

国 文 研

National Institute of
Japanese Literature

NIJL

Workshop on Japanese old and
rare books in HNL2017
Feb. 17, 2017

Future of the Network for Research on Japanese Classical Books

National Institute of Japanese Literature
Research Department
Associate Professor

Junko Koyama (presenter)

prepared by Center for Collaborative Research on Pre-modern Texts,
National Institute of Japanese Literature



National Institute of Japanese Literature
National Institutes for the Humanities
Inter-University Research Institute Corporation

<http://www.nijl.ac.jp/>

1. Current situation of the project

- Project for advancing the frontiers of large-scale academic research efforts

Project to Build an International Collaborative Research Network for Pre-modern Japanese Texts

- Term: **2014-2023**

- Goals:

- (1) **To make publically available on the internet a collection of 300,000 digitized Japanese classic books from various fields, along with their bibliographies (71,000 books will be digitized by March 2016).**
- (2) To promote collaborative international research projects through the use of digitized materials.
- (3) To build networks for collaborative international research.



A new public image viewing system starts in April, 2017.

2. Collaborative institutions

20 universities

National Institute for the Humanities

- National Museum of Japanese History
- National Institute of Japanese Language and Linguistics
- International Research Center for Japanese Studies

Other institutions

Digitization of classic books in other institutions
Cooperation in building databases
by the **National Institute of Informatics.**

国 文 研

NIJL



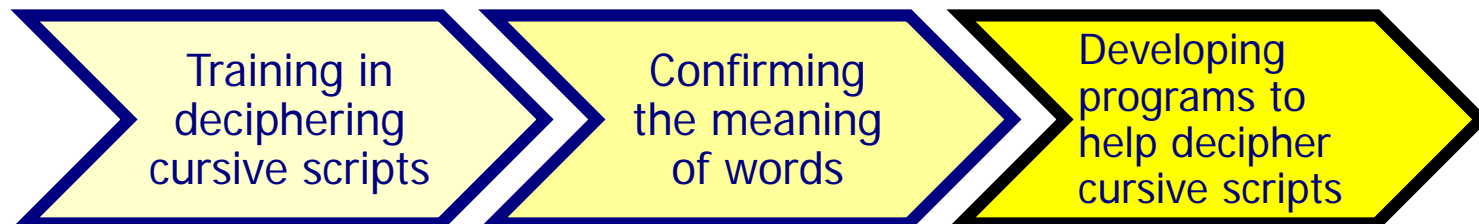
Foreign institutions with collaborative agreements

Collège de France
Korea University
University of Florence
Beijing Foreign Studies University
Leiden University
University of Venice
University of Naples
Sapienza University of Rome
Vatican Apostolic Library
University of Cambridge
University of British Columbia
Columbia University
C. V. Starr East Asian Library at
University of California

**Collaborative efforts will be strengthened
and broadened as this project progresses**

3 .Two points of view for classic studies

■ How to read a classic



The Hentaigana App

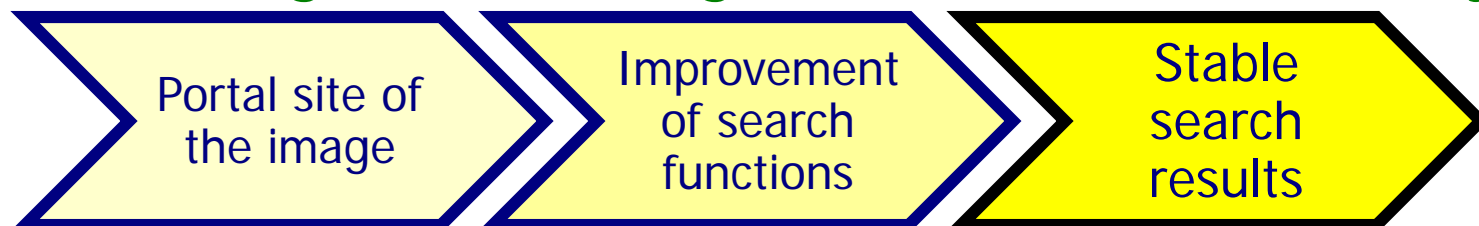
KuLA

JapanKnowledge

SMART-GS

Classical Books
OCR

■ How to gather intelligence data for the study



N
I
J
L

User-friendly search

Cross-searching

Tagging

TEI-conformant texts

Embedding DOI



+ Tool development support :

