Platform Perspective for Library Digital Scholarship Support

• Xiaohua Shi (Visiting Librarian, PhD)
Platform Perspective for Library DS Support

CONTENTS

1 Background

2 DS in HYL

3 Platform Perspec.

4 Conclusion
BACKGROUND

Harvard-Yenching Library

East Asian studies

The largest university library for East Asian research outside of Asia.
Scholars across disciplines are increasingly adopting digital tools and methods in their research and teaching.

Digital Library → Digital Scholarship

The library has decades of experience with DS/DH.

The library has some of the tools and resources you need.

The library will continually break through with innovative methods to meet users’ needs.
Scoping the Field of Digital Scholarship

Open Movement

Products

Open Textbooks
Open Research

Scholarly Communication

Open Data
Digital Assets

Processes

e-Science

Open Access

Open Science
OER

## Digital Scholarship Areas of Practice

<table>
<thead>
<tr>
<th>Area of Practice</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Altmetrics</td>
<td>• Altmetrics</td>
</tr>
<tr>
<td>APCs</td>
<td>• APCs</td>
</tr>
<tr>
<td>ASNs</td>
<td>• ASNs</td>
</tr>
<tr>
<td>IPR</td>
<td>• IPR</td>
</tr>
<tr>
<td>Journal publishing</td>
<td>• Journal publishing</td>
</tr>
<tr>
<td>Licensing</td>
<td>• Licensing</td>
</tr>
<tr>
<td>Author IDs</td>
<td>• Author IDs</td>
</tr>
<tr>
<td>Bibliometrics</td>
<td>• Bibliometrics</td>
</tr>
<tr>
<td>Content hosting</td>
<td>• Content hosting</td>
</tr>
<tr>
<td>Maker spaces</td>
<td>• Maker spaces</td>
</tr>
<tr>
<td>Mediated deposit</td>
<td>• Mediated deposit</td>
</tr>
<tr>
<td>Metadata</td>
<td>• Metadata</td>
</tr>
<tr>
<td>Copyright</td>
<td>• Copyright</td>
</tr>
<tr>
<td>CRIS/RIMS</td>
<td>• CRIS/RIMS</td>
</tr>
<tr>
<td>Data citation</td>
<td>• Data citation</td>
</tr>
<tr>
<td>Monograph publishing</td>
<td>• Monograph publishing</td>
</tr>
<tr>
<td>Multimedia production</td>
<td>• Multimedia production</td>
</tr>
<tr>
<td>OER</td>
<td>• OER</td>
</tr>
<tr>
<td>Digital preservation</td>
<td>• Digital preservation</td>
</tr>
<tr>
<td>Digitization</td>
<td>• Digitization</td>
</tr>
<tr>
<td>DMPs</td>
<td>• DMPs</td>
</tr>
<tr>
<td>Open access</td>
<td>• Open access</td>
</tr>
<tr>
<td>Policy compliance</td>
<td>• Policy compliance</td>
</tr>
<tr>
<td>Publication ethics</td>
<td>• Publication ethics</td>
</tr>
<tr>
<td>DOIs</td>
<td>• DOIs</td>
</tr>
<tr>
<td>ETDs</td>
<td>• ETDs</td>
</tr>
<tr>
<td>Funding mandates</td>
<td>• Funding mandates</td>
</tr>
<tr>
<td>RDM</td>
<td>• RDM</td>
</tr>
<tr>
<td>Software development</td>
<td>• Software development</td>
</tr>
<tr>
<td>Subject archives</td>
<td>• Subject archives</td>
</tr>
<tr>
<td>GIS</td>
<td>• GIS</td>
</tr>
<tr>
<td>Grant writing</td>
<td>• Grant writing</td>
</tr>
<tr>
<td>Institutional repository</td>
<td>• Institutional repository</td>
</tr>
<tr>
<td>Text and data mining</td>
<td>• Text and data mining</td>
</tr>
<tr>
<td>Visualization tools</td>
<td>• Visualization tools</td>
</tr>
<tr>
<td>Website development</td>
<td>• Website development</td>
</tr>
</tbody>
</table>
Library as Platform

A library as platform would give rise to messy, rich networks of people and ideas, continuously sparked and maintained by the library’s resources.

A library as platform is more how than where, more hyperlinks than container, more hubbub than hub.

As a platform, a library will serve its users better, serve more users, better accomplish its cultural and educational missions, and build a bulwark of value against looming cutbacks.

