Visualizing the landscape of Digital Scholarship in Japanese Studies

Keiko Yokota-Carter
Japanese Studies Librarian

Cecilia Caride
Special assistant for Japanese Collection

“Breaking barriers – unlocking Japanese resources to the world”
EAJRS Annual Conference
September 16 – 19 2015
Acknowledgements

This presentation is based on consultations and advice from:

• Ms. Toshie Marra, C.V.Starr East Asia Library, University of California at Berkeley
• Dr. Hoyt Long, University of Chicago
• Todd Hamish, British Library
• Mr. Yuta Hashimoto, Kyoto University
• Professor Kiyonori Nagasaki, Tokyo University
• Professor Ryo Akama, Ritsumeikan University
• Mr. Jeremy York, Hathi Trust
• Professor Alessandro Bianchi, Freer & Sackler Galleries, Smithsonian Institution
• Professor Misa Nakamura, National Institute of Japanese Literature
• Kinokuniya Bookstore
• Maruzen Bookstore
1. Basic steps for Research Activities in Digital Scholarship

2. Environmental scan of the condition of resources to support digital Japanese studies scholarship outside Japan

3. Examine major projects according to the “Basic steps for Research Activities in Digital Scholarship”

4. Call for collaboration to sustain humanities/social science, and build the future of global Japanese studies
Basic Steps for Research Activities in Digital Scholarship

Professor Ryo Akama (Ritsumeikan University) & Professor Kiyonori Nagasaki (Tokyo University)

1. Digitize analog materials (including transcription and encoding in standardized form*).

2. Establish an archival environment for born-digital and digitized materials.

3. Combine digitized and born-digital materials into a database structure.

4. Using a variety of software tools, analyze the materials in the database.

5. Digitally archive the information pertaining to or resulting from the analysis, and structure it as a database.

6. Use Social Networking Services (SNS) and other appropriate tools to network the above research cycle throughout the community.

*PDF / Images more common than transcribed text

Group or Individual projects

Step 6 becomes much easier and more efficient when data is standardized.
Environment Scan

Declining...

Support for Humanities in US and elsewhere

Number of students in Japanese Studies

JAPAN 国立大学「文系学部廃止政策」

Rising...

Information technology
Digital Resources
Big Data analysis

Digital Scholarship / Digital Humanities

How can we create a brighter future for HUMANITIES / SOCIAL SCIENCES?

U-M Library Circulation 2013-2014
total volumes: 1,221,989

print sources checked out
e-book sections downloaded
e-journal articles downloaded
Student Research Cycle example: “Electoral behavior”

- Lecture/paper
  - “Public opinion and electoral behavior”

- Acquire resource
  - Print edition
  - Jiji seron chosa tokuho

- Convert Text / Data
  - Unusable Licenses
  - Unusable Formats (CD-ROM, PDF etc.)

- Writing

- Analysis / Synthesis / Discovery

- Subject librarian - Selector
  - Vendor license negotiation
  - Build collection

- Open Data / text repository
Research example: “Digital Humanities in Japanese Studies”
Courtesy of Dr. Hoyt Loyd, University of Chicago -- Presented at CEAL 2015

100,000+ Poems
4,000+ Poets
166 Journals

Hand input (worked with TEI Librarian)

Scanning
OCR

Database

3. Complex network analysis to explore patterns of relation within the data (using Excel and Gephi)
# Limited choice of Commercial E-books

