Asia Network

An API-based cyberinfrastructure for the flexible topologies of digital humanities research in Sinology

Hou Ieong "Brent" Ho*, Sean Wang**, Pascal Belouin, and Shih-Pei Chen
Max Planck Institute for the History of Science
Berlin, Germany
swang@mpiwg-berlin.mpg.de

Abstract—**Digital humanities (DH) is a burgeoning field of research in Sinology and Asian studies more broadly, and its diversity and maturity necessitate a cyberinfrastructure fit for DH-focused Sinologists' specific needs. "Asia Network" is our solution. It is a pioneering approach for resource dissemination and emerging data analytics (such as text mining and other fair-use, consumptive research techniques) in the humanities. It is a language-agnostic software that facilitates the secure linkage between third-party research tools to different third-party textual collections (both licensed and open-access ones) via application programming interfaces (APIs). It revolutionizes how scholars can work with textual sources by promoting a flexible, networked approach to e-infrastructure development. Crucially, Asia Network is a loosely-coupled software with flexible topologies; it can enable both federated or centralized linkages, and it can even "disappear" as long as its API standards remain in place to facilitate communications among databases and tools in the back-end. Thus, unlike large-scale infrastructural projects, Asia Network actively lowers the profile of centralized infrastructure and instead promotes existing tools and resources by enabling their interoperability. As a result, it allows scholars to fully leverage the potential of material digitization and digital research tools without re-creating silos of resources in the digital realm.**

Keywords—digital humanities; Sinology; cyberinfrastructure

I. Introduction

Digital humanities (DH) is a burgeoning field of research in Sinology and Asian studies more broadly. DH research techniques, including various databases from digitization efforts and growing numbers of digital research tools, have had an impact on Sinologist research communities globally. Stanford University, for example, has held an annual "Digital Humanities Asia" conference since 2016 [1], and Taiwan is holding the ninth International Conference of Digital Archives and Digital Humanities in December 2018 [2]. There are also collaborations across multiple regions, as evidenced by the enthusiastic participants at the International Conference on Cyberinfrastructure for Historical China Studies this March at Harvard Center Shanghai [3]. Such events demonstrate the diversity and maturity of DH in Sinology globally.

Outside of Sinology, DH has been grappling with issues such as long-term sustainability and interoperability. In response, many have proposed that DH needs basic infrastructures behind research projects to ensure its long-term success. In Europe, for instance, CLARIN [4] and DARIAH [5] are two such large-scale research infrastructures for humanities. While they have done a tremendous job in centralizing available digital resources, much of their infrastructures remain at the administrative level, and their generic coverage across the entire humanities meant that their utility for a specific discipline like Sinology is limited. How can we, as DH scholars and Sinologists, design a cyberinfrastructure fit for our specific needs, taking past experiences with these large-scale infrastructural projects into consideration?

"Asia Network" is our answer to this question [6]. It is a pioneering approach for resource dissemination and emerging data analytics (such as text mining and other fair-use, consumptive research techniques) in the humanities. It is a language-agnostic software that facilitates the secure linkage between third-party research tools to different third-party textual collections (both licensed and open-access ones) via application programming interfaces (APIs). It revolutionizes how scholars can work with textual sources because, under the current condition, it is impossible for scholars to use digital research tools to analyze licensed textual collections without downloading or scraping the full texts, which violates licensing terms. The Asia Network software can securely pass through these licensed texts to digital research tools, thus allowing scholars to work in a legal manner and ensuring commercial publishers the safety of their collections. Such flexible, networked approach to e-infrastructure development avoids re-creating silos of resources in the digital realm and allow scholars to fully leverage the potential of material digitization and digital research tools. Crucially, Asia Network is a loosely-coupled software with flexible topologies; it can enable both federated or centralized linkages, and it can even "disappear" as long as its API standards remain in place to facilitate communications among databases and tools in the back-end. Thus, unlike large-scale infrastructural projects, Asia Network actively lowers the profile of centralized infrastructure and instead promotes existing tools and resources by enabling their interoperability.

