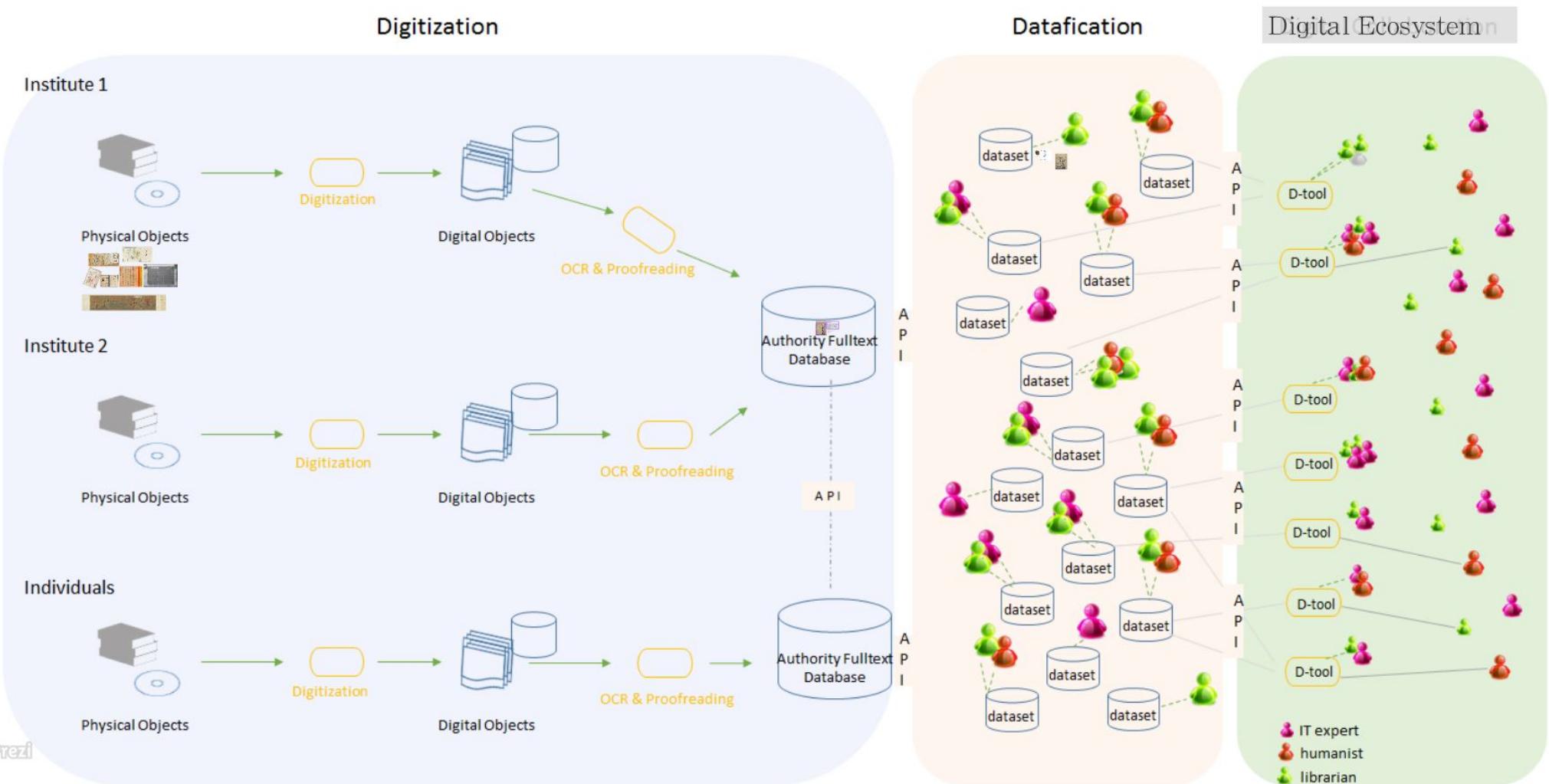
DHCI = Digitization + Datafication + Digital Ecosystem

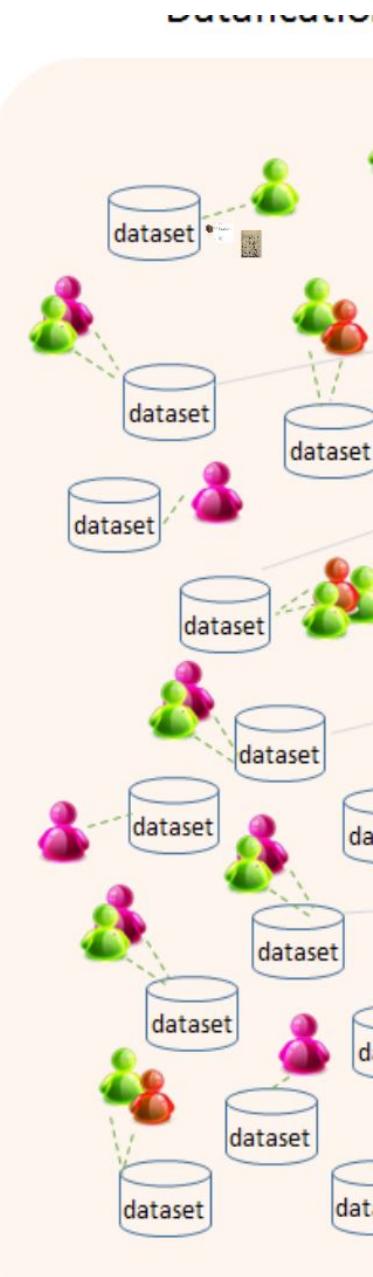
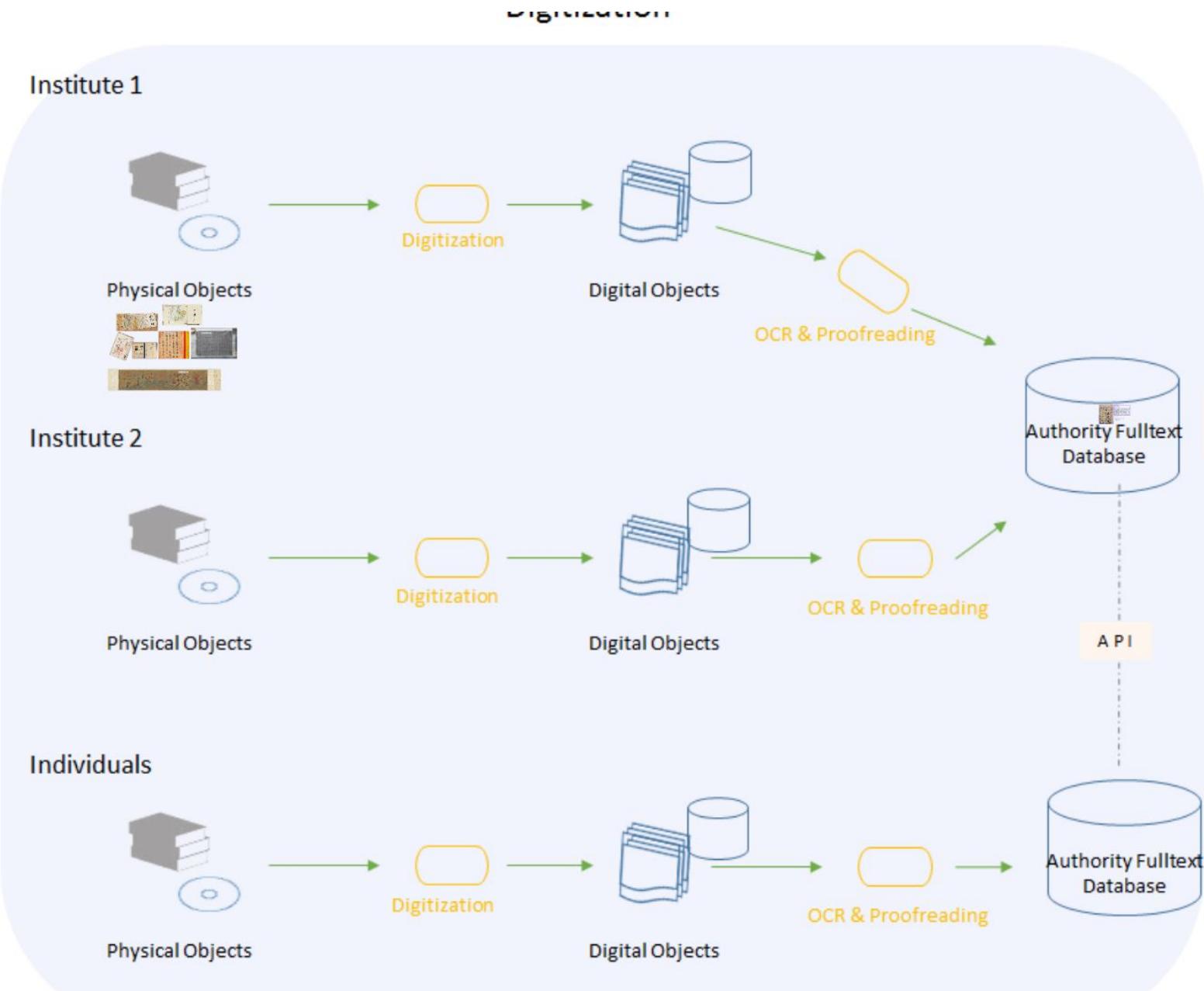
Digital Ecosystem



