A Cyberinfrastructure for Studying Chinese History

A Proposal Based on the Experience of the China Biographical Database Project (CBDB)

LIK HANG TSUI 徐力恆

China Biographical Database

CUHK DS symposium, 31 March 2017

(Thanks to the help of Peter Bol, Liu Chao-lin, Wang Hongsu)

Biographies in China

China has over 2000 years of elite biography

There are perhaps more than 250,000 biographies in the historical record: including biographies in dynastic histories, local gazetteers, tomb inscriptions, etc. Local gazetteers (Song-Qing) mention at least 2 million people!

Chinese history is an important source for the history of humanity

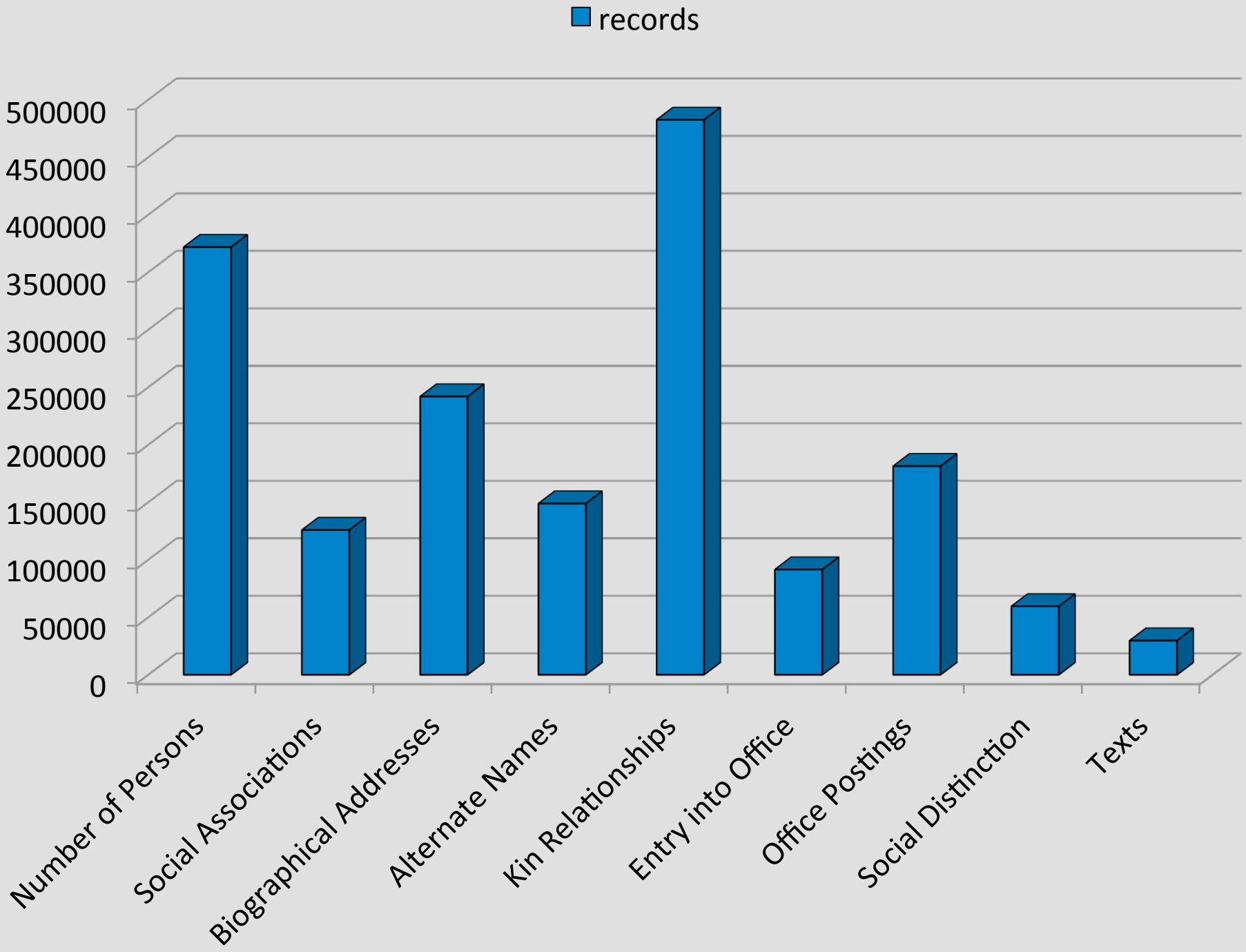
So, how do we organize and study this information with modern day technology?

Jointly owned by (since 2005):

Fairbank Center (Harvard), 中國古代史研究中心 (Peking Univ.), 史語所 (Academia Sinicia)

Free and open for academic use

http://projects.iq.harvard.edu/chinesecbdb


records
500000
450000
400000
350000
300000
250000
200000
150000
100000
50000
0
Number of Persons
Social Associations
Biographical Addresses
Alternate Names
Kin Relationships
Entry into Office
Office Postings
Social Distinction
Texts

Symposium

9:00 a.m. - 12:35 p.m.

GIS and Big Data for Urban Applications

Prof. HUANG Bo
Department of Geography and Resource Management
Faculty of Social Science

My presentation

1. A cyberinfrastructure for Chinese history: what and why?
2. Resource sharing mechanisms
3. Communication between participating actors

A Cyberinfrastructure for Historical China Studies (Hongsu Wang, Lik Hang Tsui, Peter K. Bol)

服務於中國歷史研究的網絡基礎設施

王宏甦[1]、徐力恆[2]、包弼德[3]

（第七屆數位典藏與數位人文國際研討會論文，2016 年 12 月，臺北）

摘要

數據庫、研究項目數量和參與中國文史數位研究的人員大幅增加，使得為中國歷史研究建立相應的網路基礎設施變得必要。網絡基礎設施可以起的作用在於連接對一個學科有用的電腦軟件、數據集、人才、實務做法、標準和合作模式，促進研究的進步。本文將具體論述為何要營建中國歷史研究的網絡基礎設施，以及如何從資源的共享和成員的交流兩方面實現這個目標。

關鍵詞：網絡基礎設施、中國歷史、數位人文

A Cyberinfrastructure for Historical China Studies

Hongsu Wang, Lik Hang Tsui, Peter K. Bol

(Paper for the 7th International Congress of Digital Archives and Digital Humanities, Taipei, Taiwan, Dec. 2016)

Abstract

The proliferation of databases for the study of Chinese history and the increasing numbers of researchers taking part in their development calls for a cyberinfrastructure. A cyberinfrastructure can be conceived as a network of discipline-specific software applications and data collections and also of the personnel and the set of best practices, standards, and collaborative methods they establish. This paper discusses how participants in such a cyberinfrastructure for historical China studies can share their resources and how their communication can be facilitated by various technologies and mechanisms.

[1] 哈佛大學「中國歷代人物傳記資料庫」項目經理（Project Manager, CBDB, Harvard University）。電郵為：hongsuwang@fas.harvard.edu
[2] 哈佛大學「中國歷代人物傳記資料庫」博士後研究員（Postdoctoral Fellow, CBDB, Harvard University）。電郵為：tsui01@fas.harvard.edu
[3] 哈佛大學副教務長、東亞語言與文明系查理斯·卡威爾（Charles H. Carswell）講座教授（Vice Provost for Advances in Learning and the Charles H. Carswell Professor of East Asian Languages and Civilizations, Harvard University）。電郵為：peter_bol@harvard.edu

Keywords: Cyberinfrastructure, Chinese history, digital humanities

一、引言：中文數位人文網絡基礎設施的實現

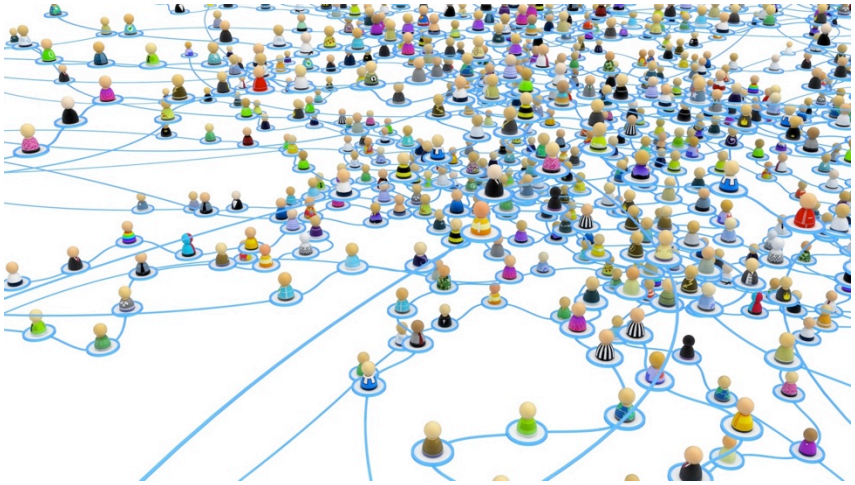
美國學術團體協會（ACLS）在 2005 年發佈的研究報告《我們的文化共同體》（Our Cultural Commonwealth）提出人文、社會科學應該像自然科學研究一般，有自己的網絡基礎設施（cyberinfrastructure）。[4] 網絡基礎設施的層次介於基礎科技和具體用於某研究項目、某學科和實踐的特定科技之間，可以說處於中層。它可以起的獨特作用，在於連接對一個學科有用的電腦軟件、數據集、人才、實務做法、標準和合作模式。[5] 我們希望在本文處理的議題如下——我們為何要在中國歷史研究領域營建這種網絡基礎設施，以及如何實現這個目標。

和自然科學相比，人文學科和部分社會科學學科（尤其是其中量化特點不明顯的學術領域）深深浸淫在語言之中，很受語言的特點影響。就以主題模型（topic modeling）為例，當這方法用於中國文史研究時會面對頗多挑戰。一般使用這種研究方法時，認定每個詞之間的空格就代表分詞的區隔，但中文文本的情況並非如此。例如，當機器閱讀「中華民國」這四個中文字時，它可以認定那是四個詞，也可以是兩個各為兩個字的詞，也可以是一個四字詞。怎麼讓機器獲得判斷的能力，需要人文學者利用他們的知識介入。

因此，這意味著我們必須深入調查、瞭解語言的特點（例如是古代漢語、現代漢語），才能建立專門用於中國歷史研究的數位人文工具。如果不正視這一點，數位研究計劃之間的溝通和連接將很難進行，數據的分享也會面對很多障礙，導致閉門造車的弊病。不過，近年學界對我們會議的主題——數位人文的興趣越來越濃，用於統計、社會網絡分析、地理空間分析和地圖繪製、文本標註和挖掘、製作主題模型，還有建立關係型數據庫（relational databases）或物件導向式數據庫（object-oriented databases）的電腦軟件如雨後春筍。過去很難用，或者很難獲得的軟件，現在已經是隨手可得，其操作也簡便許多了。項目數量和參與數位研究的人員大幅增加，也為中國研究數位人文建立相應的網路基礎設施開始有了成熟的條件。就如其他學科和區域研究的專家一樣，我們領域裡不少學者都感到應該開展這種工作，因此開始了本文的寫作。[6]