Further, and crucially, a library platform can continuously increase its value by providing access to that which is built on it.

Dr. David Weinberger is a senior researcher at Harvard's Berkman Center for Internet & Society, co-director of the Harvard Library Lab

https://www.libraryjournal.com/?detailStory=by-david-weinberger
Library Platform Prospective

Library Manage System (ILS)

- Aleph 500
- Sierra
- Symphony
- ....

Library Service Platform (LSP)

- Alma
- WMS
- BLUEcloud
- FOLIO

Next generation and New Platforms

- Beyond Client/Server Computing
- Cloud-based Architecture
- SaaS/DaaS Platforms
- Web-Based Interfaces
- Multi-Tenant Platforms
- Services-Oriented Architecture
- APIs Exposed for Extensibility & Interoperability

DS Platform!
Support DS with more Platforms

Mass digitization - coupled with new media, technology and distribution networks - has transformed what’s possible for libraries and their users.

Smart DS Platform will support more digital data, tools and environment.
Harvard-Yenching Library

East Asian studies

The largest university library for East Asian research outside of Asia.

DS IN HYL
The Digital Scholarship Support Group (DSSG) provides faculty, students, and staff interested in incorporating digital methods into their teaching and research with a single point of entry to the many resources available at Harvard.

https://dssg.fas.harvard.edu/about/
The First Four Years of the DSSG

Providing Instruction in Digital Scholarship Theory, Tools, and Methods
- Foundations Seminar Series
- Workshops
  - Introduction to Digital Humanities: an Online Course for HarvardX
- Digital Teaching Fellows program
- DSSG Office Hours
  - Individualized Course and Research Project Support

Supporting Teaching and Research
- Advocacy and Building Community
  - Training for Librarians
  - First Annual Boston-Area Digital Scholarship Symposium
  - Digital Futures Discovery Series

Developing Infrastructure
- Omeka as a Service
- Digital Toolkit

https://dssg.fas.harvard.edu/about/first-four-years/
## Digital Toolkit (DTK)

### Anaconda Python
- A Python distribution provides for a Python project quickly and easily, particularly in data science work.

### Cytoscape
- Originally for biological networks visualization, and it is also a general platform for complex network analysis and visualization.

### Gephi
- Network visualization and analysis

### Mallet
- Command line package for natural language processing

### QGIS Desktop
- Open source GIS program

### Zotero
- Research tool that helps for research output management

[https://atg.fas.harvard.edu/toolkit](https://atg.fas.harvard.edu/toolkit)
East Asian Digital Scholarship Lab
At the Harvard-Yenching Library

Home

"Libraries have always been in the business of knowledge creation and transfer, and the digital scholarship incubator within the library can serve as a natural extension of this essential function" - The University Library as Incubator for Digital Scholarship

The East Asian Digital Scholarship Lab aims to serve as a starting point for scholars and students interested in digital scholarship in East Asian studies. It exhibits ongoing or completed projects that can be used as models and provides information about databases, tools, and methods that are used in creating these projects. The rest of the site is arranged according to the usual workflow of digital scholars, starting from finding and managing data and moving subsequently to analysis and presentation.

News and Highlights

- Doing Digital Scholarship at SSRC

https://labs.ssrc.org/dds/

Doing Digital Scholarship offers a self-guided introduction to digital scholarship, designed for digital novices. It allows you to dip a toe into a very large field of practice. It starts with the basics, such as securing web server space, preserving data, and improving your search techniques. It then moves forward to explore different methods used for analyzing data, designing digitally infected teaching assignments, and creating the building blocks required for publishing digital work.

Incubator for East Asia Digital Scholarship / Digital Humanities in Harvard

Platform Perspective for Library Digital Scholarship Support
A Cyberinfrastructure for Historical China Studies

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.