1. Conceptual Model in An Ecosystem Perspective

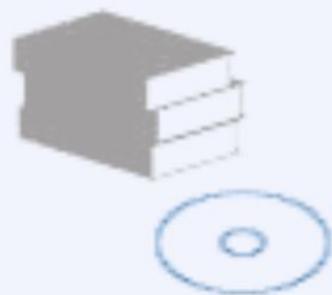
DHCI = Digitization + Datafication + Digital Ecosystem



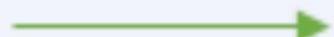


Digitization

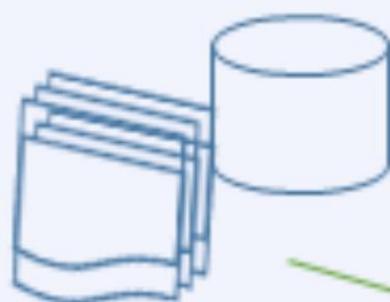
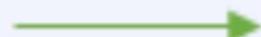
Institute 1



Physical Objects



Digitization



Digital Objects

OCR &

Institute 2

都黃曉峯校刊

會九萬一
二萬八千

山海經

槐蔭草堂藏版

弟子口
誦百此
母字口
法存之

稿題

山海經序

晉記室參軍郭璞撰

世之覽山海經者皆以其闕誕迂誇多奇怪似僞之言
莫不疑焉嘗試論之曰莊生有云人之所知莫若其所
知吾於山海經見之矣夫以宇宙之寥廓羣生之紛
繁陰陽之煦蒸萬殊之區分精氣渾清自相演薄遊魂
靈怪觸象而構流形於山川麗狀於木石者惡可勝言
子然則總其所以乖歧之於一響成其所以變混之於
一象世之所謂異未知其所以異世之所謂不異未知
其所以不異何者物不自異待我而後異異果在我非
物異也故胡人見布而疑賡越人見罽而駭毳夫觀所