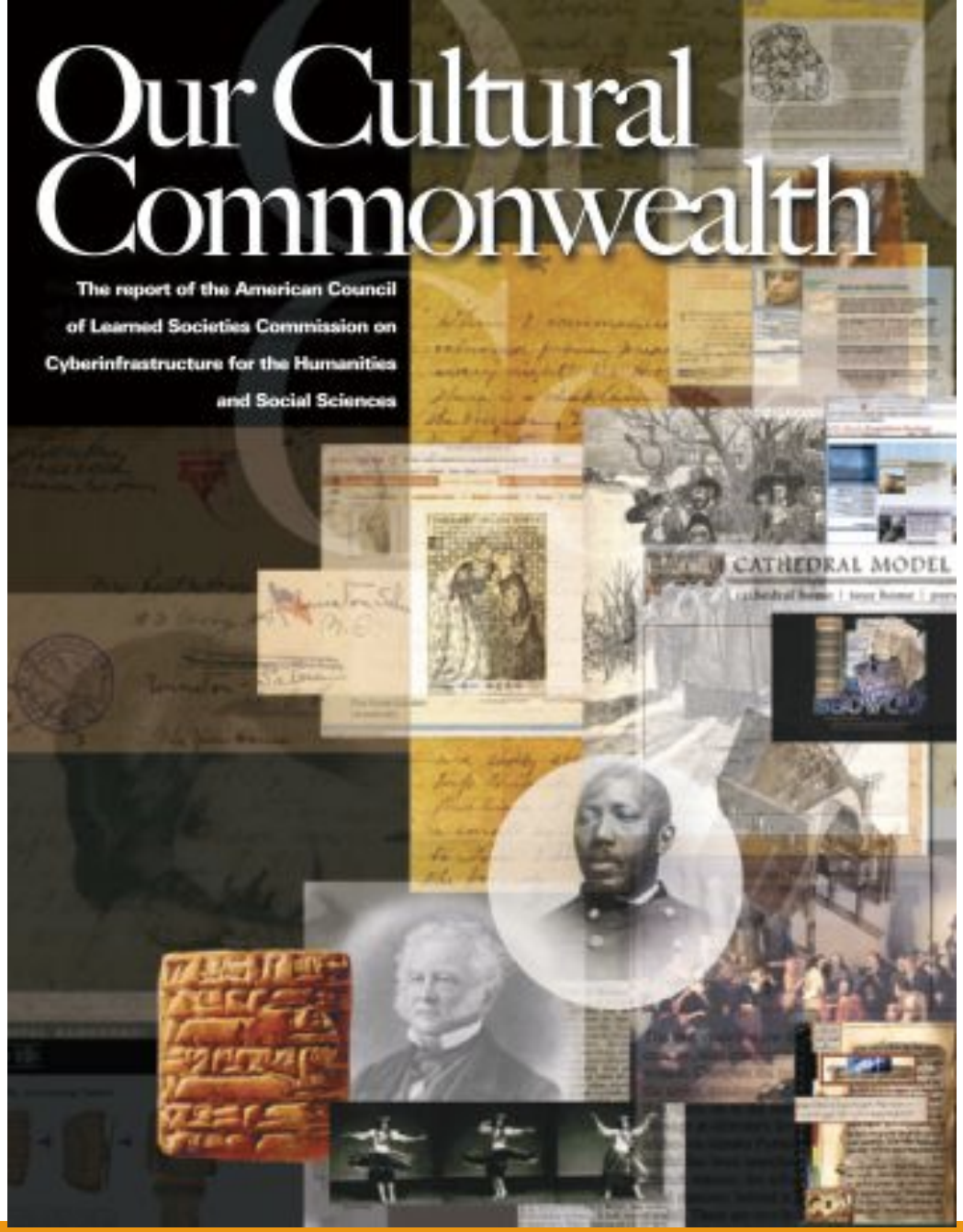
[4] 參閱：http://www.acls.org/cyberinfrastructure/OurCulturalCommonwealth.pdf。
[5] 參閱：https://www.nsf.gov/cise/sci/reports/atkins.pdf。
[6] 比較相關的是佛教研究數據的基礎設施，參閱 Christine L. Borgman, *Big Data, Little Data, No Data: Scholarship in the Networked* World (Cambridge, MA: The MIT Press, 2015), 186-200.

What is a cyberinfrastructure?



"…research funders use the term **cyberinfrastructure** to describe research environments that support advanced data acquisition, data storage, data management, data integration, data mining, data visualization and other computing and information processing services distributed over the Internet beyond the scope of a single institution."

(https://en.wikipedia.org/wiki/Cyberinfrastructure)



ACLS report (2005)

- Tangible:
  - network and means of storage in digitized form, discipline-specific software applications and project-specific data collections.

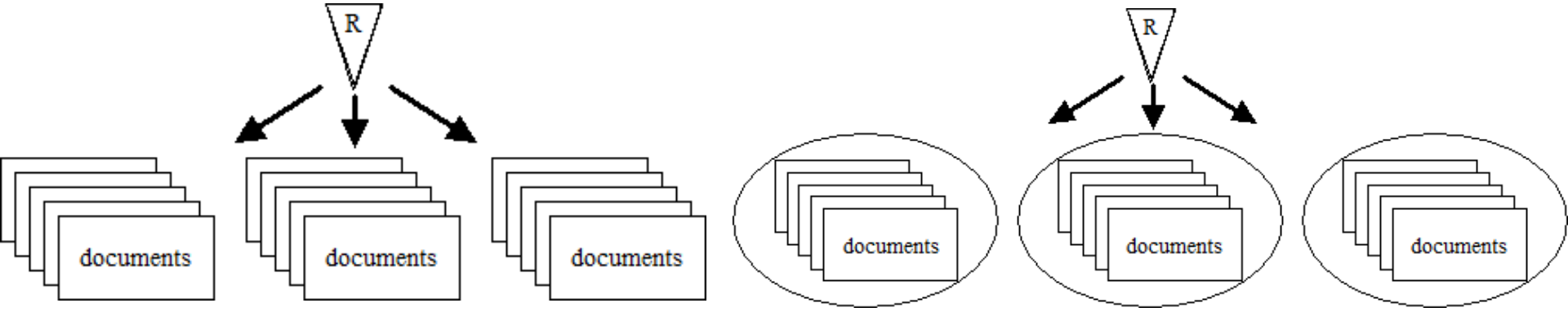
- Intangible:
  - layer of expertise and the best practices, standards, tools, collections and collaborative environments that can be broadly shared across communities of inquiry.

盡平上思田黃勝許與交趾表裏寇邊二十九年函梁率兵討之賊走象山山近交趾深林不可入乃度其出入列堠圍之徐伐山通道且戰且進二年拔其寨擒許走交趾國傑以巢地爲屯田募度遠諸獞人耕之以爲兩江蔽障蕭泰登字則平吉安路人至元三十一年僉廣西廉訪司事柳州左道謀叛論死者二百人泰登錄之而釋不知情者百三十人謝七人他所辨雪無算凡黜貪謬者二百一十人謝讓字仲和韻昌人大德中爲湖廣行省左右司郎中時廣西兩江岑雄黃勝許等屢相讐殺爲邊患讓謂此豈第可懷柔不宜力競寬其法以羈縻之使不至跳梁可也因書榜招諭以攜其黨也兒吉尼字尚文唐兀氏人至正末爲廣西行省平章政事兼廉訪使時紅巾擾攘吉尼以公費修築城池民不知勞洪武元年楊璟取廣西吉尼堅壁不下城破執送京師不屈死郡人感其德立廟祀之陳瑜字仲庸雷州人廣西中書省都事城破以佩刀自刎有劉永錫者潭州人與瑜同事率妻子溺於白龍池死瑜爲義兵千戶洪武元年明兵圍靜江僉事守西城城陷身中數鎗知不敵自

Traditionally…

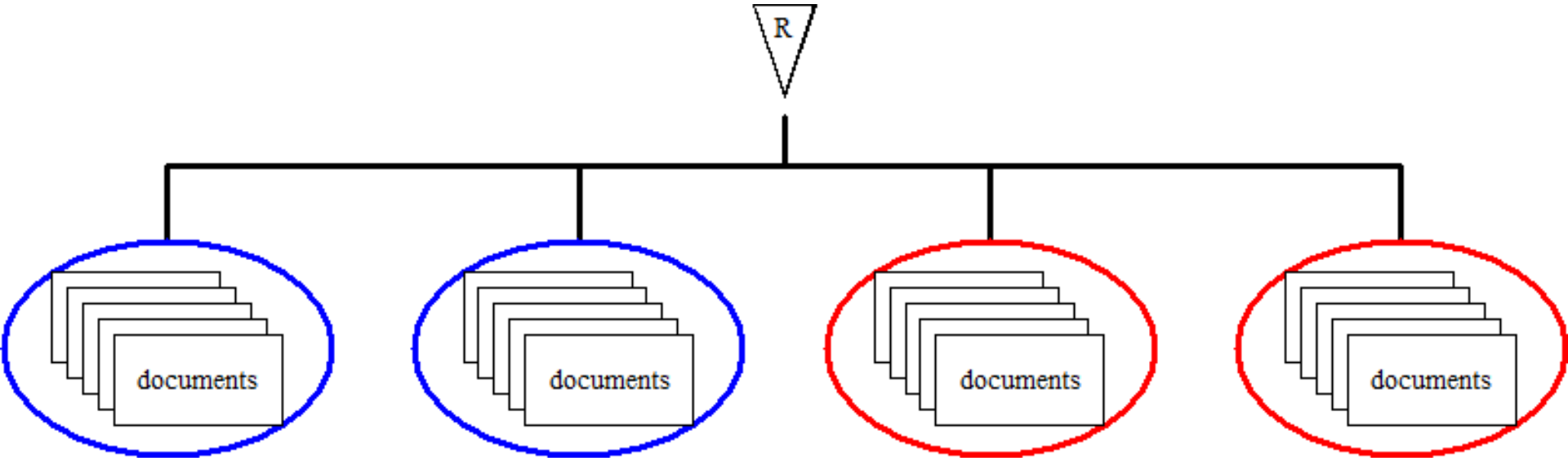
Researchers gather their own research material by themselves, through their own ways.

Sometimes they also relied on collections in libraries.



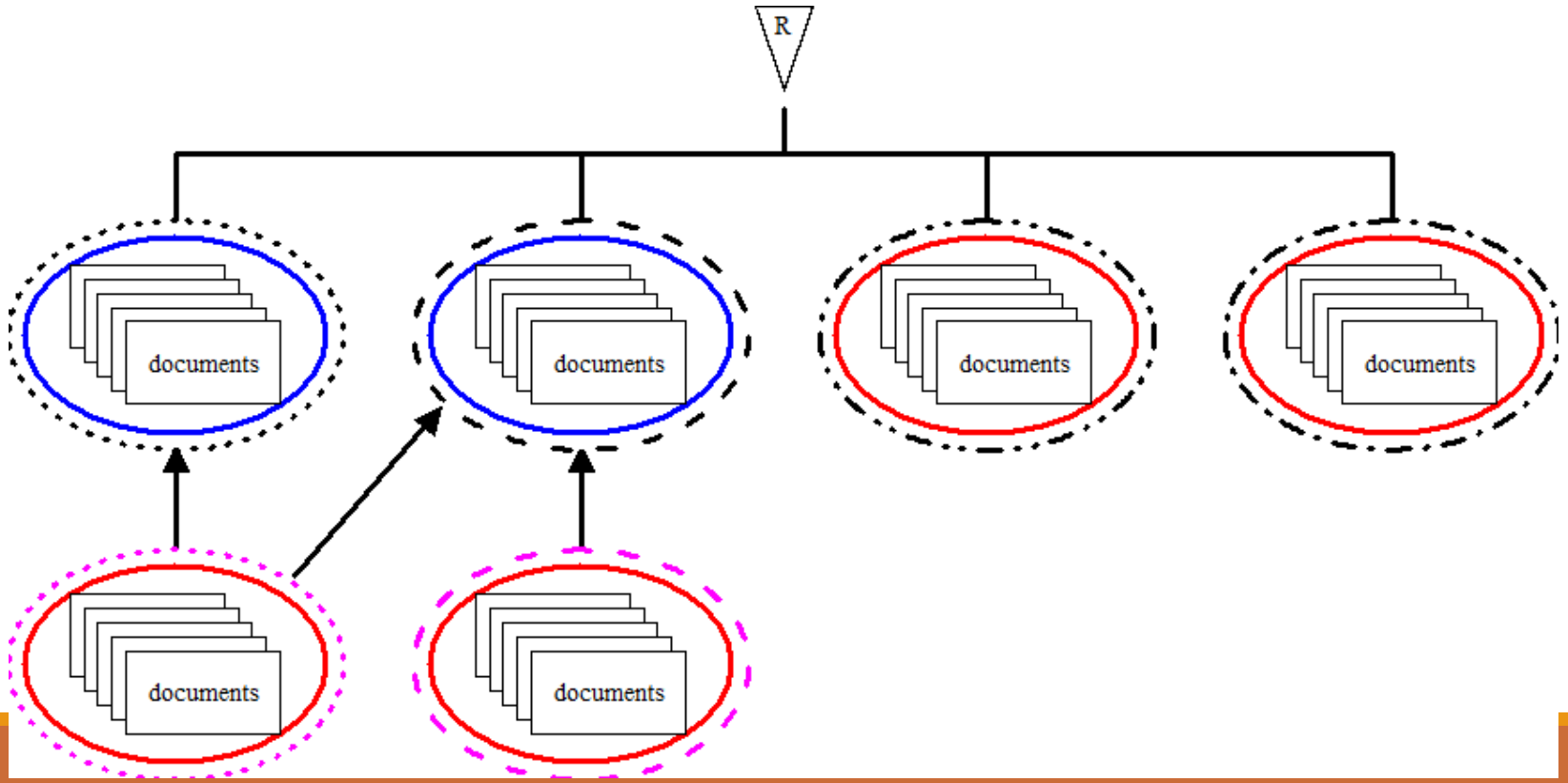
Then, with the Internet and the digital age…