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。

王宏甦，徐力恆，包弼德 CBDB, Harvard University
Open and linkable API system

Data-bases

Tools

Platforms

16

Platform Perspective for Library Digital Scholarship Support

April 23, 2020
APIs Linking Backend and Frontend in Online Systems

FOR THE OWNER

Servers
Systems Maintenance
Control Over Data

FOR THE USER

Adding Data
Correcting Data
Querying Data
Data Dumps
Visualization

= Supervised Crowdsourcing
Library Support

Cyber infrastructure

Metadata Design

Platform Co-founder

Data Support

Database API Broker
<table>
<thead>
<tr>
<th>The Center for Geographic Analysis (CGA) of Harvard University and the China Data Institute</th>
<th>A cloud-based data analysis platform for data gathering, management, analysis, visualization, and sharing.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goal: to build a cloud platform for data research and training</td>
<td>To build a core infrastructure for disseminating and utilizing spatiotemporal data, particularly data from China, to support case-based study and training.</td>
</tr>
<tr>
<td></td>
<td>Support case-based training programs on data analysis for economic, social, public health, urban planning, and other research subjects.</td>
</tr>
<tr>
<td></td>
<td>Provide a secure dissemination channel for data providers, and eliminate the burden of downloading data, installing analytical software, and managing data and tools locally from researchers.</td>
</tr>
</tbody>
</table>

https://projects.iq.harvard.edu/chinadatalab/home
Existing Workflow in China Data Lab

Network analysis and text mining on innovation data with **KNIME**

Machine learning for industrial co-agglomeration analysis with **KNIME**

Regional dynamics of China with **KNIME & Alteryx**

Space-time analysis of high education in China with **KNIME & Alteryx**

OCR and text mining on historical books with **KNIME**

Air quality analysis on social media data with **Jupyter**

...
Network analysis in CDL with CBDB data

KINME
## Network analysis in CDL with CBDB data

### Quantitative Indicators

<table>
<thead>
<tr>
<th>朝代</th>
<th>宋</th>
<th>明</th>
<th>元</th>
<th>唐</th>
<th>清</th>
</tr>
</thead>
<tbody>
<tr>
<td>节点（人）数</td>
<td>13021</td>
<td>6305</td>
<td>4784</td>
<td>5688</td>
<td>2347</td>
</tr>
<tr>
<td>边数</td>
<td>102645</td>
<td>26201</td>
<td>16869</td>
<td>16721</td>
<td>4394</td>
</tr>
<tr>
<td>平均度</td>
<td>7.88</td>
<td>4.16</td>
<td>3.53</td>
<td>2.94</td>
<td>1.87</td>
</tr>
<tr>
<td>核心节点1</td>
<td>3678(朱熹)</td>
<td>318(李維楨)</td>
<td>511(吳澄)</td>
<td>309(白居易)</td>
<td>149(紀昀)</td>
</tr>
<tr>
<td>核心节点2</td>
<td>1635(蘇軾)</td>
<td>317(王世貞)</td>
<td>499(王惲)</td>
<td>251(杜甫)</td>
<td>80(錢大昕)</td>
</tr>
<tr>
<td>核心节点3</td>
<td>1426(周必大)</td>
<td>297(呂柟)</td>
<td>417(虞集)</td>
<td>205(李白)</td>
<td>60(張繹英)</td>
</tr>
<tr>
<td>最大联通组</td>
<td>12087(92.3%)</td>
<td>5809(92.1%)</td>
<td>3747(78.3%)</td>
<td>4815(84.7%)</td>
<td>1231(52.4%)</td>
</tr>
<tr>
<td>平均距离</td>
<td>13.90</td>
<td>9.68</td>
<td>8.69</td>
<td>8.46</td>
<td>14.96</td>
</tr>
<tr>
<td>最大距离</td>
<td>70</td>
<td>42</td>
<td>32</td>
<td>24</td>
<td>36</td>
</tr>
<tr>
<td>Assortativity</td>
<td>-0.1983</td>
<td>-0.2525</td>
<td>-0.3180</td>
<td>-0.3291</td>
<td>-0.1677</td>
</tr>
<tr>
<td>聚类系数</td>
<td>0.1317</td>
<td>0.1211</td>
<td>0.1766</td>
<td>0.1113</td>
<td>0.0476</td>
</tr>
</tbody>
</table>
Network analysis in CDL with CBDB data

Visualization

Platform Perspective for Library Digital Scholarship Support

April 23, 2020
Harvard-Yenching Library

East Asian studies

The largest university library for East Asian research outside of Asia.