卷之二十一

十一

有光明照二百里梵志種父名供友母字足
色子曰金剛集侍者曰寶受上首智慧弟乙

卷之二十一

印弗在廿時人壽

易五贊

原象

一肇判。陰降陽升。陽一以施。陰雨而承。

雲峰胡氏

畫前之易

此

則能施陰虛則能承乾日太一以其極至則日太極也
則未畫之前天行施與承雖見於乾坤已畫上而承其
皆兩而承其理可見陽之象也

Prezi

太

王羲之

七月帖

七月一日 羲之

得去月七日書

不可言 吾故羸

羲之帖

褚氏良

古月琴之帖

日但多念家事

自也多忙

諸君考之以得之

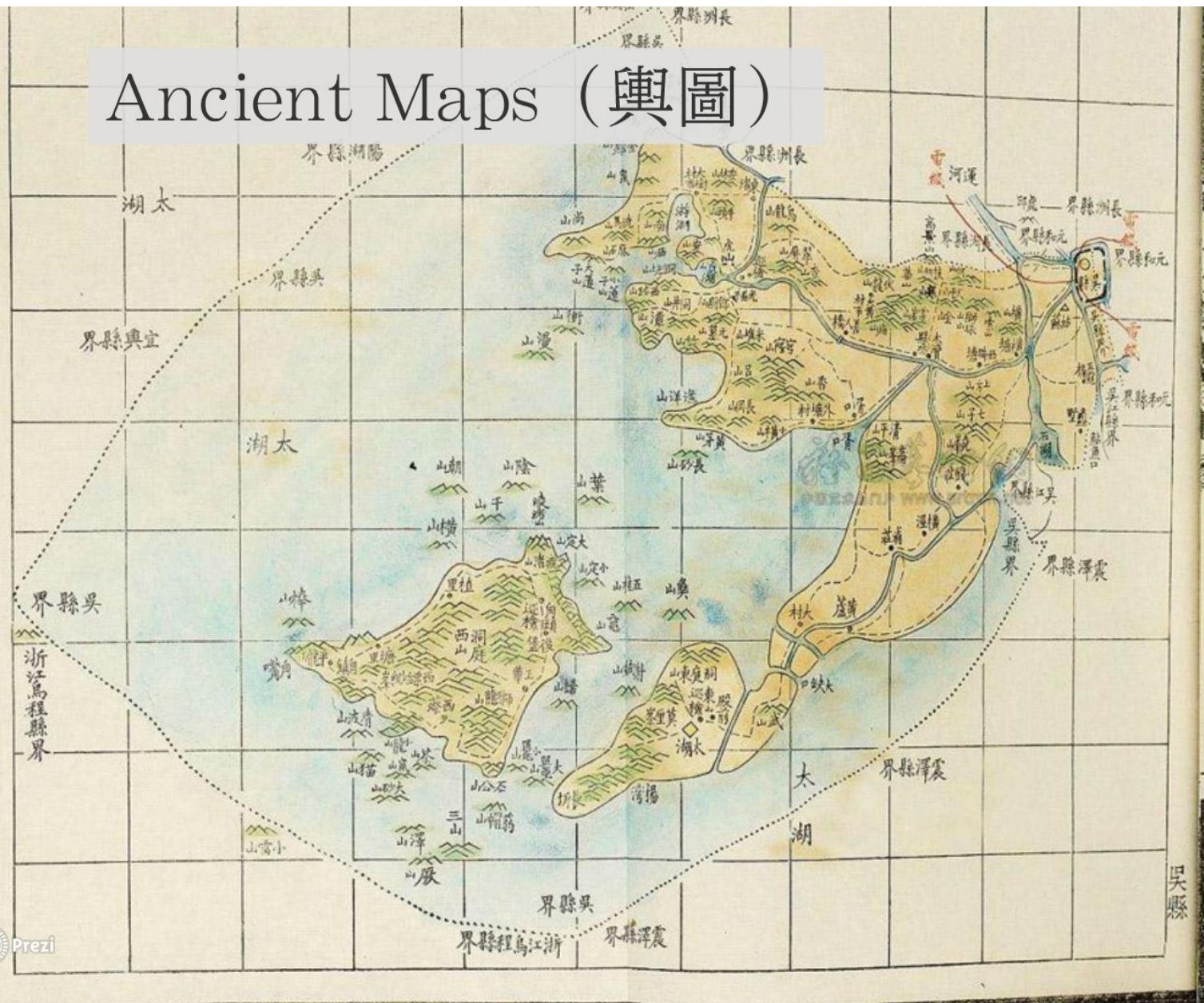
不亦可矣

王羲之

Calligraphy (書法)

江
山
湖
全
圖

Ancient Maps (輿圖)



Archives (檔案)



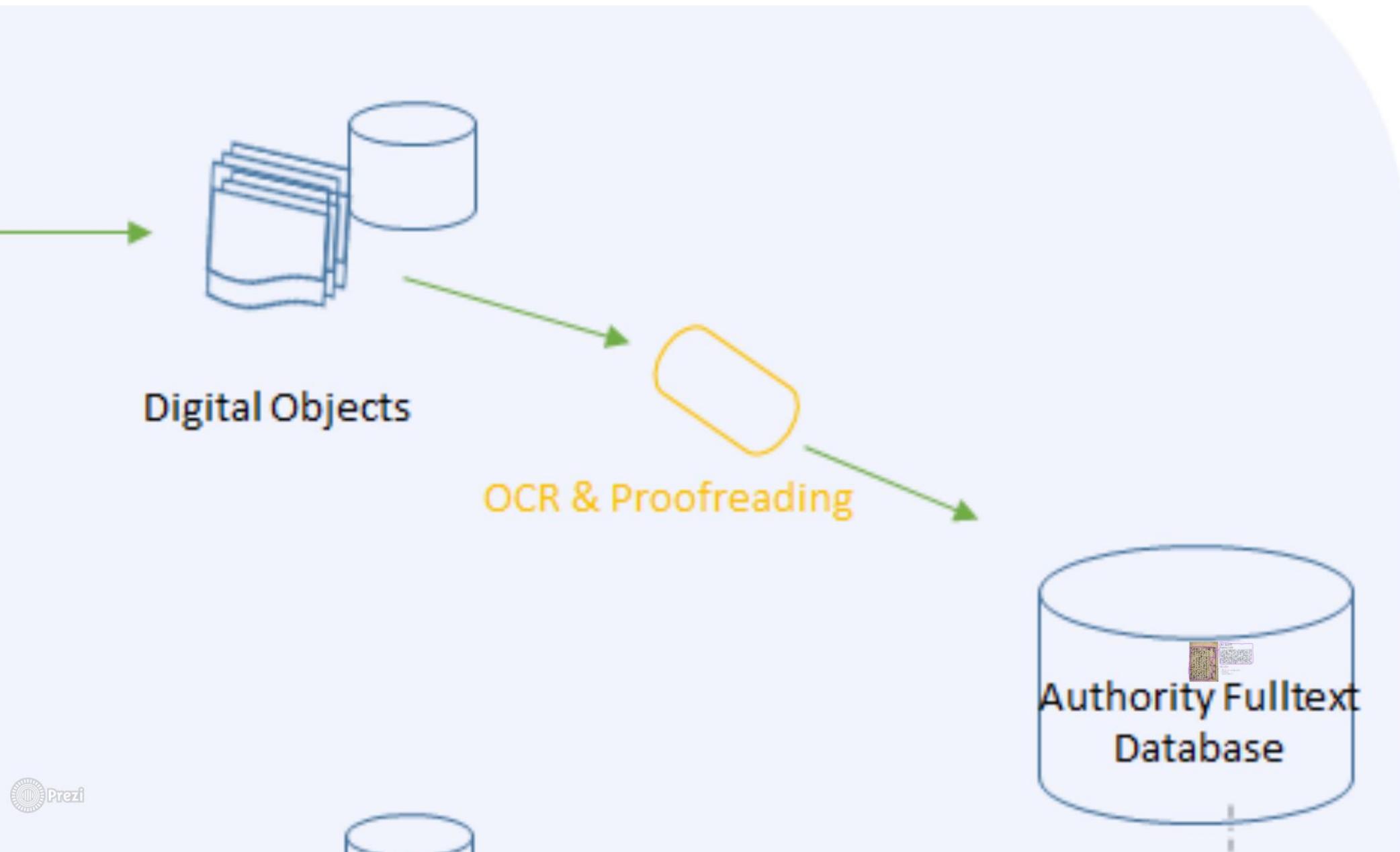
家庭概况表									
被保人姓名									
姓	名	性別	年齡	職業	地址	關係	被保人	被保人	被保人
陳	立夫	男	三十	中央組織部部長	上海	親屬	陳立夫	陳立夫	陳立夫
胡	若謙	女	二十一	國府隆高處	南京	配偶	胡若謙	胡若謙	胡若謙
戚	培植	女	二十	中央組織部部長	南京	配偶	戚培植	戚培植	戚培植
佳	均格	女	二十二	國府隆高處	南京	配偶	佳均格	佳均格	佳均格