Since our inception in May 2017, Asia Network has progressed to the beta development stage and we plan to release it publicly by the end of this year. At the time of writing, Asia Network is linked via APIs to the following resources: CBETA [7], the Taiwan History Digital Library [8], the Kanseki

* Presenting author

** Corresponding author

Repository [9], the Chinese Text Project [10], a small set of Staatsbibliothek zu Berlin's classical Chinese collections, and Perseus (Greek and Latin materials only) [11]. The only linked research tool is MARKUS [12], though DocuSky [13] and other tools are on our immediate development horizon. Here in this paper, we outline (1) the current landscape of DH in Sinology and what we see as the main challenges it presents to researchers; (2) a basic summary of Asia Network's functions and features design; and (3) a call for collaborators to develop common API and metadata standards for DH in Sinology.

II. Current Landscape of DH in Sinology

Digitization of historical materials has dramatically transformed how Sinologists gather research sources and approach research questions. As more and more archives, libraries, and other research institutions embrace digital technologies, DH-focused projects and initiatives in Sinology would only continue to proliferate. This growth, however, cannot be assumed as a foregone conclusion, as the current landscape of DH in Sinology is incredibly fractured and already presents many roadblocks to seamless access and sustainability. In this section, we briefly survey this fractured landscape and show how our Asia Network infrastructure bridges the fault lines within it.

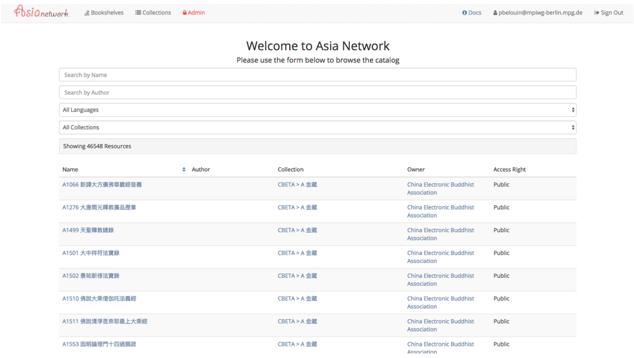
The current landscape is fractured both geographically and thematically. It is not a hyperbole to say that Sinology is a global discipline today, as China studies departments and research centers exist in many countries. However, a core-periphery relationship among the roles and foci of these nodes of global Sinology research community persists. In Mainland China, where much of the primary sources still reside in various archives and institutions, the commercialization of resource digitization reigns supreme. Despite the fact that many of these sources originate in the public domain (or have long passed their copyright protection terms), commercial publishers build proprietary databases from the digitization and charges high royalties for access. Such commercial models are being copied by university libraries and presses as well. While open-access movements have been gaining steam in recent years and the Chinese DH community has grown dramatically, as seen in the third DH symposium at Peking University [14], this sector has continued to rely on this commercial model and shows little signs of movement. In particular, subscription access often does not include provisions for text mining, full-text access, and other standard DH techniques today, as the pricing model still prioritizes read-only access. It should be noted that we are not advocating for eradicating commercial database vendors from this landscape; rather, it is to point out that their existing business model (and the accompanying lack of technical improvements) make it difficult to leverage the full potential of their digitized materials, even for researchers who have subscription access to their materials.

In Taiwan, DH's first strong foothold in the Chinese-speaking world, intersecting scholarly expertise in humanities and computer science resulted in a strong environment for the development of research databases and research tools development. Many of these databases on driven by thematic interests, such as collections of Buddhist texts or Taiwanese historical documents, just to name a few [7, 8]. Also of note is the DocuSky platform developed by the National Taiwan University, which allows individual users to organize and work on their own set of materials in one place [13]. While there are certainly monetizing tendencies in Taiwan as well, it is notable that many of these databases and tools are built on open-access principles and, if there is a fee, it is usually only for high-usage clients and/or to cover basic maintenance costs. These practices encourage development and sharing of new tools and resources, and Taiwan hosted the first Asia-focused DH international conference in 2009 [15]. Since then, the International Conference of Digital Archives and Digital Humanities (DADH) has become an annual event, and the Taiwanese Association for Digital Humanities became a constituent organization of the global Alliance of Digital Humanities Organizations in 2018 [16].