Open Data Ideathon · hackathon

4. Portal site of Classical Books

Improvement of information and image availability
Not individually but comprehensively

Cross-searching

via NIJL's authorized-management site, not through individual sites.



encouraging unexpected findings

NIJL (Platform for Japanese classics)

Partner

Partner

Partner

Partner



and
others

5. The reason why NIJL has taken on this project

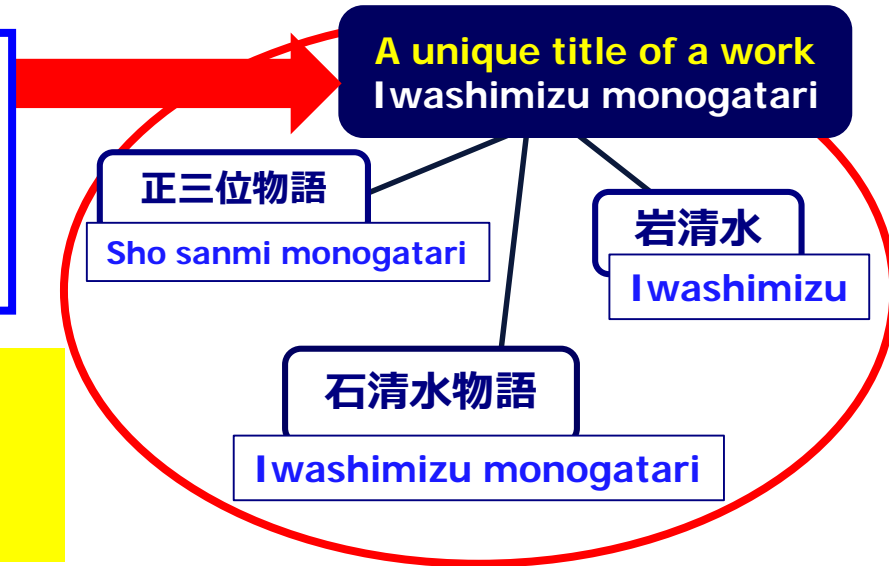
《To arrive at necessary information》

- Each library has developed a catalog of its collection of classic books.
- NIJL has authorized management regarding the catalogs and micro films of classic books.