Some materials are open, some are commercial, some are restricted

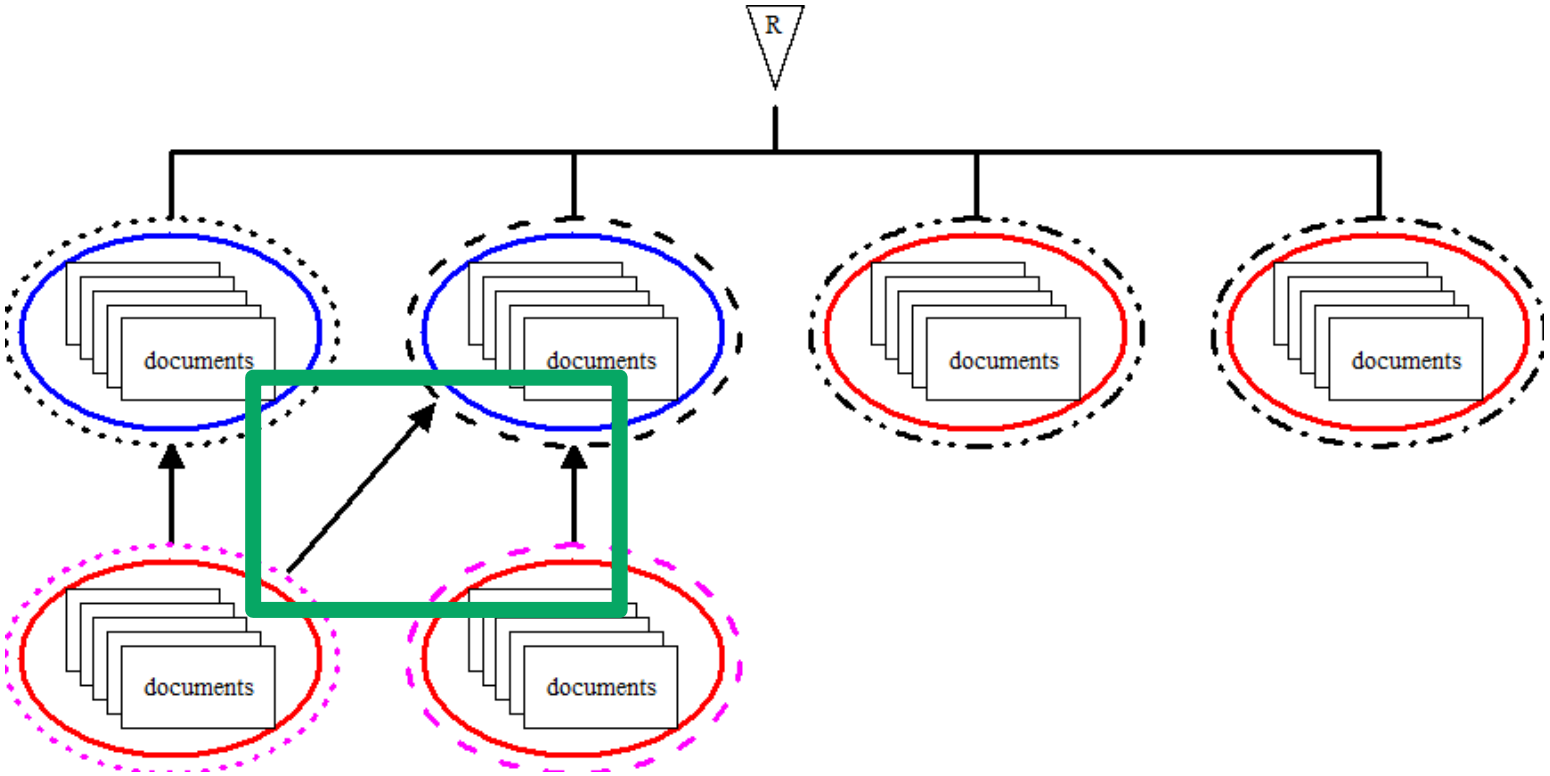


And usually more like this:

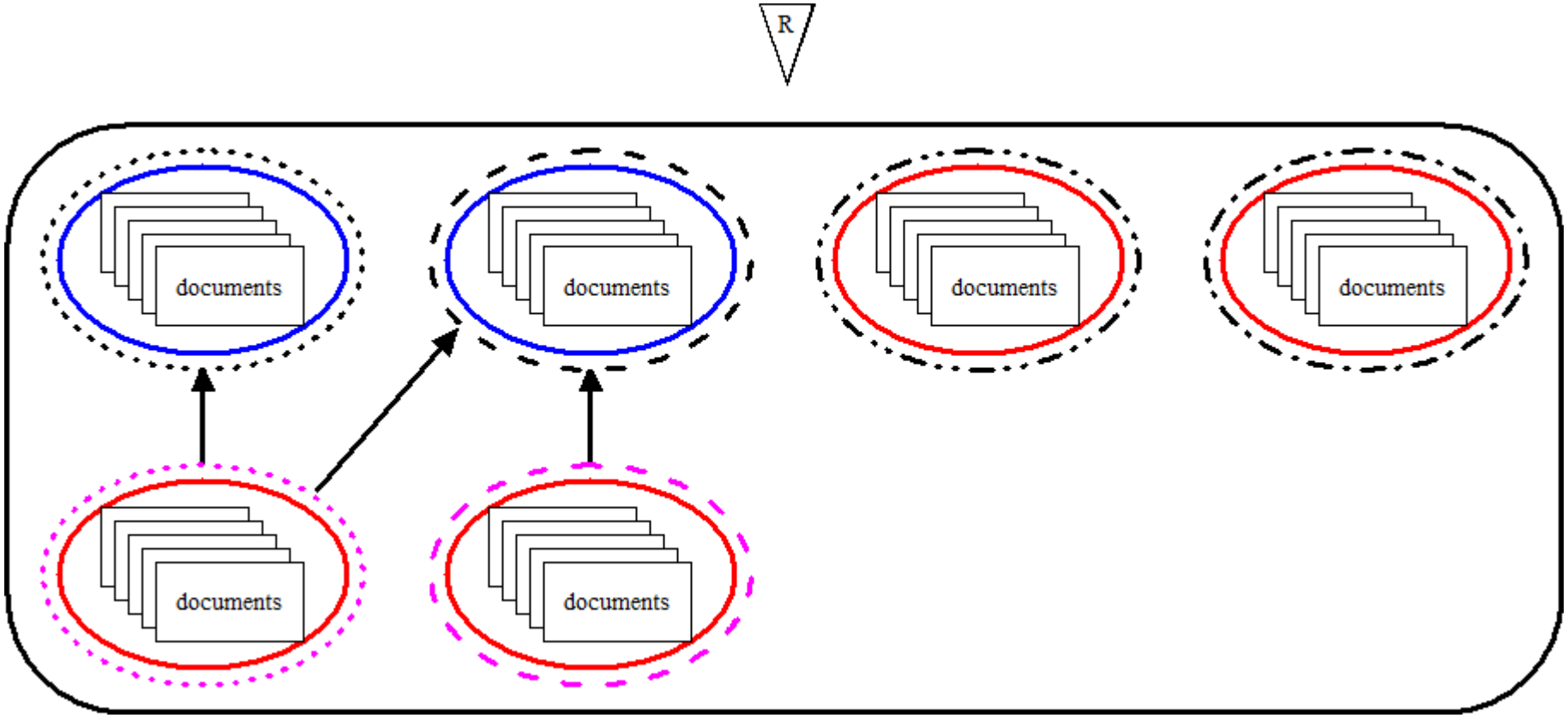
Different collections, different formats, different levels of access, different interfaces for using research data



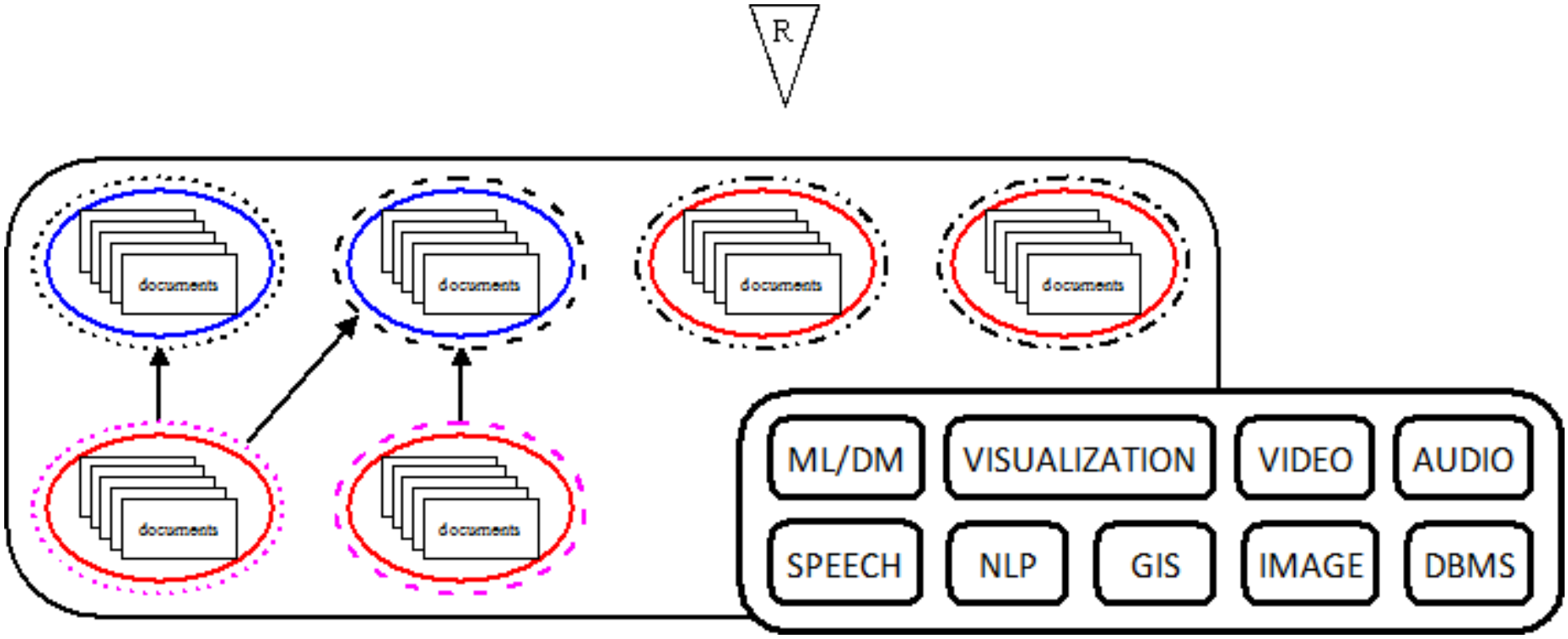
Libraries help us integrate some of these interfaces



Ideally, a cyberinfrastructure helps us access different data sources in compatible ways



And also provide tools for using these data



So,how?