Elsewhere in the world, especially in North America and Europe, thematic DH research projects dominate the landscape. Close alliances between scholars and librarians, aided by international funding bodies like the Andrew W. Mellon, Chiang Ching-kuo, and Luce Foundations create diverse projects that largely consume primary sources for producing research results and presentations. Nonetheless, long-term preservation and sustainability remain serious issues, as is the linking of individual silos of project repositories and making them interoperable. In the context of Sinology-focused DH projects, Harvard's China Biographical Database (CBDB) exemplifies both the success and the pitfalls, as its continued growth over almost two decades of open-access development was threatened by funding uncertainty and, just this year, sold its distribution rights in Mainland China to a commercial publisher [17]. In Europe, centralized governmental funding agencies like the European Research Council provides more stability, but similar stories exist as well. Nonetheless, many top research projects and tools (such as MARKUS [12], Ten Thousand Rooms [18], and Ming Qing Women's Writings [19]) come from Sinologists based in North America and Europe and continue to base at libraries and research institutions there.

So, how should an individual researcher in Sinology approach this fractured landscape? If one is interested in primary sources, there are many individual databases that must be searched one by one. While there are open-access, full-text databases like the Kanseki Repository [9], the majority remain proprietary and read-only, making the usage of digital research tools on full texts incredibly difficult, as well as synthesis of sources from mixed copyright origins. In the rare cases where digital analyses and manipulation are possible via tools like MARKUS [12] or LoGaRT [20], sharing of results is challenging tool. If one is interested in integrating research products from many DH projects, many employ static silos of project websites or repositories without appropriate technical linkages that enable interoperability. It is heartening that DH in Sinology has progressed to such a point where a critical mass of research and researchers meant that we must consider cyberinfrastructure, and our "Asia Network" is a proposed solution that bridges these complex fault lines in this fractured global landscape.

III. “ASIA NETWORK”: A BASIC CYBERINFRASTRUCTURE FOR DH RESEARCH IN SINOLOGY



Asianetwork   Bookshelves   Collections   Admin   Docs   pbelouin@mpiwg-berlin.mpg.de   Sign Out

Welcome to Asia Network

Please use the form below to browse the catalog

Search by Name

Search by Author

All Languages

All Collections

Showing 46548 Resources

Name	Author	Collection	Owner	Access Right
A1066 新譯大方廣佛華嚴經音義		CBETA > A 金藏	China Electronic Buddhist Association	Public
A1276 大唐開元釋教廣品歷章		CBETA > A 金藏	China Electronic Buddhist Association	Public
A1499 天聖釋教總錄		CBETA > A 金藏	China Electronic Buddhist Association	Public
A1501 大中祥符法寶錄		CBETA > A 金藏	China Electronic Buddhist Association	Public
A1502 景祐新修法寶錄		CBETA > A 金藏	China Electronic Buddhist Association	Public
A1510 佛說大乘僧伽吒法義經		CBETA > A 金藏	China Electronic Buddhist Association	Public
A1511 佛說清淨毘奈耶最上大乘經		CBETA > A 金藏	China Electronic Buddhist Association	Public
A1553 因明論理門十四過類疏		CBETA > A 金藏	China Electronic Buddhist Association	Public

Fig. 1. Asia Network’s web interface landing page.

Asia Network is an infrastructure for DH projects (e.g., databases, tools, research platforms) to link with one another. Based on a collaborative process involving experienced stakeholders like system developers, DH researchers in Sinology, and librarians, our goal is to produce a general, reusable APIs that can cover common DH projects activities, such as log-in mechanism, contents discovery, tools discovery, contents and tools matching, and personal online research workspace (linked to researchers’ individual storage). While there is a front-end web user interface that we have developed in-house, it is not essential for the API-linked ecosystem to function. This loosely-coupled infrastructural design and its flexible topologies are Asia Network’s distinguishing feature from other large-scale, centralized research infrastructures.