唐周文矩含樂圖無上神品

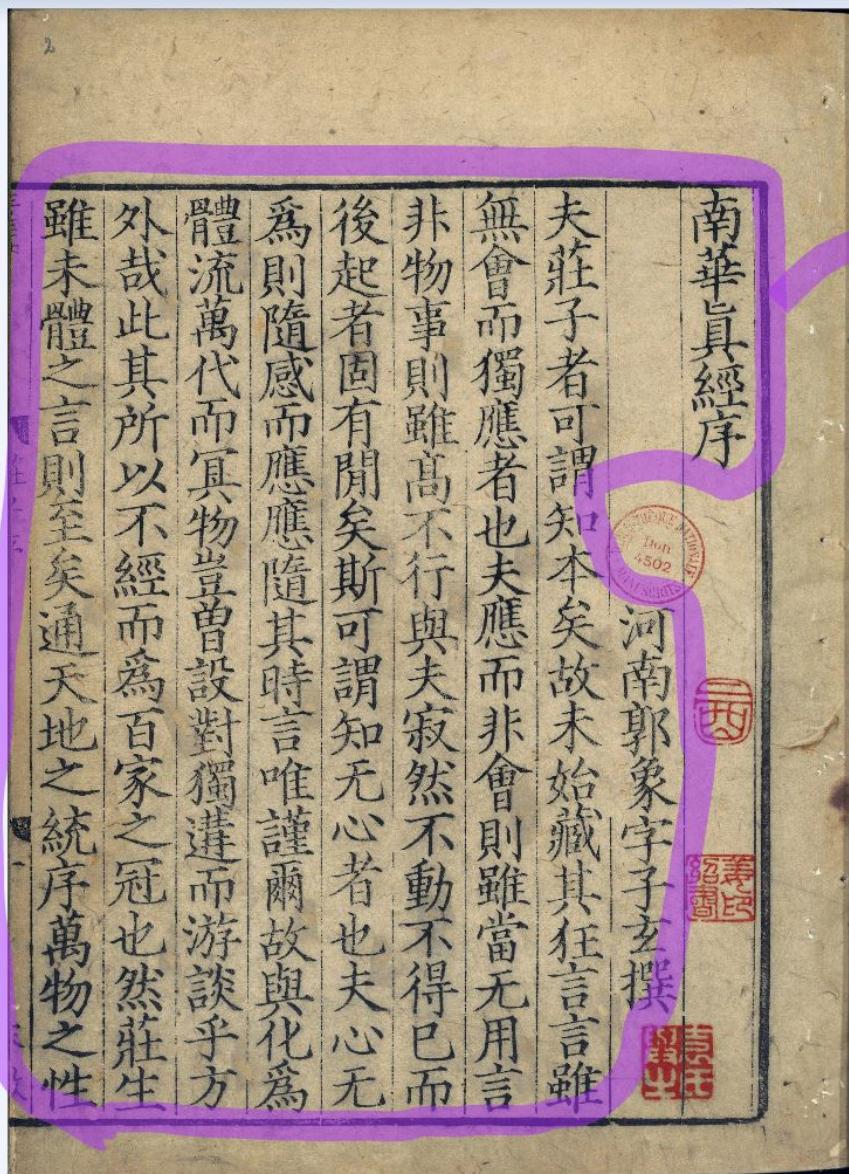
也

Paintings (繪畫)





A
P
I



南華真經序

河南郭象子子玄撰

雖言而無為方生
其狂言用己心化乎莊
言無得夫與談然
藏其始會則不動也
而獨應者也非寂然故
夫行與可謂時設對百家之冠也
非物事則雖高不行與夫寂然不動不得已而
後起者固有間矣斯可謂知无心者也夫心无
爲則隨感而應應隨其時言唯謹爾故與化爲
體流萬代而冥物豈曾設對獨遘而游談乎方
外哉此其所以不經而爲百家之冠也然莊生
雖未體之言則至矣通天地之統序萬物之性

Make sure:

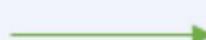
- Full text is proofread
- correct
- authoritative



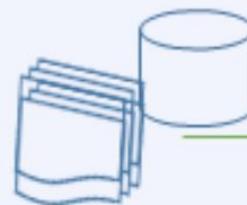
Institute 2



Physical Objects



Digitization



Digital Objects

OCR & Proofreading



OCR & Proofreading



Authority Fulltext
Database

A
P
I



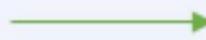
Authority Fulltext
Database

A
P
I

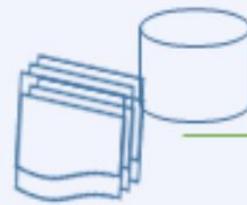
Individuals



Physical Objects



Digitization



Digital Objects

OCR & Proofreading

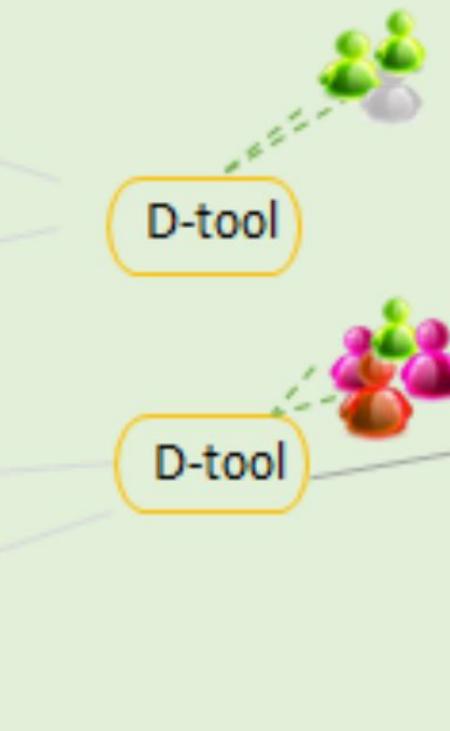
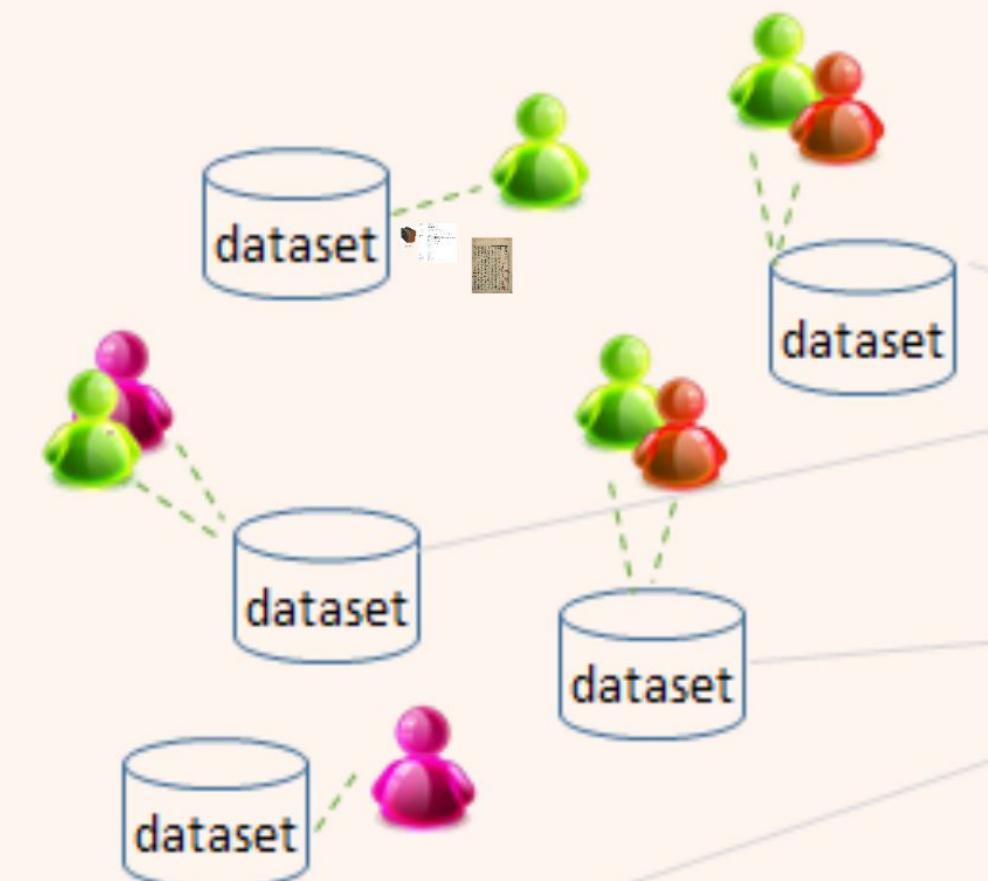