Advantage of NIJL:
authorized management
of classics

【Renewed URL of Integrated Catalogue of
Early Japanese Books】

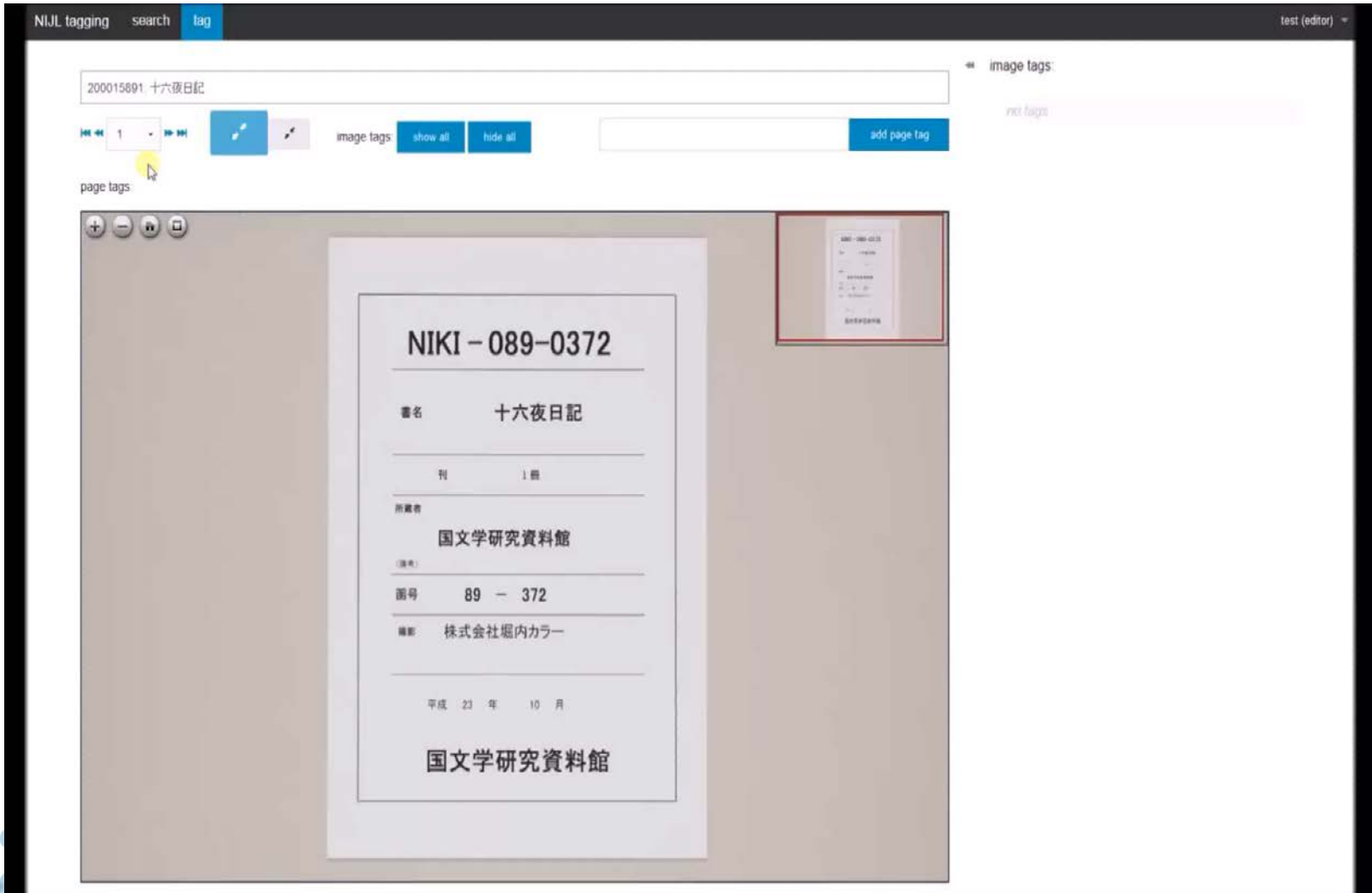
<http://base1.nijl.ac.jp/~tkoten/>



- NIJL has been developing a comprehensive bibliographical database of Japanese classics, which includes the "Integrated Catalogue of Pre-modern Japanese Books" , an index of classic books from all fields.