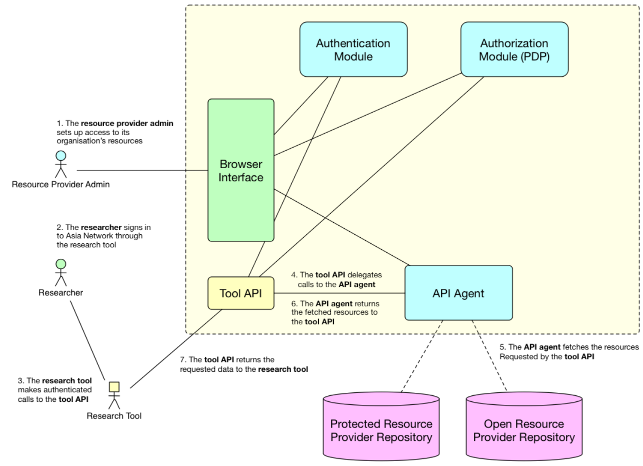
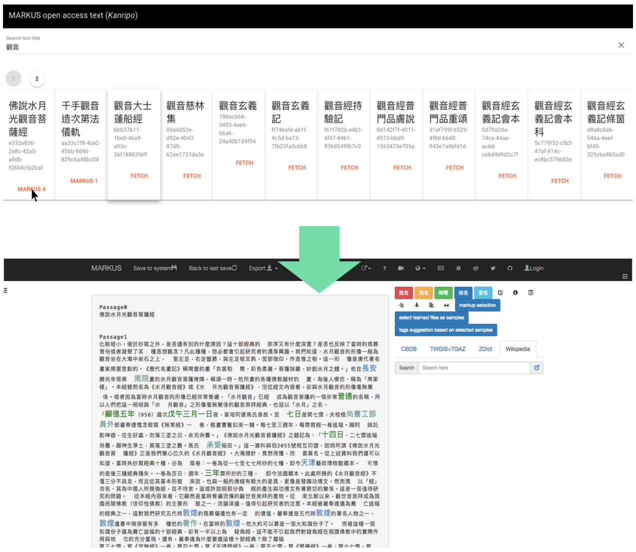

Authentication Module
Authorization Module (PDP)
1. The resource provider admin sets up access to its organisation’s resources
Resource Provider Admin
Browser Interface
2. The researcher signs in to Asia Network through the research tool
Researcher
Tool API
4. The tool API delegates calls to the API agent
6. The API agent returns the fetched resources to the tool API
API Agent
5. The API agent fetches the resources Requested by the tool API
7. The tool API returns the requested data to the research tool
3. The research tool makes authenticated calls to the tool API
Research Tool
Protected Resource Provider Repository
Open Resource Provider Repository

Fig. 2. Asia Network’s API-based architecture, based on the flexible topologies of DH in Sinology.

Despite its current name, the Asia Network software can handle resources in all languages. Many projects and infrastructures have proposed similar ideas (including many European Commission-funded e-infrastructures), creating complex new initiatives like CLARIN and DARIAH [4, 5]. Ours, by comparison, is a modular solution that works, adapting and growing with research projects. Research and structural design remain intimately connected. This demonstrates the significant returns from our early investment into DH research in Sinology.

The flexible topologies based on APIs enable diverse DH tools and contents development because they allow decentralized, role-based collaborative growth; said more simply, as long as individual stakeholders implement common API standards, everyone can just focus on their specific tasks knowing that results will be interoperable. This allows each DH project to focus on its own critical, unique contributions. At the same time, individual researchers can still speed up their research using their existing research tools to search across multiple databases without necessarily going to a centralized portal. This workflow based on APIs apply to DH projects based on computational (rather than text-based) methodologies as well. The Asia Network APIs are designed to be flexible for different topologies (*ad hoc*, centralized, federated/distributed). In all cases, Asia Network maintains critical activities (i.e., audit, security check, transaction monitoring and encryption) for interactions among licensed materials and research tools.