Resource sharing + communication

Resource sharing:

1. Sharing through APIs

2. The sharing of documents

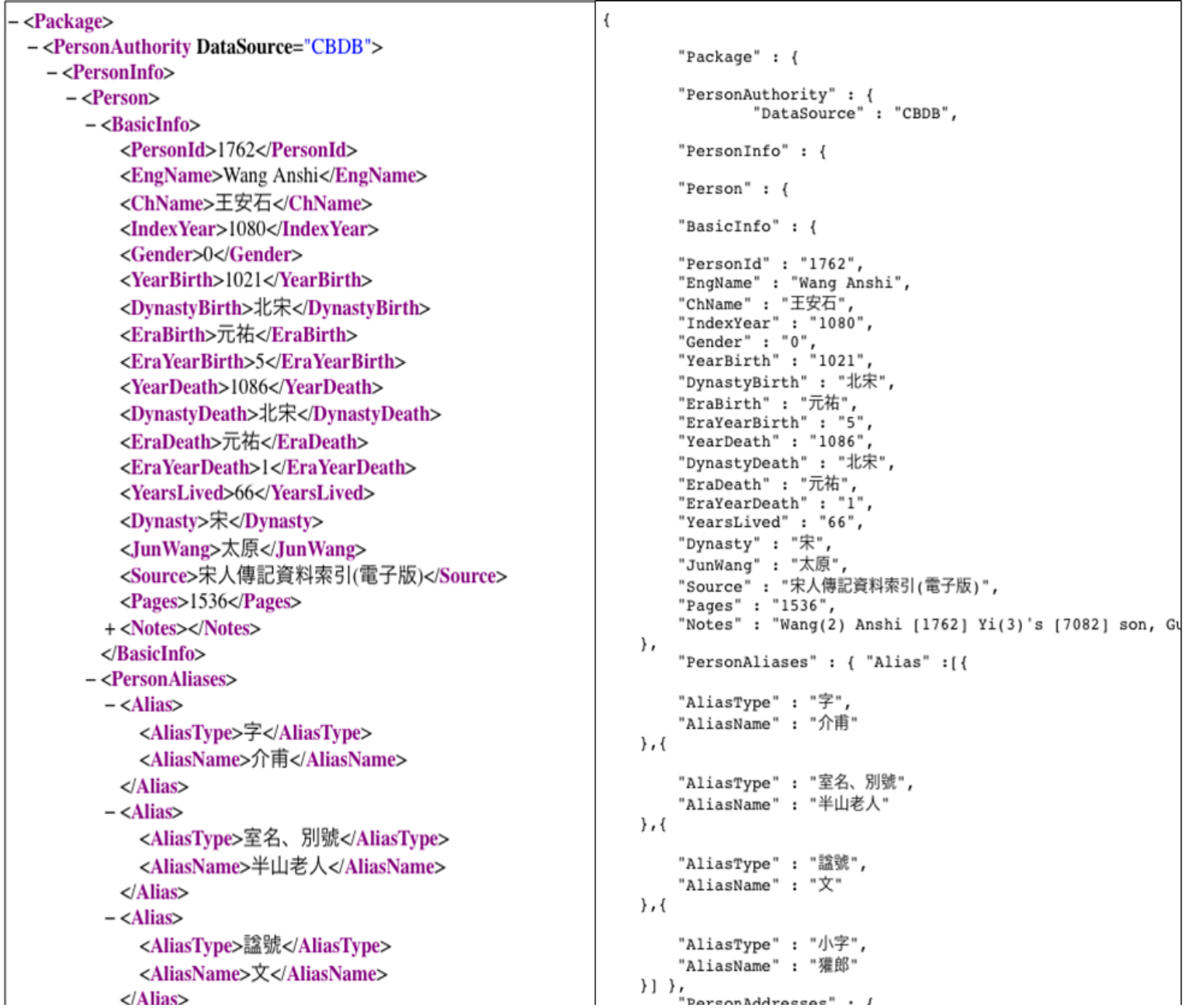
APIs: Application Programming Interface

APIs allow users access resources through a relatively stable program interface.

When the providers modify, enhance, or expand the resources, the APIs generally remain the same, thereby relieving users of unnecessary technical involvements.

Users specify the formats of their requests via an input interface, and obtain the requested resources via an output interface.

The syntax and semantics of the interfaces must be defined precisely.



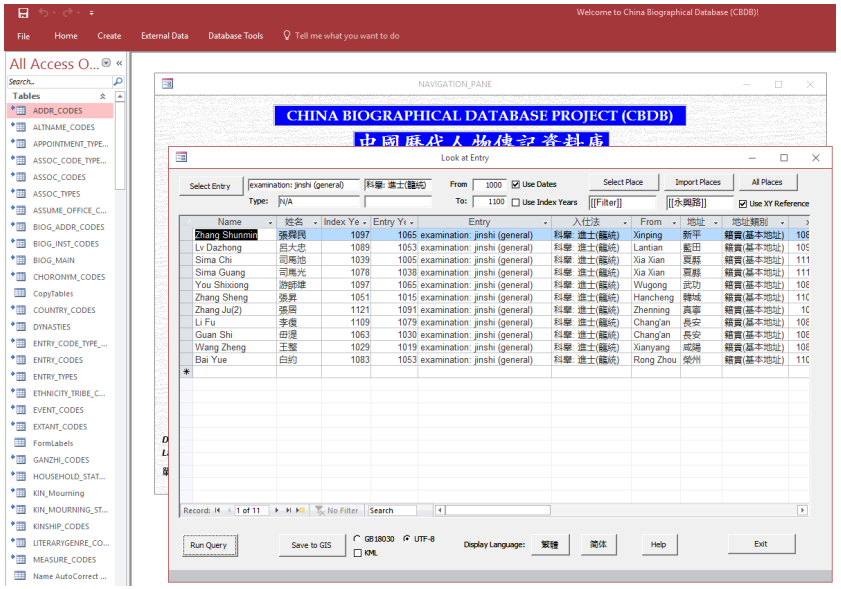
```
- <Package>
  - <PersonAuthority DataSource="CBDB">
    - <PersonInfo>
      - <Person>
        - <BasicInfo>
            <PersonId>1762</PersonId>
            <EngName>Wang Anshi</EngName>
            <ChName>王安石</ChName>
            <IndexYear>1080</IndexYear>
            <Gender>0</Gender>
            <YearBirth>1021</YearBirth>
            <DynastyBirth>北宋</DynastyBirth>
            <EraBirth>元祐</EraBirth>
            <EraYearBirth>5</EraYearBirth>
            <YearDeath>1086</YearDeath>
            <DynastyDeath>北宋</DynastyDeath>
            <EraDeath>元祐</EraDeath>
            <EraYearDeath>1</EraYearDeath>
            <YearsLived>66</YearsLived>
            <Dynasty>宋</Dynasty>
            <JunWang>太原</JunWang>
            <Source>宋人傳記資料索引(電子版)</Source>
            <Pages>1536</Pages>
          + <Notes></Notes>
          </BasicInfo>
        - <PersonAliases>
          - <Alias>
              <AliasType>字</AliasType>
              <AliasName>介甫</AliasName>
            </Alias>
          - <Alias>
              <AliasType>室名、別號</AliasType>
              <AliasName>半山老人</AliasName>
            </Alias>
          - <Alias>
              <AliasType>謚號</AliasType>
              <AliasName>文</AliasName>
            </Alias>
```

```
{
        "Package" : {

        "PersonAuthority" : {
                "DataSource" : "CBDB",

        "PersonInfo" : {

        "Person" : {

        "BasicInfo" : {

        "PersonId" : "1762",
        "EngName" : "Wang Anshi",
        "ChName" : "王安石",
        "IndexYear" : "1080",
        "Gender" : "0",
        "YearBirth" : "1021",
        "DynastyBirth" : "北宋",
        "EraBirth" : "元祐",
        "EraYearBirth" : "5",
        "YearDeath" : "1086",
        "DynastyDeath" : "北宋",
        "EraDeath" : "元祐",
        "EraYearDeath" : "1",
        "YearsLived" : "66",
        "Dynasty" : "宋",
        "JunWang" : "太原",
        "Source" : "宋人傳記資料索引(電子版)",
        "Pages" : "1536",
        "Notes" : "Wang(2) Anshi [1762] Yi(3)'s [7082] son, Gu
    },
        "PersonAliases" : { "Alias" :[{

        "AliasType" : "字",
        "AliasName" : "介甫"
    },{

        "AliasType" : "室名、別號",
        "AliasName" : "半山老人"
    },{

        "AliasType" : "謚號",
        "AliasName" : "文"
    },{

        "AliasType" : "小字",
        "AliasName" : "獾郎"
    }] },
```

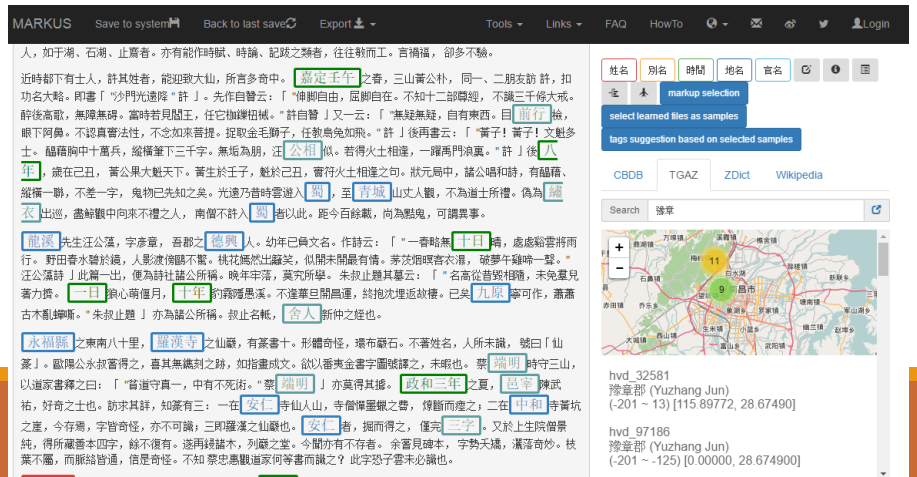
E.g. CBDB’s API

CTEXT

CBDB



MARKUS



China Historical GIS (TGAZ) API
中國歷史地理信息系統API

TGAZ API

intro

usage

examples

credits

cga

:: home

Chinese historical records the valid years of the database are -222 to 1911.

Canonical Placename

- http://maps.cga.harvard.edu/tgaz/placename/hvd_32180

Canonical Placename Formats

- **json** http://maps.cga.harvard.edu/tgaz/placename/json/hvd_80547
- **rdf** http://maps.cga.harvard.edu/tgaz/placename/rdf/hvd_135744
- **html** http://maps.cga.harvard.edu/tgaz/placename/html/hvd_9732
- **xml** http://maps.cga.harvard.edu/tgaz/placename/xml/hvd_96066

Faceted Search

parameters allowed:

- **n**: name (the spelling of the placename)
- **yr**: year of existence (the year during which the placename existed)
- **ftyp**: feature type (the the feature type, or class of placename)
- **src**: source (the data source, such as CHGIS, RAS)
- **p**: part of (the immediate parent jurisdiction where the place was located)
- **fmt**: format (the output format returned: xml, json, html)

Placename Examples

- Pinyin: http://maps.cga.harvard.edu/tgaz/placename?n=Tianbian
- Chinese: http://maps.cga.harvard.edu/tgaz/placename?n=晋阳
- Tibetan: http://maps.cga.harvard.edu/tgaz/placename?n=རྒྱ་གཞུང་
- Russian: http://maps.cga.harvard.edu/tgaz/placename?n=Вятское Наместничество

Sharing (or even crowdsourcing) place name data?