NEW PLATFORMS DESIGN
The Dilemma of Humanities Scholars

- The original form of resources is not conducive to develop research
- The huge number of resource
- Limited ability to obtain available information
- Hard for them to process data with advanced technologies
- Library cooperated with the related departments of Humanities
## digital humanities research platforms

<table>
<thead>
<tr>
<th>Function comparison of different DH research platforms</th>
<th>CABDHRP</th>
<th>CULTURA</th>
<th>CText</th>
<th>Scripta Sinica database</th>
<th>THDL</th>
<th>CBETA research platform</th>
<th>MARKUS semi-automatic text annotation system</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full-text data collection</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>X</td>
</tr>
<tr>
<td>Automatic segmentation of Chinese words</td>
<td>○</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Automatic annotation</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>X</td>
<td>○</td>
<td>X</td>
<td>○</td>
</tr>
<tr>
<td>Term search</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>External reference</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Word frequency statistics</td>
<td>○</td>
<td>○</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>○</td>
</tr>
<tr>
<td>Bookmark note</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>○</td>
<td>○</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Social network visualization tool</td>
<td>○</td>
<td>○</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>○</td>
</tr>
<tr>
<td>Recommender</td>
<td>X</td>
<td>○</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Recording use process</td>
<td>○</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Open API</td>
<td>X</td>
<td>X</td>
<td>○</td>
<td>X</td>
<td>X</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>

A Chinese ancient book digital humanities research platform

The system architecture of the developed CABDHRP
Transform the physical form into electronic form to store and use

Two levels of digitalization

Repaired first and then scanned

Conductive to long-term preservation
Easy to use and disseminate

Yin, Q. and Zhuoyuan, X. and Xiaohua, S.*. From Collection Resources to Intelligent Data: Construction of Intelligent Digital Humanities Platform for Local Historical Document. Digital Scholarship in the Humanities. (Accepted, SSCI)
Self-built Metadata Scheme

From “reading” to “analysis”

Based on the reuse of DC, 18 metadata elements had been defined, including 4 custom elements

Realizing Multiple Associations

“Chinese Year Number”

Physical Resource Association

Chinese-Western Calendar Association

Chinese Year Number ➔ 1920

Gregorian Year

House Land/Mountain Goods, Rights, etc.

From "reading" to "analysis"
Consideration on DS Platform Construction

- Retrieve and Discover Data
- Community Communication
- Application of Artificial Intelligence
- Open Data Service Support
- Text Analysis and Statistics
- Preserve and Manage Research Data
- Analyse and Display Research Data
- Application of Artificial Intelligence
- Community Communication
- Open Data Service Support
- Text Analysis and Statistics
- Preserve and Manage Research Data
- Analyse and Display Research Data

Platform Perspective for Library Digital Scholarship Support
Platform Architecture Design

4. Resource Utilization

3. Data Processing

2. Resource Storage

1. Resource Collection
Platform Function Module Design

Intelligent Digital Humanities Platform

- Scholar Application Module
- Management Application Module
- Support Service Module
Scholar Application Module

Resource Utilization
- Documents Retrieval
- Documents Browse
- Documents Download
- Document Data Feedback

Research Projects Management
- Research Teams Management
- Research Documents Management
- Research Outcomes Management

Personal Information Management
Management Application Module

- User Identity Authentication
- Access Privilege Control
- Maintenance and Management of Document Data
  - Metadata Management
  - Data Upload
  - Full Text Input (Image upload)
  - Data Feedback Review
Support Service Module

- User Behavior Data Collection
- Personalized Resource Recommendation
- Visualization Tools Analysis Tools
- Community Communication
- Data Statistics
- Others
  - Image Cutting
  - Character Recognition
  - ....
上海交通大学地方历史文献数据平台
View by Different Categories

Platform Perspective for Library Digital Scholarship Support

Contract Documents

Litigation Documents
<table>
<thead>
<tr>
<th>页码</th>
<th>内容</th>
</tr>
</thead>
<tbody>
<tr>
<td>38</td>
<td>Platform Perspective for Library Digital Scholarship Support</td>
</tr>
</tbody>
</table>
View by Different Categories

Data Management

- Metadata Management
- Data Upload
- Full-text Input
- Image Upload
- Data Feedback Review
View by Different Categories

Project Management Page

- Project Related Settings
- Project Team Management
- Project Documents Management
- Project Outcomes Management
Data Interaction

- Add Annotations
- Decide Who Can See
Harvard-Yenching Library

East Asian studies

The largest university library for East Asian research outside of Asia.

CONCLUSION
Conclusion

Digital Scholarship can provide a new research paradigm and a new academic perspective for the university research, especially in arts and humanities.

Digital Scholarship is the systematic extension of Library traditional service.

Library will play an important role in resource organization, resource preservation (management), technology support and platform service in the process of Digital Scholarship research support.
References


7. Digital Scholarship at Harvard University. https://dssg.fas.harvard.edu/


Thank you

Q&A