<table>
<thead>
<tr>
<th></th>
<th>Maruzen E-Book Library</th>
<th>Net-Library (Kinokuniya) at EBSCO</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Meta-data (US MARC)</strong></td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td><strong>Available titles</strong></td>
<td>Most are e-reprints of multi-volume sets or monographs published before 2013</td>
<td>Almost same titles as Maruzen</td>
</tr>
<tr>
<td><strong>New publication</strong></td>
<td>Started * E-books available half-year after print version</td>
<td>Started</td>
</tr>
<tr>
<td><strong>Humanities &amp; Social Science</strong></td>
<td>東大出版会</td>
<td>勝応出版会</td>
</tr>
<tr>
<td></td>
<td>みすず書房</td>
<td>勁草書房</td>
</tr>
<tr>
<td></td>
<td>勁草書房</td>
<td></td>
</tr>
<tr>
<td></td>
<td>吉川弘文館</td>
<td></td>
</tr>
<tr>
<td></td>
<td>有斐閣</td>
<td></td>
</tr>
<tr>
<td><strong>Pricing</strong></td>
<td>Books 2.5 times of prints, 6 times of print journal.</td>
<td>Books 2.5 times of print version</td>
</tr>
<tr>
<td></td>
<td>$ / Y exchange rate price of original Japanese price</td>
<td>Fixed $ price of original Japanese price</td>
</tr>
<tr>
<td><strong>Discovery Tools</strong></td>
<td>Summon, EBSCO, ExLibris</td>
<td>Worldcat</td>
</tr>
<tr>
<td><strong>Demand Driven Acquisition</strong></td>
<td>?</td>
<td>YES</td>
</tr>
<tr>
<td><strong>Text-mining</strong></td>
<td>?</td>
<td></td>
</tr>
</tbody>
</table>
Scattered resources

- Aozora bunko
- Research institutions
  - Shibata History Database
- Museums
- Japanese Text Initiative
- Digital texts in universities
- National Institute of Japanese Language Research Project
- Individual projects
- Library of Congress
- National Diet Library
- Archives JCAR
- Japan Knowledge

What materials are available?

Union list of resources ??
Europeana  “Japan version” possible ???

Matches for: Japan

View through waves off the coast of Kan...
Krab met kersenbloesem
De toren en de grote poort van Asak...
Japanese champion boxer arrested f...

Matched results:

Results per page: 24

Results 1 - 24 of 51,648

Search: Japan

Refine your results:

By media type:
- IMAGE (46611)
- TEXT (4322)
- VIDEO (518)
- SOUND (190)
- 3D (7)

By language of description

By year

By providing country

Can I use it?

By copyright

By provider

By data provider

Include content contributed by users
Effort - Gateway to resources : Discovery tool

Example: University of Michigan Library search portal

- Japanese books in Library
- JapanKnowledge
- 国立国会図書館
  - 近代デジタルライブラリー
  - 青空文庫
- Cinii
- Hathi Trust

Gather them at ONE place to Break barriers – gateway to Japanese resources
Step 1, 2, 3 for the Digital research activities:

Gather digital source in ONE place, preserve and make them available

Preservation and Access managed by Academic Community
Union repository of digitized texts from libraries in US and beyond

Hathi Trust
Page images
Digital text
Print original
in ONE place!

Digitized by Google
+ Digitized by University
+ Preserve paper original

University of California

Harvard

Michigan

Keio

And more!
and unlocking Japanese resource to the world

Worldwide public domain works in Hathi Trust

- Works published before 1875 (over 140 years ago)
- 8,682 titles total (Keio University contributions = 87% of total)

Worldwide view verified by Todd Hamish, Head of East Asia Collection at the British Library
USE Hathi Trust KEIO University materials through Michigan catalog
Hathi Trust: Features and Limitations

- Public Domain
- Page images available for download
- Transcribed text
- OCR accuracy still needs improvement
- No text at all for kuzushiji
Step 4,5,6 for the Digital research activities:
Hathi Trust Research Center

Gather them at ONE place and
Breaking barriers – unlocking Japanese resources
to the world, **Create new knowledge**

**Modes of Access**

HTRC hides the complexity of computational investigation of the massive HathiTrust corpus. A researcher engages with HTRC through a complexity-hiding interface as shown in the figure. The interface provides a web portal and a software programmatic interface. HTRC brings together various text mining tools, the HathiTrust corpus, aggregated and statistical feature information about the corpus, and other data sources that are needed for text mining. Text mining tools are then run on compute resources that are co-located with the HathiTrust data.