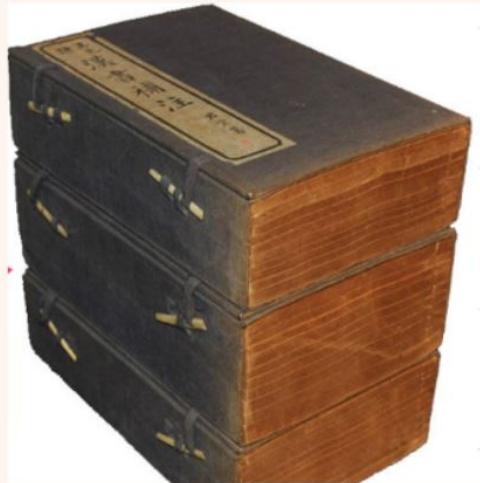
Authority Fulltext
Database

A
P
I

Datafication

Digital E





catalog

Vernacular title	南華真經 : 10卷
Title	Nan hua zhen jing : 10 juan
Attribution	郭象注 ; 陸德明音義. Guo Xiang zhu ; Lu Deming yin yi.
Published	[China : 顧春世德堂, 明嘉靖12 i.e. 1533] [China : Gu Chun Shi de tang, Ming Jiajing 12 i.e. 1533]
Description	12 v.
Series / collection	Harvard-Yenching Library Chinese Rare Books Digitization Project-Philosophy >
Language	Chinese
Notes	明嘉靖12年[1533]顧春世德堂刊《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」, 框高 19.8 x13.5. Hart collection, in rare books room. <> Double leaves, oriental style, in case.
Subjects	莊子 -- Commentaries. > Zhuangzi -- Commentaries. > 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. -- >
Author / Creator	
Other title	六子書. Liu zi shu.
Creation Date	Ming Jiajing 12 i.e. 1533]
HOLLIS number	990077550710203941
Permalink	http://id.lib.harvard.edu/almna/990077550710203941/catalog
Source	HVD - Core

南華真經序

西

卷一

卷二

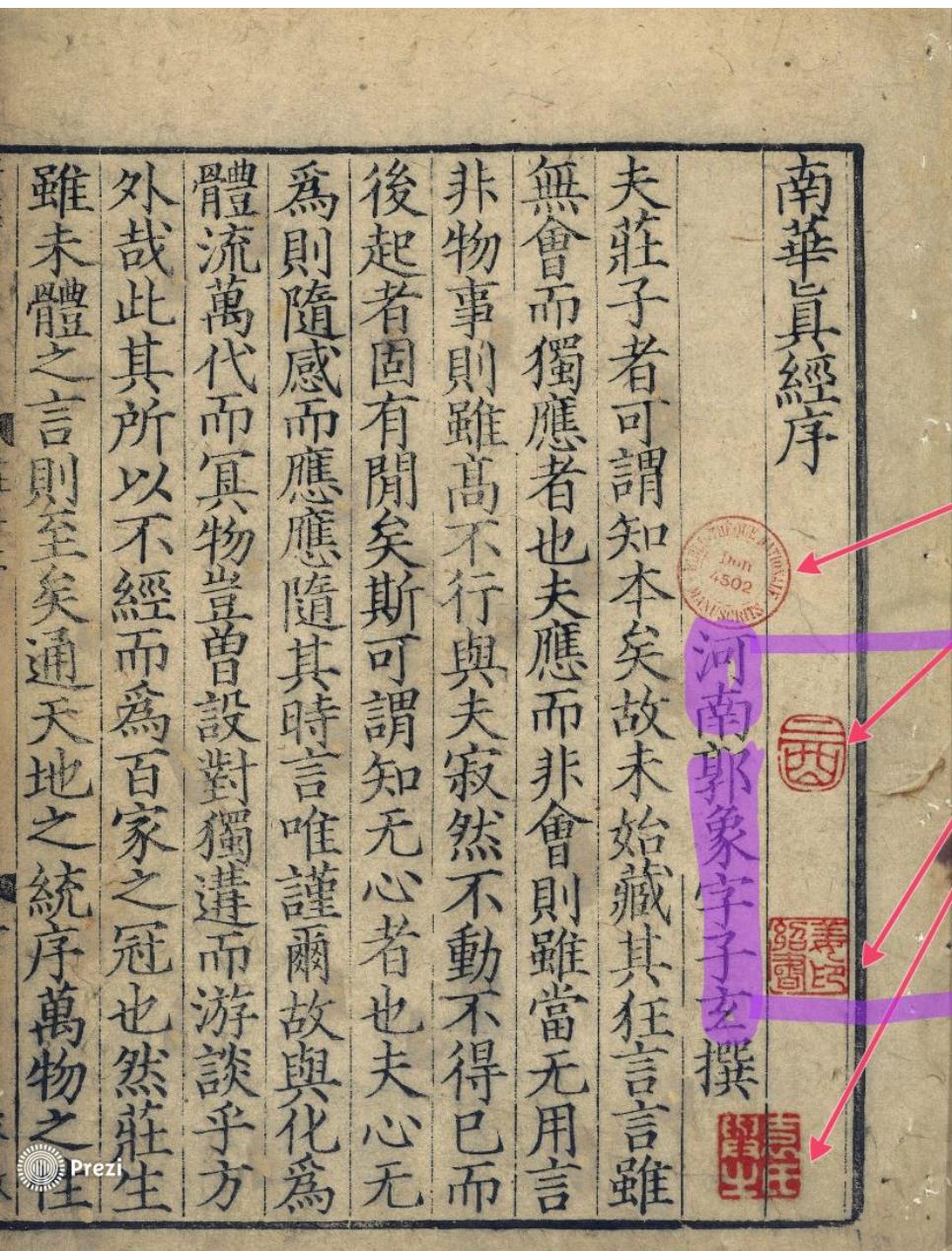


河南郭象字子玄撰

卷三

卷四

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无用言非物事則雖高不行與夫寂然不動不得已而後起者固有閒矣斯可謂知无心者也夫心无爲則隨感而應應隨其時言唯謹爾故與化爲體流萬代而冥物豈曾設對獨達而游談乎方外哉此其所以不經而爲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