6. Improved search functions①

Addition of tags to individual images

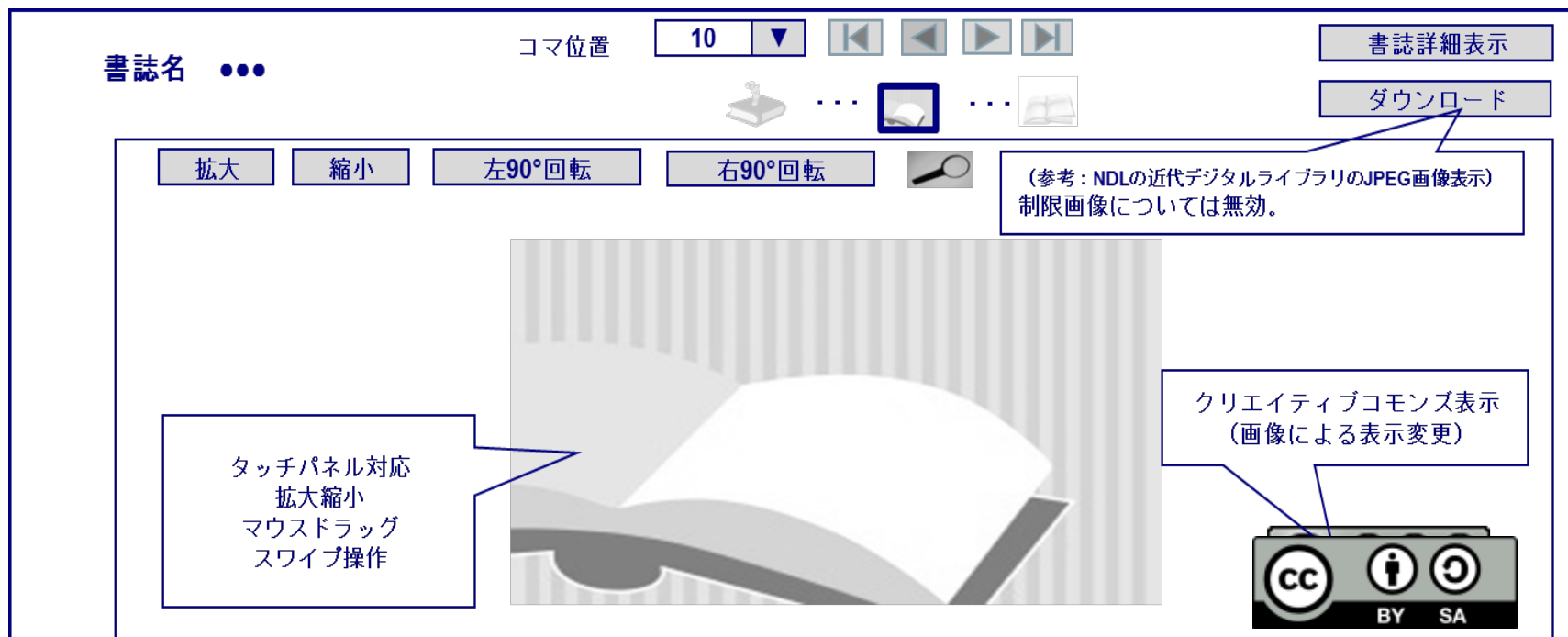


click

Developing a tagging tool on the Web

7. Improved search functions②

Social tagging will commence in April, 2017 (initially by researchers)



ソーシャルタギング

(表示されている画像の検索語の提案)

提案

Social tagging

図4：画像ビューア画面例

8. Improved search functions③

An example of research by Dr. Kengo Terasawa

Application:

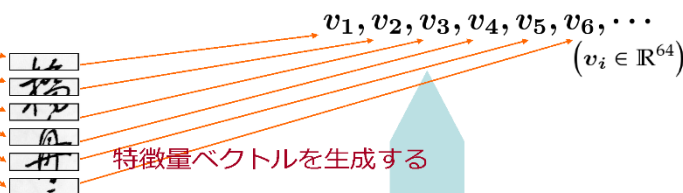
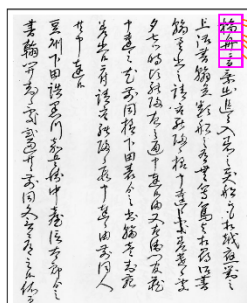
Locate a range of images which share similar shapes in a given image