CHGIS

(Harvard CGA+ Fudan 史地所)

Base maps, times series data for place names、

http://www.fas.harvard.edu/~chgis/

APIs and Data Crawling

- To enable big data methods
  - e.g. distant reading
- Crawling/mining data $\neq$ stealing data
- More complex needs in terms of data:
  - not just full-text searches
- Agreements between projects to establish data crawling rights

Moretti：“遠讀”

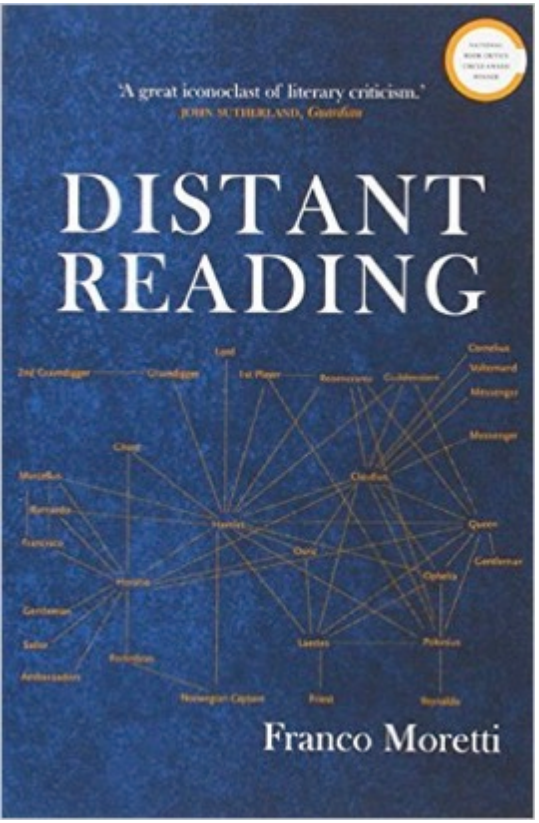
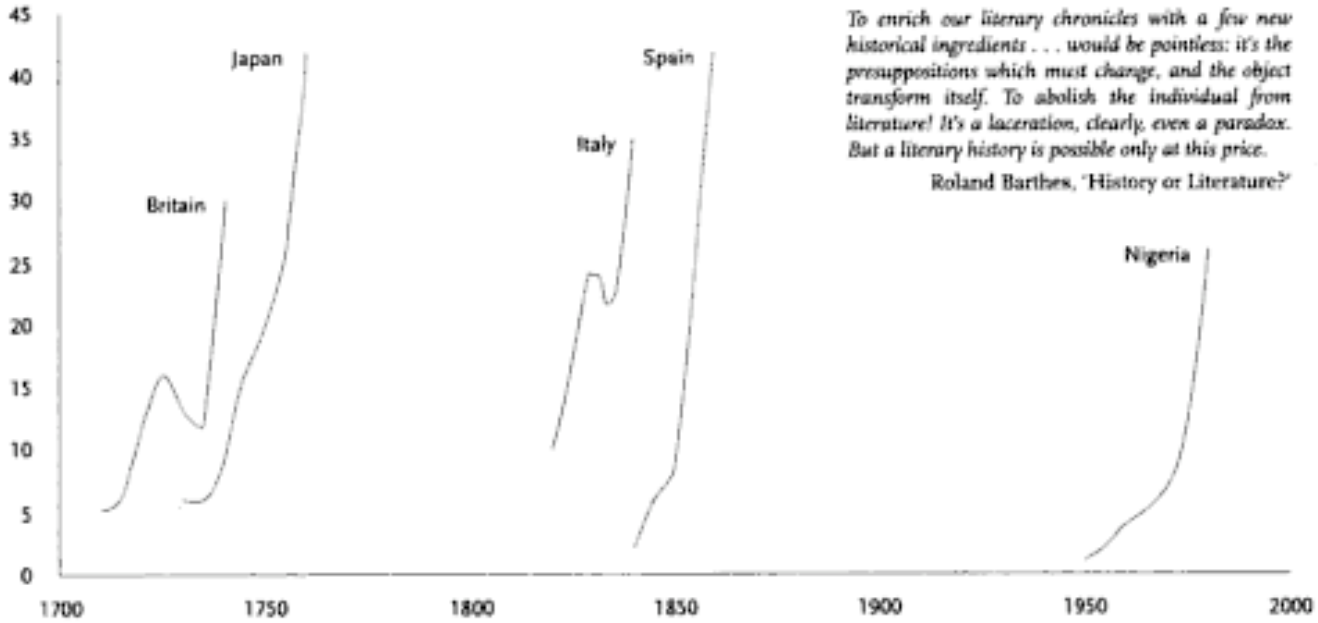


FIGURE 1: *The rise of the novel, 18th to 20th century*



To enrich our literary chronicles with a few new historical ingredients . . . would be pointless: it's the presuppositions which must change, and the object transform itself. To abolish the individual from literature! It's a laceration, clearly, even a paradox. But a literary history is possible only at this price.

Roland Barthes, 'History or Literature?'

New novels per year, by 5-year average. Sources: For Britain: W. H. McBurney, *A Check List of English Prose Fiction, 1700–39*, Cambridge, MA 1960, and J. C. Beasley, *The Novels of the 1740s*, Athens, GA 1982; both partly revised by James Raven, *British Fiction 1750–70: A Chronological Check-List of Prose Fiction Printed in Britain and Ireland*, London 1987. For Japan: Jonathan Zwicker, 'Il lungo Ottocento del romanzo giapponese', in *Il romanzo*, vol. III, *Storia e geografia*, Torino 2002. For Italy: Giovanni Ragone, 'Italia 1815–70', in *Il romanzo*, vol. III. For Spain: Elisa Martí-Lopez and Mario Santana, 'Spagna 1843–1900', *Il romanzo*, vol. III. For Nigeria: Wendy Griswold, 'Nigeria 1950–2000', *Il romanzo*, vol. III.

The right to read is the right to mine: Text and data mining copyright exceptions introduced in the UK.

New copyright exceptions to text and data mining for non-commercial research have recently come into effect and this is welcome news for UK researchers and research, argues **Ross Mounce**. *Here he provides a brief overview of the past issues discouraging text and data mining and the what the future holds now that these exceptions have been introduced. But despite legal barriers being removed, many technical barriers still remain. Furthermore it remains to be decided what formally constitutes 'non-commercial' research.*

After eight long years including not one but two expert-led reviews of intellectual property; new copyright exceptions, some of which in particular will enable and empower UK academic research

DIGGING INTO DATA CHALLENGE

Mining biographies in 2,000 local gazetteers
地方志

Automating Data Extraction from Chinese Texts

PROJECT
about
team
links

ANALYSIS
tools
data

EVENTS
workshops
release

Structured datasets extracted from 41 early gazetteers held by the Institute of History and Philology at Academia Sinica and 2,000 later gazetteers will be made available here.

Gazetteers

Local gazetteers are massive repositories of data pertaining to mountains and rivers, roads and bridges, administrative divisions, taxes and population, agricultural production, disasters and epidemics, notable local people, government officials, public and private schools, tombs and monuments, religious sites, selected writings of local authors, and a host of other topics. They were first introduced at the prefectural level (about 350 prefectures), were slowly adopted by counties (about 1300), and eventually were even written for some towns and religious sites. Authors of gazetteers focused primarily on the amassing of data rather than interpretation of events, and the categories of information they collected remained fairly consistent since the twelfth century. Individual researchers have used gazetteers to gather information on particular topics, such as natural disasters, examination candidates, and religious sites. However, systematic, national-level studies based on gazetteer data are very rare because the process of manually collecting data is extremely laborious, cannot be easily altered once begun, and is difficult to replicate in related studies.

Recent Tweets

@Tom Mullaney
@DID_ACTE
Now Live #DHAsia @cesta_stanford A Macroanalysis of Modern Japanese and Chinese Texts @RichardJeanSo @hoytlong on https://t.co/Sx8WePJYXV
3 weeks, 1 day ago

@DID_ACTE
Chinese text analysis and reading platform MARKUS featured on H-ASIA's Digital Asia resources @HNet_Humanities https://t.co/IN95Ts16dl #DH
3 weeks, 3 days ago

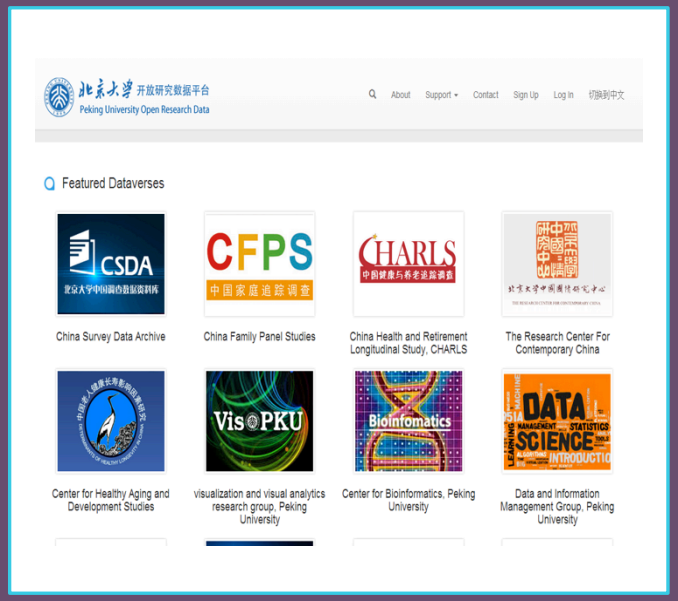
@DID_ACTE
Also featuring Chinese humanities & social sciences, and #MARKUS ! https://t.co/1CifZPiqTR


蒙古
朝鲜
汉城
서울
韩国
日本海
黄海
札幌
日本
名古屋
東京
大阪
中国
不丹
孟加拉国
缅甸
老挝
台北
台湾
香港
泰国
地图数据 ©2016 GS(2011)6020 Google, ORION-ME, SK telecom, ZENRIN
地图数据 ©2016 GS(2011)
中国方志库

Book1	Book2	overlap.count	count.adj.1	count.adj.2	percentage overlap	percentage overlap		Y: overlap N:separate
平湖縣志	平湖縣志	266	307	2688	86.64495114	9.895833333	86.64495114	Y
元城縣志	大名縣志	422	496	1766	85.08064516	23.89580974	85.08064516	Y
宣城縣志	宣城縣志	462	583	550	79.24528302	84	84	Y
滄州志	重修天津府志	506	605	7037	83.63636364	7.190564161	83.63636364	Y
烏程縣志	烏程縣志	437	688	546	63.51744186	80.03663004	80.03663004	Y
固始縣志	固始縣志	203	254	449	79.92125984	45.21158129	79.92125984	Y
雷州府志	徐聞縣志	297	2019	432	14.7102526	68.75	68.75	Y
重修天津府志	南皮縣志	357	7037	534	5.073184596	66.85393258	66.85393258	Y
溫州府志	平陽縣志	485	3578	811	13.55505869	59.8027127	59.8027127	Y
義烏縣志	金華府志	450	866	1771	51.9630485	25.40937324	51.9630485	Y
海豐縣志	陸豐縣志	103	632	302	16.29746835	34.10596026	34.10596026	Y
長泰縣志	八閩通志	133	483	14762	27.53623188	0.900961929	27.53623188	Y
八閩通志	漳平縣志	24	14762	90	0.162579596	26.66666667	26.66666667	Y
八閩通志	內黃縣志	28	14762	132	0.189676196	21.21212121	21.21212121	N

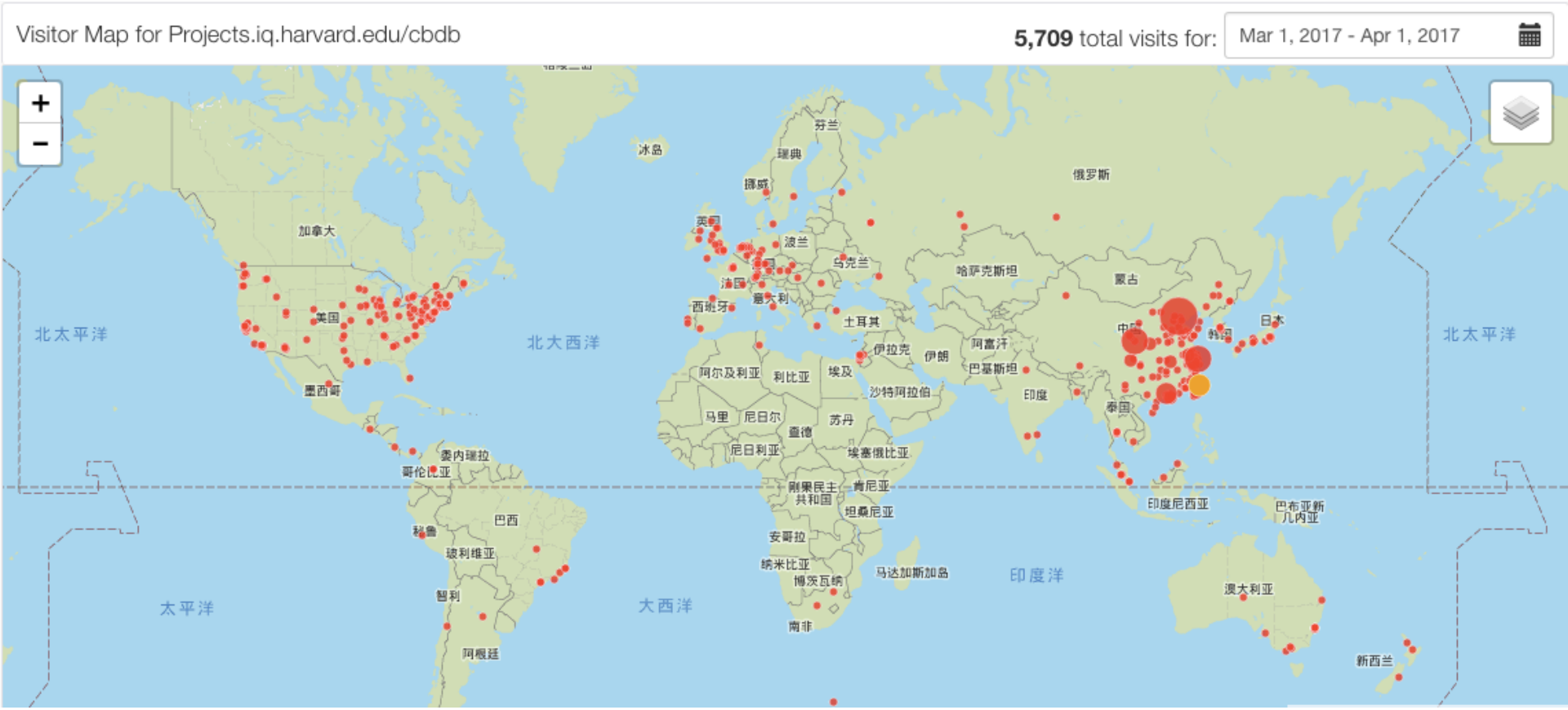
Sharing of Documents

- Sustainability of data:
  - standards for long term storage e.g. Dataverse, ICPSR
- Version control:
  - not fixed entities
  - encouraging citation of these resources
- Description:
  - user's guide, metadata descriptions etc.