Seals Authority Dataset

- recognize the characters
- created time
- belongs to whom
- meaning
- size
- appearance
- related works

Geo-name Authority Dataset

- name
- alternative name
- start time
- end time
- upper level
- longitude
- latitude
- kml file

Time Authority Dataset

Personage Authority Dataset

Official Title Authority Dataset



- character (印面文字)
- image (印文圖片)
- owner (印章主人)
- sculptor (治印者)
- source (來源)
- dynasty (時間)
- pinyin (印文拼音)
- abstract (提要)
- palindrome (是否回文印)
- character_order (印文字序)
- official_or_private (官印還是私印)
- moral (寓意)
- seal_shape (印面形狀)
- seal_font (印面字體)
- seal_character_type (印面文字類型)
- Concave_or_Convex (陰文還是陽文)
- seal_outline (邊欄)
- seal_size (印面大小)
- Seal_material (印章材質)
- Seal_handle_Appearance (印章鈕式)
- Seal_Edge (印章邊欄)
- seals_on_same_work (鈐同一作品上之其他印)
-
-

昵晏樓



張大千

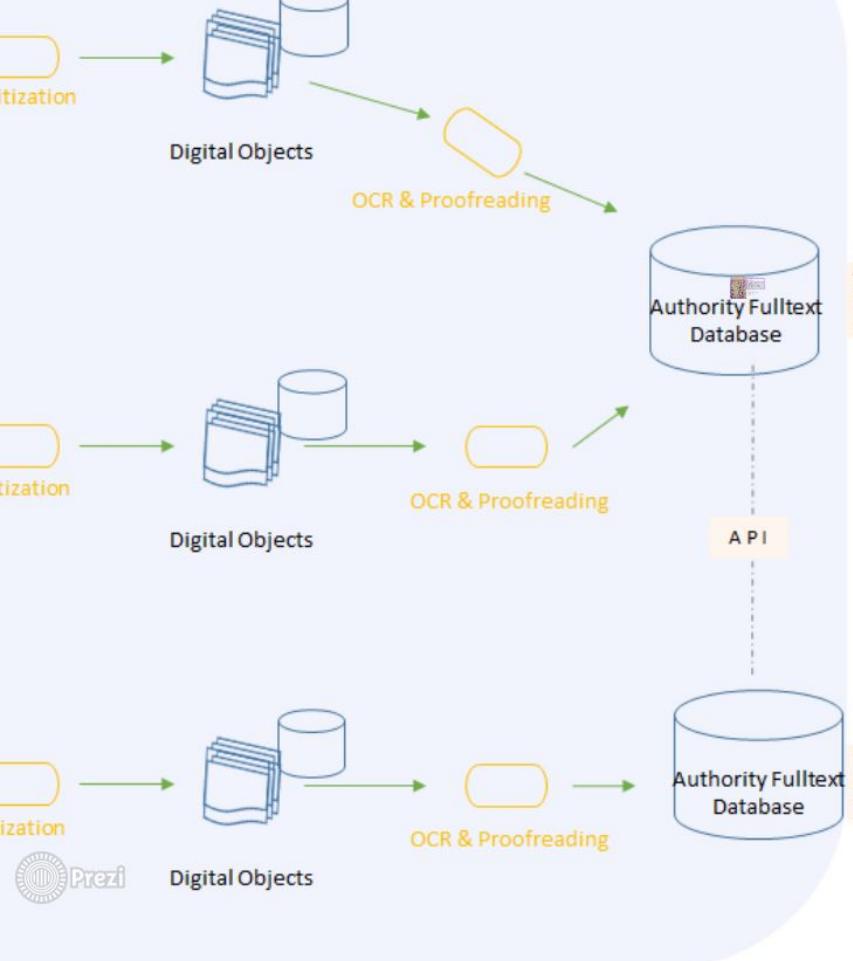
曾紹傑

從繪畫中析出

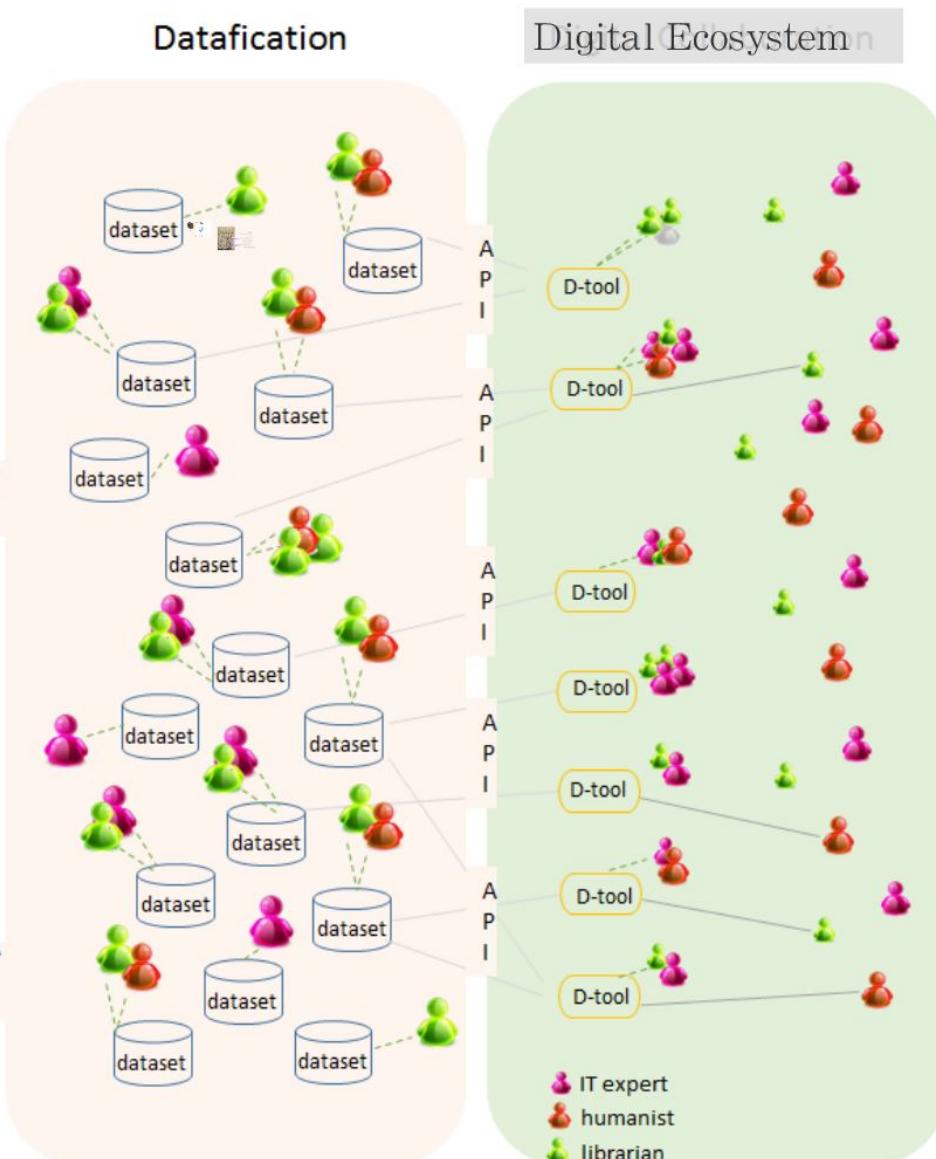