**•Virtual workspace**
**•Tools**
**•open beyond partner institutions—ASK Hathi!**

https://www.hathitrust.org/htrc
Share your comments and digitized resources: feedback@issues.hathitrust.org
Best Practice for Data/Research cycle

SAT Daizōkyō Text Database

Data and Research Cycle

- Transcription and encoding
- Creative Commons Database
- Share research around the world
- Improve texts and database

Manuscripts, woodcuts, commentary, translations...

大正新脩大蔵経

サーガニッティタム・タイストリピタカメ
The SAT Daizōkyō Text Database
Best Practice of classical Japanese materials

国文学研究資料館
（国文研，National Institute of Japanese Literature）

Directory of Overseas Collections of
Old and Rare Japanese Books, Other Print Materials and Manuscripts
在外日本古典籍所蔵機関ディレクトリ

• Sub-committee of Japanese rare books of the Japanese Material Committee, Council of East Asia Libraries

• European university libraries

• Collaboration with 164 institutions outside Japan

http://base1.nijl.ac.jp/~overseas/index-e.html

Asian and African Studies, The British Library
Asian Art Museum of San Francisco, C. Laan Chun Library
Asian Languages Collection, University of Auckland Library and Learning Services
Asian Library, University of British Columbia
Asian Studies Research Collection, Sir Louis Matheson Library, Monash University
Auckland, N.Z. University Library
Biblioteca Don Bosco, Fondo Marega
Bibliothek, Japanisch-Deutsches Zentrum Berlin
Bibliothèque du musée national des arts asiatiques Guimet
Bibliothèque nationale de France, Département des manuscrits
Blacker-Wood Collection of Zoology and Ornithology
Bodleian Japanese Library, University of Oxford
Brigham Young University, Harold B. Lee Library
British Library, Asian and African Studies
C. Laan Chun Library
C. V. Starr East Asian Library (Columbia University)
国文研 Union Catalogues

- 国文学研究資料館
- European universities
- Harvard University, University of California at Berkeley and Los Angeles

古典籍大型プロジェクト
日本の大学図書館デジタル資料

HATHI TRUST
Digital Library

古典籍書誌情報を加えデジタル資料にリンクは可能
世界最大の古典籍データベース?
SMART-GS: Cloud collaboration space for classical texts

Diagram courtesy of Yuuta Hashimoto, Tokyo University
Translation by Cecilia Caride
The Tale of Genji (絵入源氏物語, published in 1650)

Project members

Role
Owner
Contributor

User name
ynashimoto
user1

SMART-GS interface

Kiritsubo No Koi (桐壌更衣). The mother of Hikaru Genji. Try searching this image.
We work together!
Scattered projects

Aozora bunko
Research institutions

Japanese Text Initiative
Museums

Individual projects
Universities

(Projects hub) ??

What are others working on?

Library of Congress
国会図書館
archives
Connect and share

Hubs for sharing projects, updates, information...
Back to Reality outside Japan

1. Digitize analog materials (including transcription and encoding in standardized form*).

2. Establish an archival environment for born-digital and digitized materials.

3. Combine digitized and born-digital materials into a database structure.

4. Using a variety of software tools, analyze the materials in the database.

5. Digitally archive the information pertaining to or resulting from the analysis, and structure it as a database.

6. Use Social Networking Services (SNS) and other appropriate tools to network the above research cycle throughout the community.

Conditions in U.S. Libraries
- license problems
- unusable text formats
- statistical data as PDFs...
- Publishers don't talk with researchers
“We are in the RIGHT direction”
Brake barriers among sectors
Discuss and move forward in virtual spaces....

Apply to join JpnLib email group
...and in person

Great ideas

new "Union Conference"

- Researchers
- Industry (platforms, software, licensing)
- Archives
- Libraries
- Museums
- Government (grants, legal issues)

Tech prowess

Fundraising skills

Gather in ONE place and break barriers to unlock Japanese resources to the world

Create new knowledge for humanity
Dank u wel!

Since April 19th, 1600 -, breaking barriers and open to the world.

World wide use Full texts