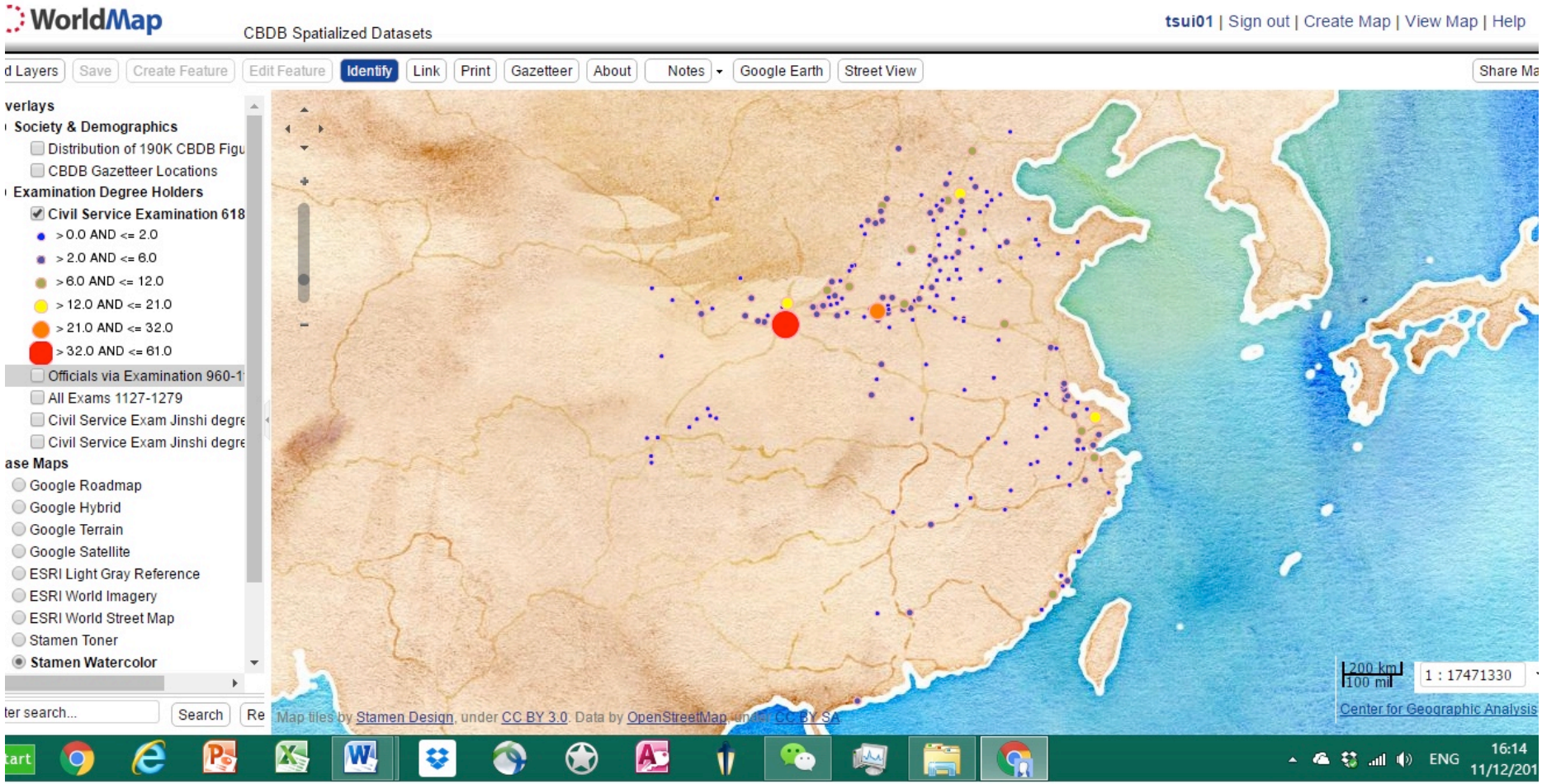


Data Repositories and their standards

Tracking user stats for optimizing database user experiences



GIS data platform: WorldMap
e.g. ChinaMap etc.

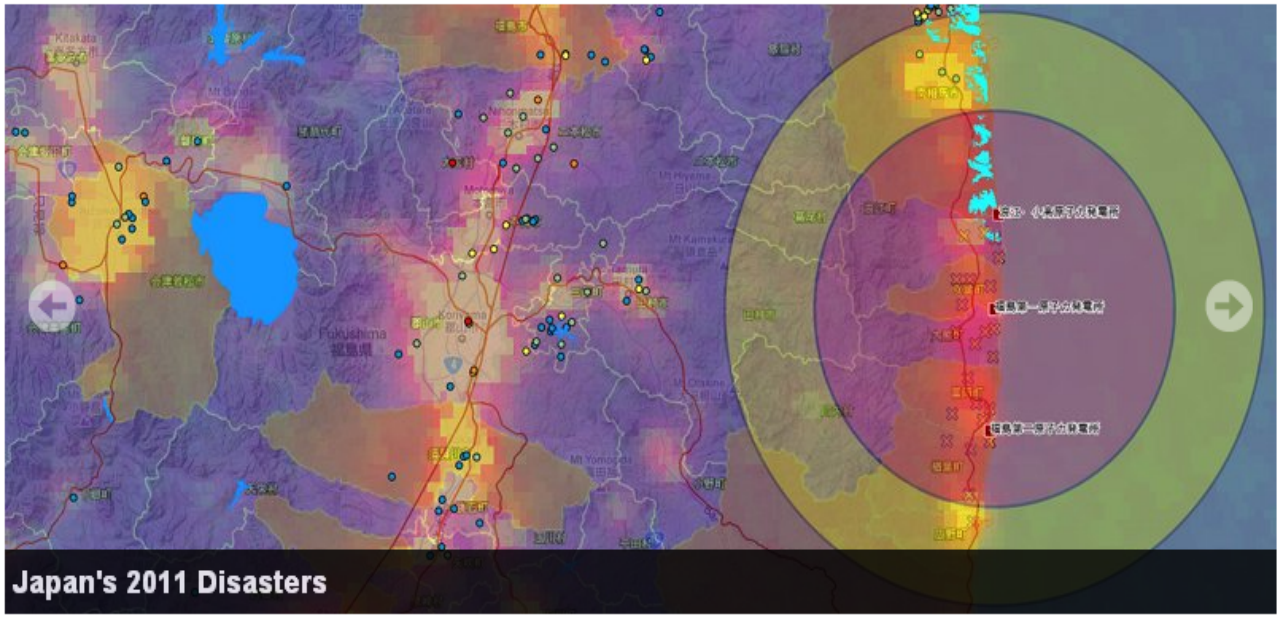


WorldMap

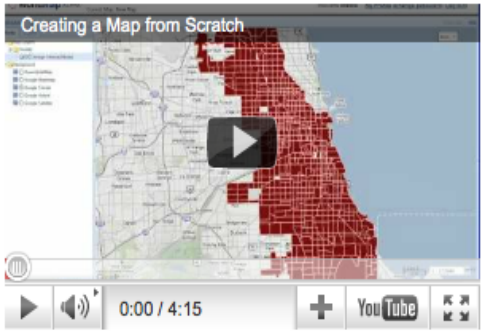
Explore, Visualize, and Publish Geographic Information

Sign in | Create Map | View Map | Help

Create a Map | View a Map | About



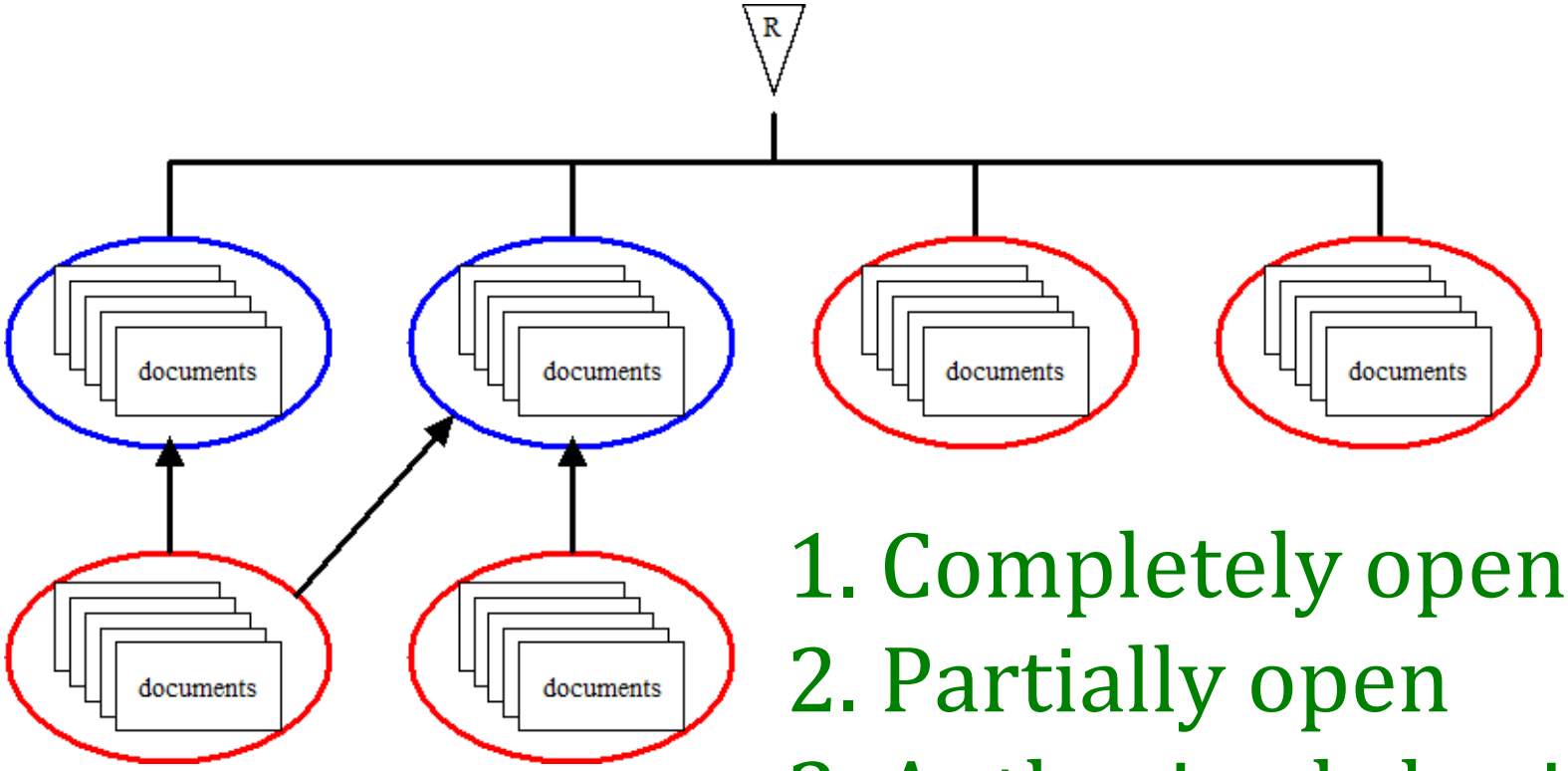
Build your own mapping portal and publish it to the world or to just a few collaborators. WorldMap is open source software.