文字認識されていない典籍の画像から、与えられた単語画像と似ている画像を検索し、出現位置を特定する。

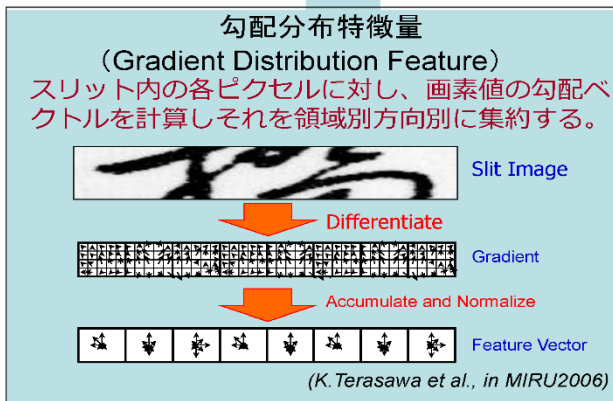
⇒ **semi-automated tagging**

Algorithm for **word spotting**

[1] Scanned Image [2] Clip by Sliding Window [3] Descriptors (Feature Vectors) → [4] Matching

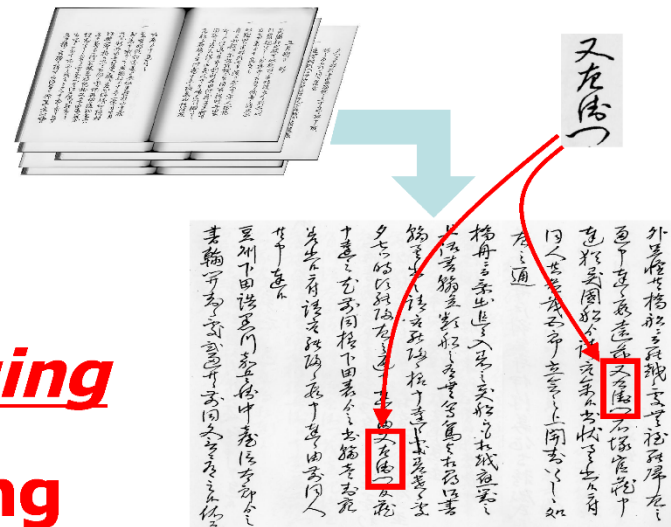


特徴量ベクトルを生成する



(K. Terasawa et al., in MIRU2006)

文書画像をスリット状に切り出した後、各スリット画像を特徴ベクトルで記述。



特徴量ベクトルを用いてベクトルのシーケンスのマッチング問題として解く。

Advantages:

No need to manually plot a range of tags on a given image.

Feasible even for cursive writing.

Applicable for various writing systems.

9 . Improved search functions④

■Reduced labor for reprinting classical Japanese books

[http://doi.org/](http://doi.org/10.1241/johokanri.58.819)

10.1241/johokanri.58.819

- ≡ A character which the software was unable to decode
- A character which the software decoded incorrectly

■Efficient transcription of necessary books

■TEI-conformant texts are currently under consideration.

click

10. Consistent search results

DOI

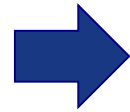
Embedding **digital object identifiers (DOI)** into each digitalized image of a classical book, and recommendation to cite the DOIs in academic papers.

This makes it easy for readers to refer back to digital images of books referred to in academic papers.

This process was very difficult in previous research.

⇒ Research should be verifiable.

| | |
|-------|-------------------|
| タイトル | 白縫譚. 第65編 |
| 著者 | 柳亭種彦, 柳水亭種清 作 |
| 出版者 | 柳亭種彦 |
| 出版年月日 | [18--] |
| 請求記号 | WA19-6 |
| 書誌ID | |
| DOI | 10.11501/2545205 |
| 公開範囲 | インターネット公開(保護期間満了) |



Toward verifiable humanities

ex.NDL

■ Most work in the humanities has depended on drawing upon fragments and citations of other researchers. However, we must now work to verify these fragments for ourselves.