Asia Network includes three main user roles in its design: resource providers, tool developers, and institutionally-affiliated users. The current API definition was designed to connect resource providers and tools developers [21]. In an *ad hoc* topology, resource provider and tool developer must directly interact with each other’s APIs. However, in most general cases (centralized and federated/distributed), Asia Network functions as a hub to maintain the most up-to-date API standards and to facilitate interactions among hooked-up resources and tools. Audit and other authorization actions are done via Asia Network’s web interface as well. This infrastructure enables speedy back-end integration among existing DH resources and tools and does not reinvent the wheel at a large scale. It also enables researchers to freely manipulate and analyze resources they have access to in different research tools without violating licensing terms.



MARKUS open access text (*Kanripo*)

Search text title

觀音

佛說水月光觀音菩薩經 e332e836-2e8c-42a0-a0db-f0604cfa2baf MARKUS 4

千手觀音造次第法儀軌 aa33c1f8-4ce2-456b-9d46-829c6a48bc04 MARKUS 1

觀音大士蓮船經 6bb37b11-1be0-46a9-a93c-36f18882f6ff FETCH

觀音慈林集 00a4d52e-d92e-4043-87d5-62ee1737da3e FETCH

觀音玄義 198ac666-3403-4ae6-b6a6-24a40b184f94 FETCH

觀音玄義記 ff74eafe-a61f-4c5d-ba13-7fb23fa3c6b8 FETCH

觀音經持驗記 f61f782b-e4b1-4f47-84b1-936d549fb7c0 FETCH

觀音經普門品膚說 0d142f7f-4011-4973-bbd9-1565473e705a FETCH

觀音經普門品重頌 31ef759f-b529-4f8d-b640-943e7a4bfd1d FETCH

觀音經玄義記會本 0d79a2da-74ca-44ae-ac6d-ce6d4d9d2c7f FETCH

觀音經玄義記會本科 5c779f52-cfb2-47af-814c-ec8bc579b03e FETCH

觀音經玄義記條箇 d8a8c6ab-54da-4eef-bf45-325cba465ad0 FETCH

Passage0
佛說水月光觀音菩薩經

Passage1
比較短小，便於抄寫之外，是否還有別的什麼原因？這十部經典的 排序又有什麼深意？是否也反映了當時的喪葬習俗或者凝聚了某 種思想觀念？凡此種種，想必都會引起研究者的濃厚興趣。我們知道，水月觀音的形像一般為觀音坐在大海中岩石之上， 垂左足，右足豎膝，與左足相叉鉤，面部微仰，作思惟之相。這一形 像是唐代著名畫家周窗首創的。《歷代名畫記》稱周窗的畫「衣裳勁 簡，彩色柔麗。菩薩端嚴，妙創水月之體。」他在長安勝光寺塔東 南院畫的水月觀音菩薩掩障，稱頌一時。他所畫的各種佛教題材的 畫，為後人模仿，稱為「周家樣」。本經雖然名為《水月觀音經》或《水 月光觀音菩薩經》，但從經文內容看，卻與水月觀音的形像毫無關 係。或者因為當時水月觀音的形像已經非常普遍，「水月觀音」已經 成為觀音菩薩的一個非常普通的名稱，所以人們把這一明明與「水 月觀音」之形像毫無關係的觀音崇拜經典，也冠以「水月」之名。
「顯德五年（958）歲次戊午三月一日夜，家母阿婆馬氏身故。至 七日是開七齋，夫校檢尚書工部員外郎翟奉達憶念敬寫《無常經》一 卷，敬畫寶髻如來一鋪。每七至三週年，每齋寫經一卷追福。願阿 娘託影神遊，往生好處，勿落三塗之災，永充供養。」《佛說水月光觀音菩薩經》之題記為：「十四日，二七齋追福供養，願神生淨土，莫落三塗之難。馬氏 承受福田。」這一資料與伯2055號相互印證，說明所謂《佛說水月光觀音菩 薩經》正是我們關心已久的《水月觀音經》。大海撈針，竟然得獲，欣 喜莫名。從上述資料我們還可以知道，當時共抄寫經典十種，分為 兩卷：一卷為從一七至七七所抄的七種，即今天津藝術博物館藏本。 可惜的是後三種經典殘失。一卷為百日、週年、三年齋所抄的三種， 即今法國藏本。此處所錄的《水月觀音經》不僅三分不具足，而且從其基本形態 來說，也與一般的佛經有較大的差異，更像是發願功德文。然而竟 以「經」命名，其為中國人所撰偽經，自不待言。這或許說明部分偽 經的產生與功德文有著密切的關係。這是一個值得研究的問題。 從本經內容來看，它顯然是當時普遍流傳的觀世音崇拜的產物。從 南北朝以來，觀世音崇拜成為我國民間佛教（信仰性佛教）的主要形 態之一，流韻深遠，值得引起研究者的注意。本經被翟奉達選為薦 亡追福的經典之一，這對我們研究五代時敦煌的喪葬儀禮也有一定 的價值。翟奉達是五代時敦煌的著名人物之一，敦煌遺書中現保留有多 種他的著作。在當時的敦煌，他大約可以算是一個大知識份子了。 而被這樣一個知識份子選為薦亡追福的十部經典，卻有一半以上為 疑偽經，這不能不引起我們對疑偽經在我國佛教中的實際作用與地 位的充分重視。還有，翟奉達為什麼要選這樣十部經典？除了篇幅
第三七齋，寫《咒魅經》一卷；第四七齋，寫《天請問經》一卷；第五七齋，寫《閻羅經》一卷；第六七齋，寫