Watch the WorldMap Quick Start video

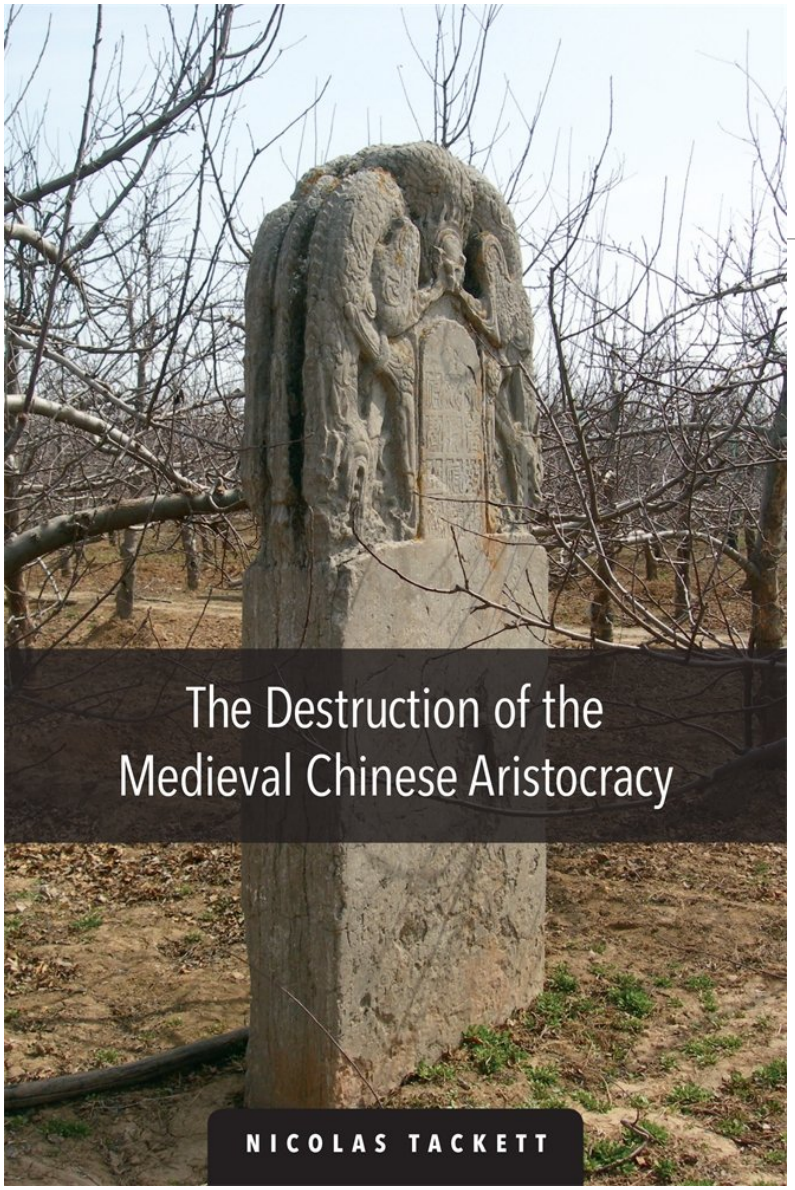
http://worldmap.harvard.edu/

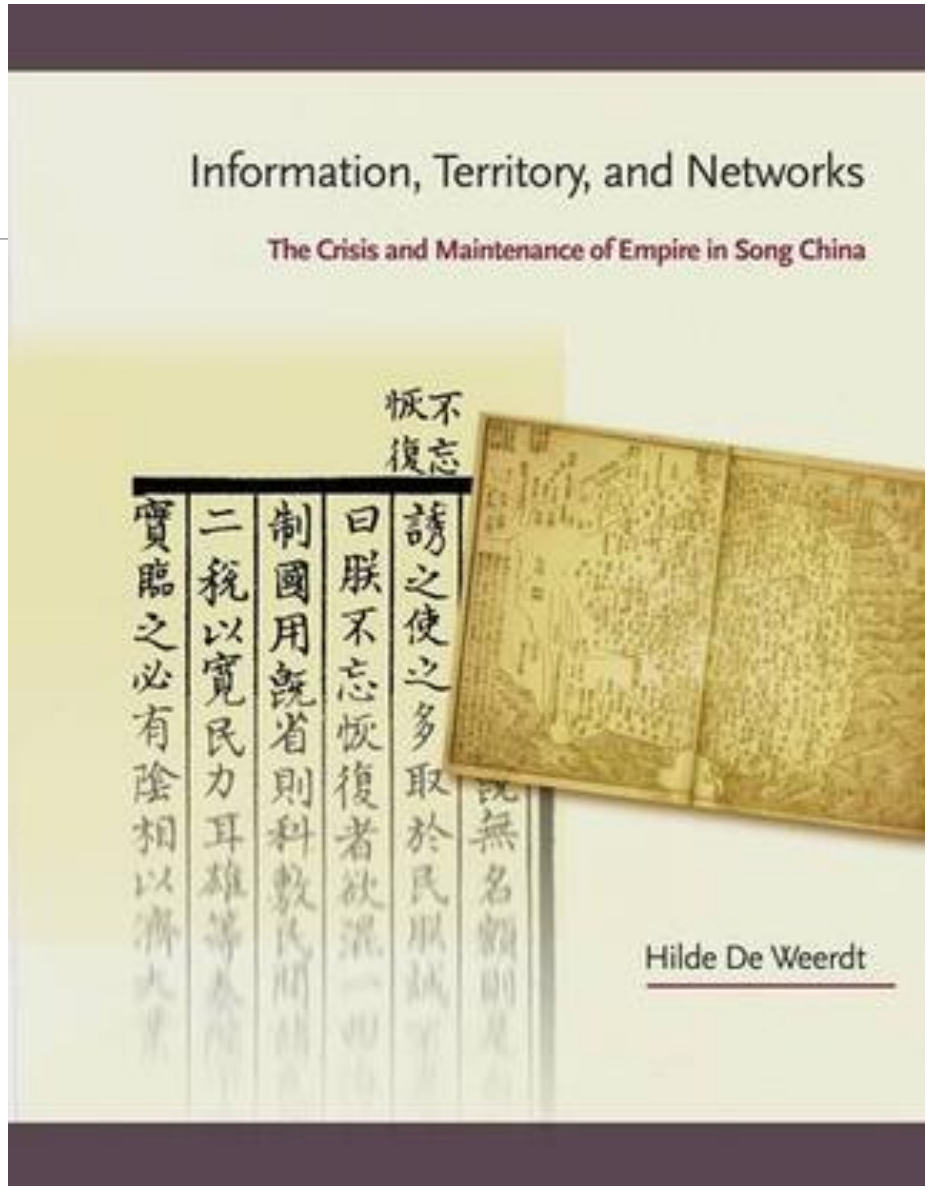
Institutional Authorization of Sharing



1. Completely open
2. Partially open
3. Authorized sharing

esp. full text databases


The Destruction of the
Medieval Chinese Aristocracy
NICOLAS TACKETT


Information, Territory, and Networks
The Crisis and Maintenance of Empire in Song China
Hilde De Weerdt

Other relevant tools of infrastructure importance

1. 跨庫書目檢索系統 Cross-Catalogue Query System of Ancient Chinese Books

2. OCR技術與中文文本資源的開放 OCR Techniques and Textual Resources e.g. CTEXT

3. 標記與可視化工具 Tagging and Visualization Tools e.g. MARKUS

4. 代碼表/權威檔 Code Tables/Authority Files e.g. data in CBDB

热点 互联网

计算机告诉你，唐朝诗人之间的关系到底什么样？

前进四先生 发表于 2017-03-15 11:28

在我还念中学的时候，每当心情不好，就靠读诗词来排遣，慢慢读得多了，就发现唐朝诗人之间存在着微妙的关系。比如杜甫非常喜欢李白，到了做梦都想见李白的地步：三夜频梦君，情亲见君意（《梦李白》）。而李白向孟浩然表过白：吾爱孟夫子，风流天下闻（《赠孟浩然》）。孟浩然的好基友则是王昌龄：数年同笔砚，兹夕间衾裯（《送王昌龄之岭南》）。

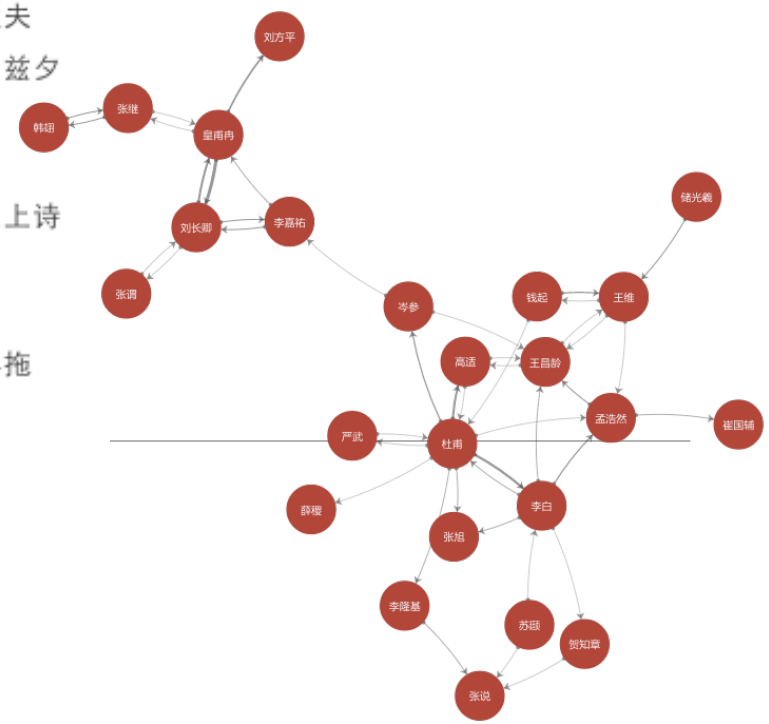
出于好奇心，我一度想理清楚他们之间的关系。但是全唐诗一共四万多首，再加上诗人之间经常称呼对方的别称，整理起来非常麻烦，慢慢的也就绝了这个念头。