晚清民國

= Digitization + Datafication + Digital Ecosystem

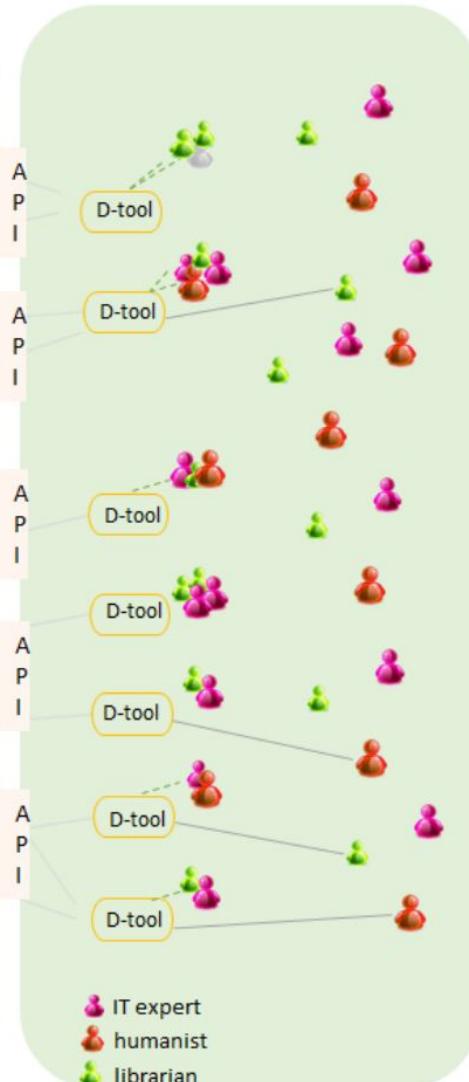
Digitization



Datafication

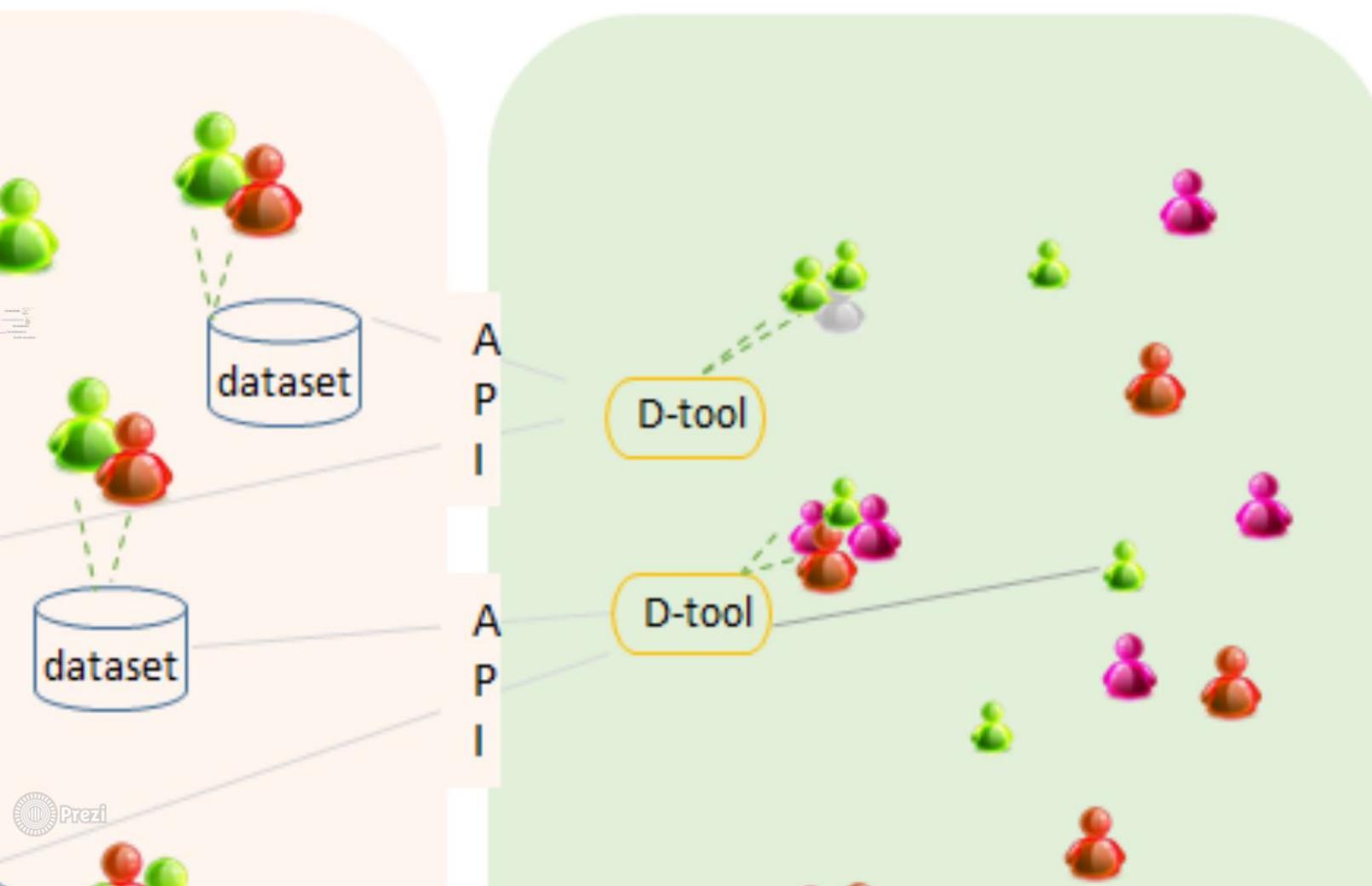


Digital Ecosystem

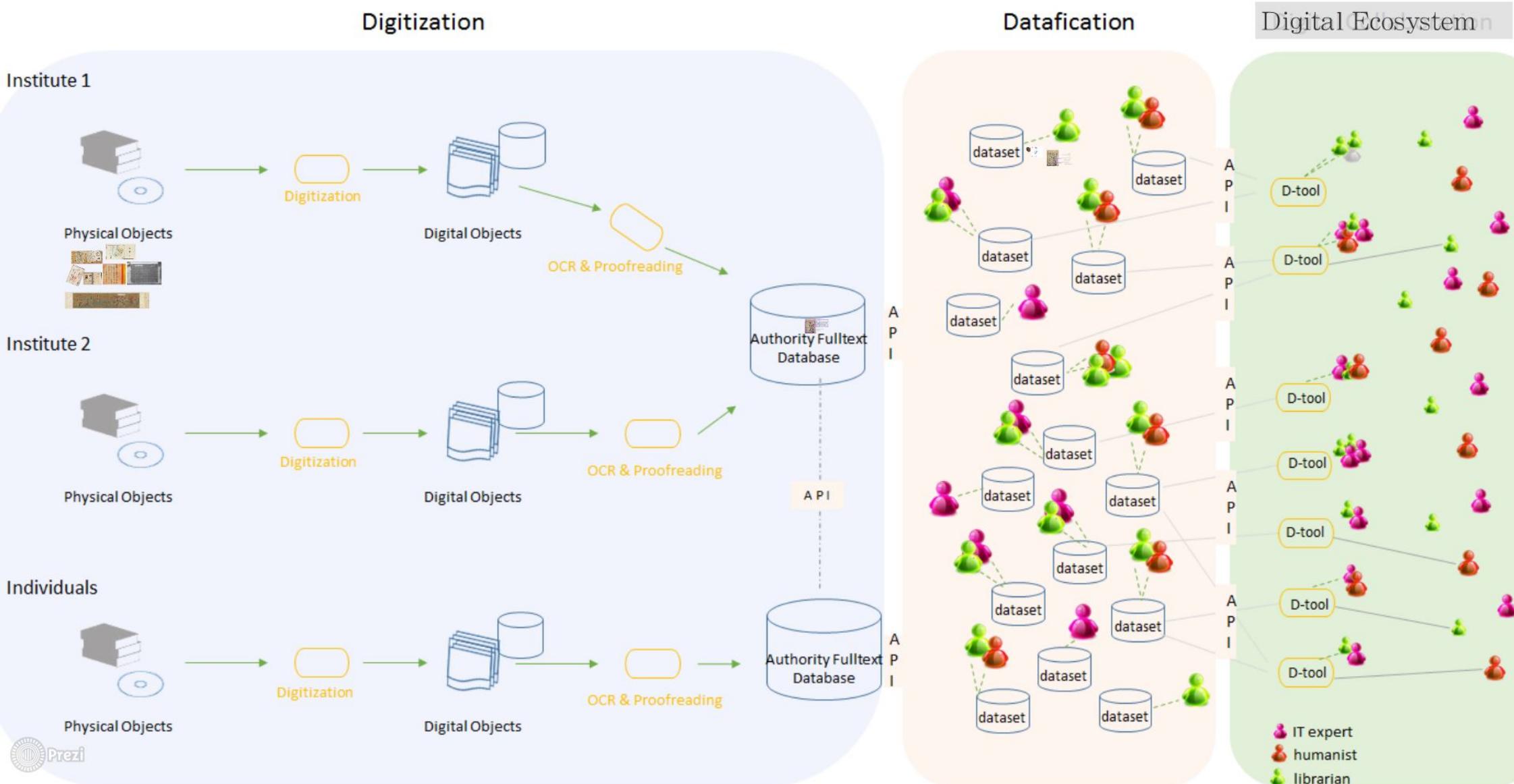


ification

Digital Ecosystem



DNUCI = DIGITIZATION + DATAFICATION + DIGITAL ECOSYSTEM



2. Imaginary Scenarios

- Libraries, museums, archives, and galleries
- Humanists
- Make traditional research more easier
- Research paradigm can be changed



宣統鑑賞



乾隆御覽之寶

此印為「乾隆五璽」之一。所謂「乾隆五璽」，是指乾隆皇帝鈐蓋在清內府藏品上皇帝收藏印，按通常的說法，這五璽的印文分別是「乾隆御覽之寶」、「石渠寶笈」、「三希堂精鑒璽」、「乾隆鑒賞」、「宣子孫」。



中國歷代印鑑規範庫

中國研究數字人文網絡基礎設施 (DIGITAL HUMANITIES CYBERINFRASTRUCTURE FOR CHINA STUDIES)

宣統御覽之寶

治印時間 ▾ 印文用於表征 ▾ 寶體 ▾ 官私印 ▾ 形制 ▾ 文字類型 ▾ 雕刻手法 ▾ 印章材質 ▾ 印章外形 ▾ 關於中國歷代印鑑數據庫

王羲之

此印為「乾隆五璽」之一。所謂「乾隆五璽」，是指乾隆皇帝鈐蓋在清內府藏品上皇帝收藏印，按通常的說法，這五璽的印文分別是「乾隆御覽之寶」、「石渠寶笈」、「三希堂精鑒璽」、「乾隆鑒賞」、「宣子孫」。



乾隆鑑賞