Fig. 3. A typical workflow passing textual resources to research tools via Asia Network APIs, in this case from the Kanseki Repository to MARKUS.

IV. Conclusion

While open-access remains the ideal end goal in any DH endeavors whenever possible, the reality is that many digitized resources in the humanities are still sold by publishers or private vendors. In Sinology especially, a fractured landscape as described above creates difficult conditions for DH scholars. We have had to navigate this complex licensing terrain during our everyday work, and Asia Network is now a prototype primed for transforming DH in Sinology. Besides our core collaborators at Leiden University and the Staatsbibliothek zu Berlin, we have now linked up with collaborators in the United States, Taiwan, Japan, Mainland China, Singapore, and beyond. We look forward to launching Asia Network later this year, and we encourage collaborative development and constructive feedback from all those who wish to contribute to building basic cyberinfrastructure in Sinology.

Acknowledgment

We thank Dagmar Schäfer, Matthias Kaun, and Hilde de Weerdt for their ongoing work and support for this project. We also thank all of our collaborators, especially Jieh Hsiang who invited us to be a part of this panel. Finally, we thank our student assistant Vitaly Lyapunov for his invaluable work in Asia Network's ongoing technical development.

References

[1] http://dhasia.org/

[2] http://dadh2018.dila.edu.tw/

[3] https://projects.iq.harvard.edu/cbdb/international-conference-cyberinfrastructure-historical-china-studies

[4] https://www.clarin.eu/

[5] https://www.dariah.eu/

[6] See https://asia-network.mpiwg-berlin.mpg.de/ for the web user interface for Asia Network's beta prototype.

[7] http://www.cbeta.org/

[8] http://thdl.ntu.edu.tw/index.html

[9] https://www.kanripo.org/

[10] https://ctext.org/

[11] http://www.perseus.tufts.edu/hopper/

[12] https://dh.chinese-empires.eu/markus/beta/

[13] https://docusky.digital.ntu.edu.tw/DocuSky/ds-01.home.html

[14] https://www.lib.pku.edu.cn/portal/cn/news/0000001622

[15] http://www.dadh-record.digital.ntu.edu.tw/Scope.php?LangType=en&His=D09k2

[16] https://www.adho.org/announcements/2017/adho-welcomes-new-organizations-0

[17] https://projects.iq.harvard.edu/cbdb

[18] https://tenthousandrooms.yale.edu/

[19] http://digital.library.mcgill.ca/mingqing/

[20] https://www.mpiwg-berlin.mpg.de/research/projects/logart-local-gazetteers-research-tools

[21] See all three tabs at https://asia-network.mpiwg-berlin.mpg.de/pages/doc_for_resource_providers for our current API definition.