前不久又想起来这段十五年前的心事。事不宜迟，拖了这么多年的愿望，不能再拖了。

这次，我将编程完成这件事。

前面已经说过，这件事主要的麻烦在于以下两点：

- **全唐诗数量太多**，一共四万多首。
- **诗人的别称太多**，比如杜甫：按字称为子美，按排行称为杜二，按官职称为杜工部。

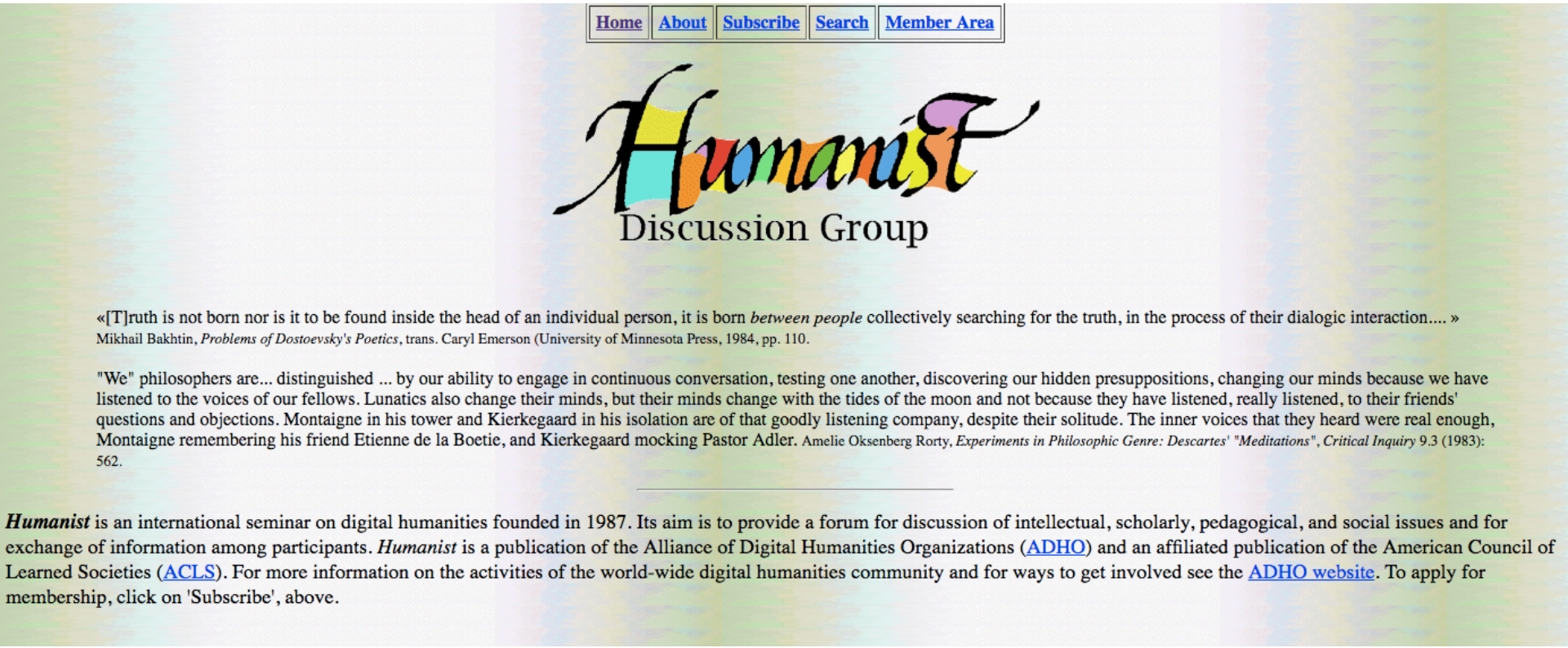


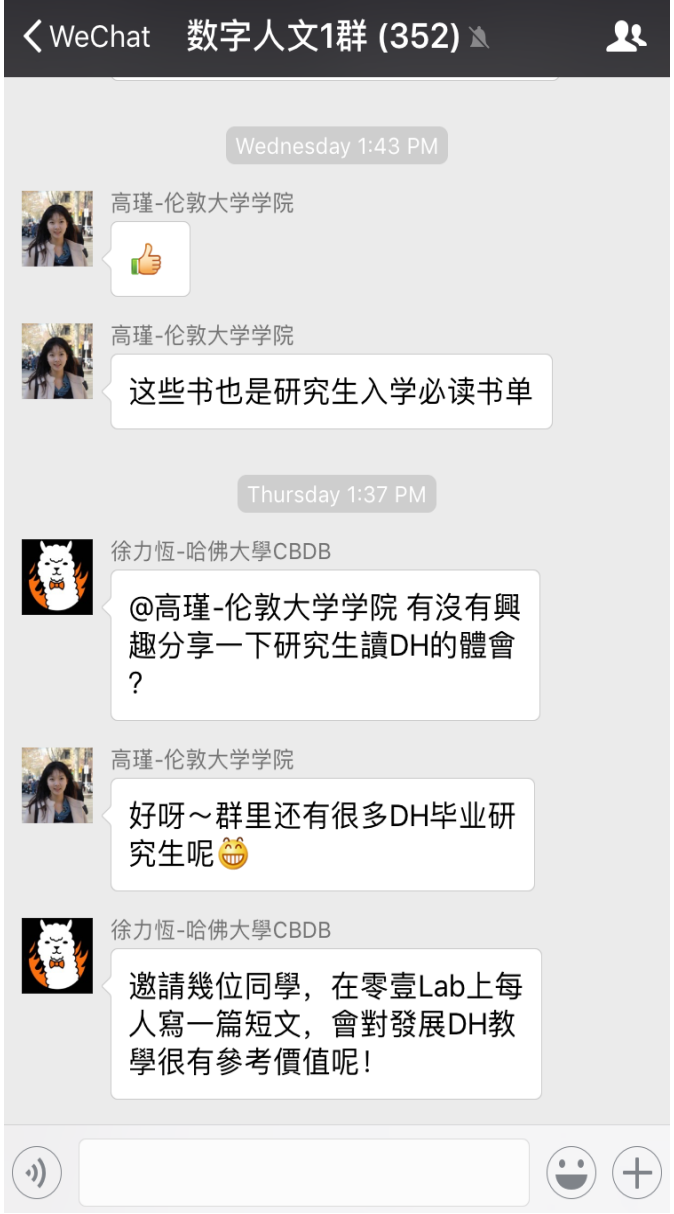
Communication between participating actors

Web-based Communication of DS/DH actors

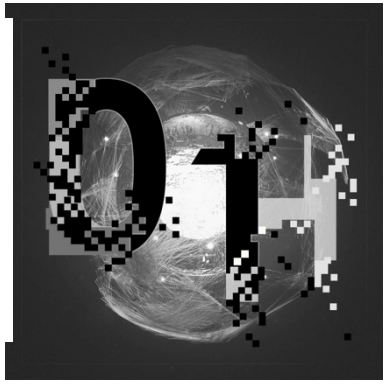
- **Cyberinfrastructure website**
  - **not centrally managed, but ideally generated and updated automatically by crawling/crowdsourcing project updates**
  - **a marketplace of ideas and resources for collaboration**
- **Mailing lists**
  - **still a good way to discuss a topic and share one's views in depth**
  - **e.g. Humanist Discussion Group**
- **Social network accounts e.g. Wechat, FB**
  - **somewhat restrictive, but users are often very active and willing to help**

http://dhhumanist.org/





Active DH scholarly communities esp. for Chinese history



零壹Lab

(ID: lingyilab)

【交换机】国内首个数字史学课程——独家推送课程资料！

原创 2016-12-12 王涛 零壹Lab

课程简介

数字工具与世界史研究

主讲：王涛

南京大学历史学院

上课地点：仙II-301

时间：2016/17学年

周二14:00-16:00

90% 02:55
× 零壹Lab ⋮

6.史学成果的网络发布：以Omeka为例

数字史学的研究成果发布超越了印刷载体的束缚，网络写作将成为一个重要的补充。为此，我们将引入元数据（Metadata）的概念，介绍以Dublin Core为代表的元数据标准。

Toolkit技能：Omeka的使用。我们将介绍网络发布平台Omeka的工作原理和相关专业术语，并通过实际的操作了解网络写作与传统写作的差异。

7.史学研究工具I：社会网络分析

我们将讨论社会网络分析（Social Network Analysis，SNA）的what，why以及how。在分析了SNA的学术发展脉络后，我们指出历史研究引入社会网络分析的方法，能够打开新的研究领域。我们将介绍SNA的分析指标，比如路径（Path）、紧密中心度（closeness centrality）、间距中心度（betweenness centrality）等，用它们来分析网络中权力的支配、权力的交换、以及社会威望等问题。

Toolkit技能：Gephi的使用。

8.史学研究工具II：自然语言处理

历史研究的重要史料是文本，在数字史学领域，这里涉及的就是自然语言处理的问题。我们将介绍自然语言的概念，并介绍基于不同算法的自然语言处理方

零壹Lab内容：http://www.iwgc.cn/list/11647

- 统计学与人文研究的结合如何可能？——清华大学统计学研究中心邓柯博士访谈
- DH年度大奖接受投票啦！请支持古文标记平台MARKUS码库思！
- 香港中文大学图书馆数码学术研究研讨会
- 你是数字人文者（DHer）么？
- 海量？智慧？整洁？混乱？——人文学科中的数据
- 《图书馆论坛》“数字人文”专栏征稿启事
- 谁是数字人文的受益者？——台湾DADH数字人文会议最新报道
- 国内首个数字史学课程——独家推送课程资料！
- 美国数字人文学者推介：霍伊特·朗(Hoyt Long，芝加哥大学)
- 独家专访Susan Schreibman：如果你习惯待在某个固定领域做研究，那最好别选数字人文

Facebook：https://www.facebook.com/DHVirtualLab

數位人文實驗室 Digital Humanities Virtual Lab

徐力恆 首頁 3

粉絲專頁 訊息

已說讚 追蹤中 分享 ···

發送訊息

數位人文實驗室 Digital Humanities Virtual Lab

@DHVirtualLab

- 首頁
- 關於
- 相片
- 按讚分析
- 活動
- 影片
- 網誌
- 貼文
- 服務內容
- 優惠

數位人文實驗室 Digital Humanities Virtual Lab

由 Ping-tzu Chu 發佈 [?] · 3月13日 23:59 ·

歡迎在花蓮地區的朋友來參加，其它地區當然也歡迎，只是無法提供旅費。

http://dhit.kktix.cc/events/corpro2017ndhu...

【庫博】Corpro文本分析工作坊

庫博是由闕河嘉老師所開發的文本分析軟體，適合分析大量文本，在很短的時間內就可以對語料庫做出一些有數量分析的觀察。歡迎大...

DHIT.KKTIX.CC

已觸及172名用戶 加強推廣貼文

讚 留言 分享

Yu Jung Cheng 、 Jing-xin Yen 和其他 2 人

留言……

顯示全部

Digital WISTM
教育

Digital chumchon lantanoi
教育

粉絲專頁的按讚

Digital Human Library

Fordham Graduate Student Di...

Digital Media & Learning Rese...

中文(台灣) · 中文(简体) · English (US) · Español · Português (Brasil) +

臺灣數位人文學會

TADH

Taiwanese Association for Digital Humanities

讀書會 @ 北京,
2017年3－7月

北京大学人文社会科学研究院
Institute of Humanities and Social Sciences, Peking University

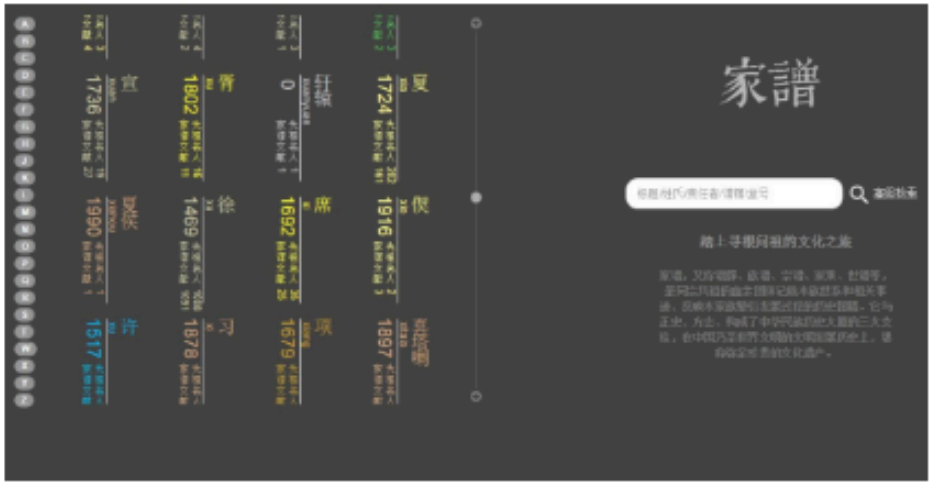
课程目标

- 掌握数字人文研究的基本概念和研究状况，并知道关注领域新进展的渠道；
- 了解重要的数字化资源和工具（尤其是针对中国文史研究的工具，如CBDB），知道利用的方法，并提高动手和解决操作问题的能力；
- 获得通过数据思考学术问题的能力，能把问题部分地转化为数字化手段能分析和呈现的课题，并摸索如何建立对自身有用的数据集；
- 了解数字人文研究的成果和新范式下学术成果的形态，并具备批判眼光，反思其研究方法和结论。

1. 数字人文的现状、基本概念和理论
2. 关系型史学数据库（上）：从用户角度看CBDB
3. 关系型史学数据库（下）：从开发者角度看CBDB
4. 电子地图和地理空间分析
5. “中国R会议”（清华大学）
6. 社会网络分析
7. “第二届北京大学数字人文论坛”（北大图书馆）
8. 文本的处理、提取和标记：MARKUS和VISUS
9. 数字人文范式下的版本目录学和书籍史
10. 总结和成果研讨

June 13, 2016, by Editor

The Digital Humanities as an Emerging Field in China



Written by Lik Hang Tsui.

The "digital humanities" (usually translated as *shuzi renwen* 数字人文 in mainland China and *shuwei renwen* 數位人文 in Taiwan) have recently received a lot of attention in Chinese academic circles, even though it took a long time for the concept to come to the attention of mainland China universities. The first digital humanities centre in China was established by Wuhan University in 2011. It remains the only mainland Chinese member of centerNet, an international network of digital humanities research centres.

But even though such research centres are so rare in China, plenty of Chinese scholars have been taking part in the field and have been conducting digital research for more than a decade. Several key universities have academic departments and centres that are developing digital projects for the humanities, such as

A cyberinfrastructure would provide the mechanisms and systems (e.g. APIs) for us to collaborate in digital scholarship for Chinese history.

It is now time for this conversation to begin for our field.

Thank you!

tsui01@fas.harvard.edu