此印為「乾隆五璽」之一。所謂「乾隆五璽」，是指乾隆皇帝鈐蓋在清內府藏品上皇帝收藏印，按通常的說法，這五璽的印文分別是「乾隆御覽之寶」、「石渠寶笈」、「三希堂精鑒璽」、「乾隆鑒賞」、「宣子孫」。



三希堂精鑑璽

此印為「乾隆五璽」之一。所謂「乾隆五璽」，是指乾隆皇帝鈐蓋在清內府藏品上皇帝收藏印，按通常的說法，這五璽的印文分別是「乾隆御覽之寶」、「石渠寶笈」、「三希堂精鑒璽」、「乾隆鑒賞」、「宣子孫」。



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Shangdong
-

寓意moral

- 師承紀學
- 世家傳承
- 淵源來歷
- 經歷紀事
- 寓言譬喻

姓名字號names

- 姓名
- 字
- 號
- 諱



EXPORT

To make drawing historical atlas more easier

export

Which government

清 (Qing)

Time point1

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 interoperate with APIs.
- Redefined metadata scheme in a DH perspective
 - for existed catalogue of common literature
 - for content annotation
- Timeline associated with geography significance

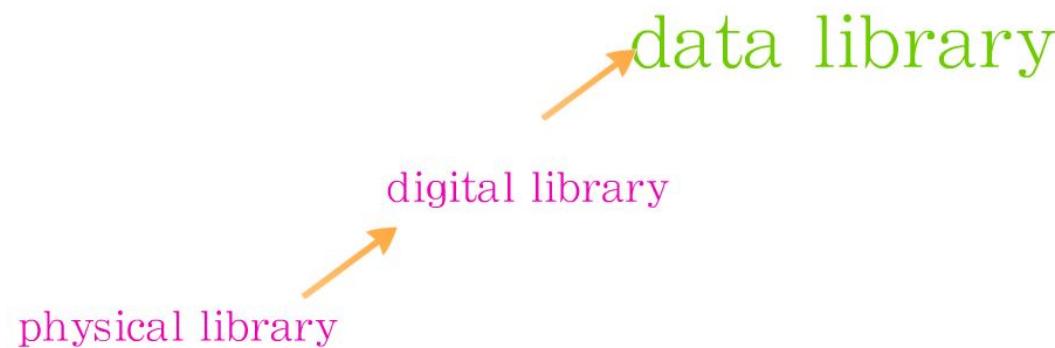
histchina.cn

IV

Future of Academic Library

2. Library Trends : Data Library

1990s VS. 2020s



[--]

Thank You !

Libraies should play an important role in DHCI

Change as soon as possible!

Thank You !