Ex. Search for the word "aurora"

Search functions mentioned above

The Database

- 刊写
 - 写本 (5)
 - 写本 (11)
 - 形態
 - 冊子 (12)
- 1 伊勢物語古意 刊 国文研
 - 2 真名伊勢物語 写 国文研
 - 3 伊勢物語残卷 写 北海道大学

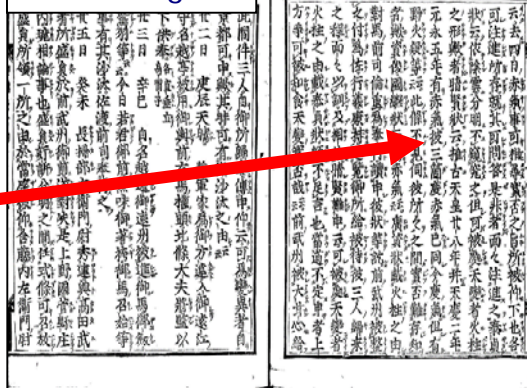
Digitized images

Tags

Finding something significant in unexpected books

We aim to promote original research via extraction of various information from the database.

Azuma-kagami



Enkouan zuikan zue



12. Encouraging further interaction with Japanese classic books

- To encourage public engagement

Published digitized images and bibliographic information under CC BY-SA: available for everyone.

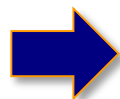


【NIJL】

As an experiment, 350 digitized classic books have been made public under CC BY-SA to promote open data searching. (In the near future, much more of NIJL's collection will be published.)

http://www.nijl.ac.jp/pages/cijproject/data_set_list.html

- From Ideathon to Hackathon



Developing an idea suggested in the ideathon

An ideathon was held by NIJL in Kyoto

| 書名 | 著者 | 刊行年 | 形式 | 公開 | 備考 |
|------|-----|------|----|----|-------|
| 源氏物語 | 紫式部 | 1091 | 漢 | ○ | 10000 |
| 源氏物語 | 紫式部 | 1091 | 漢 | ○ | 10000 |
| 源氏物語 | 紫式部 | 1091 | 漢 | ○ | 10000 |
| 源氏物語 | 紫式部 | 1091 | 漢 | ○ | 10000 |
| 源氏物語 | 紫式部 | 1091 | 漢 | ○ | 10000 |
| 源氏物語 | 紫式部 | 1091 | 漢 | ○ | 10000 |
| 源氏物語 | 紫式部 | 1091 | 漢 | ○ | 10000 |
| 源氏物語 | 紫式部 | 1091 | 漢 | ○ | 10000 |
| 源氏物語 | 紫式部 | 1091 | 漢 | ○ | 10000 |
| 源氏物語 | 紫式部 | 1091 | 漢 | ○ | 10000 |

■会場：メッセ京都6F会議室
〒600-8501 京都市中京区錦町
TEL: 075-753-7444

国文研 アイデアソン やります!

歴史的典籍オープンデータワークショップ
～古典をつかって何ができるか！じんもんそん2015～

13. Encouraging further interaction with Japanese classic books

To encourage utilization not only by researchers but also by the public

国文研データセット簡易Web閲覧

タグ連想検索

よりとも

タグ連想検索

■ 検索語
■ 関連タグ1
■ 関連タグ2

「連想」の語はVannevar Bushの1945年の論文"As we may think"によります。

グラフ中の各タグをクリックすると、そのタグが付けられているページ画像が検索表示されます。同時に、検索窓にもその単語が入りますのでそのまま「タグ連想検索」ボタンをクリックしていただければ、そのタグを中心としたタグの関係が表示されます。なお、文字の大きさは、登場頻度を表しています。矢印は、比較的近いタグであることを示しています。タグの情報が増えれば増えるほど、矢印の意味を深めていくことができます。

検索例: 鎌倉 物徂徠 御番頭 二御丸火之番

Suggested search

IIIF (International Image Interoperability Framework)
D3.js (Data-Driven Documents) etc....

A user-friendly viewer for open data searches
has been released by Dr. Nagasaki

http://www2.dhii.jp/nijl_opendata/openimages.php



Wrap up: Goals of the project

- (1) Promote integrated searching and collaboration as a platform for further research**
- (2) Encourage various institutions to engage in this sharing of information**
 - ※Assisting, collaborating with, and forming frameworks for various projects
 - ※Supporting sustainability of their results

To encourage interaction with Japanese classic books in mutually beneficial relationships

