

ROIS-DS Center for Open Data in the Humanities (CODH)



Asanobu KITAMOTO

Director, ROIS-DS Center for Open Data in the Humanities (CODH) and

Professor, National Institute of Informatics

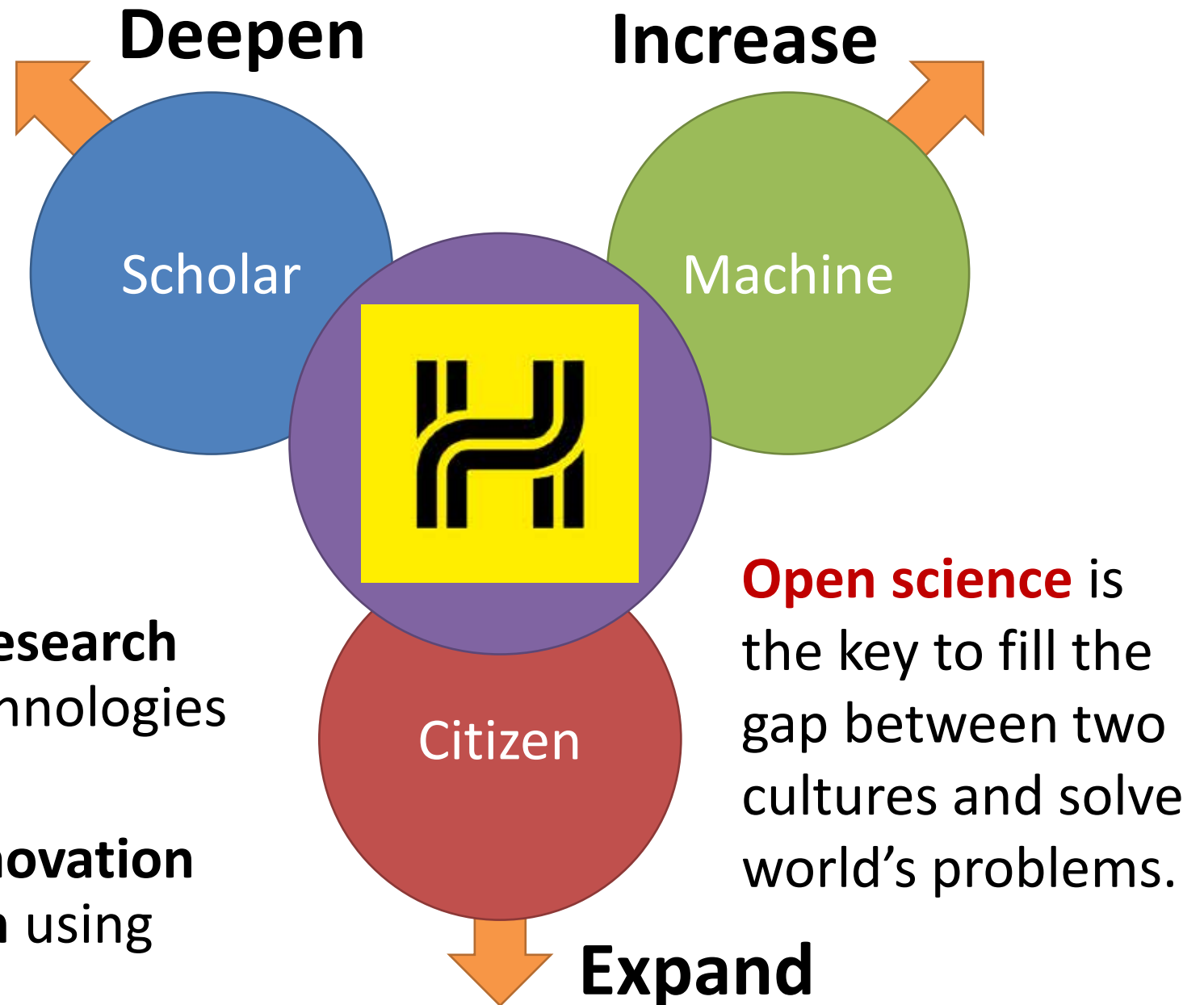
<http://codh.rois.ac.jp/> @rois_codh

ROIS-DS Center for Open Data in the Humanities (CODH)

<http://codh.rois.ac.jp/>

1. Data-driven Humanities:
Innovation in humanities research
using computer science technologies
and tools.

2. Humanities Big Data: Innovation
in non-humanities research using
humanities data.



Open science is
the key to fill the
gap between two
cultures and solve
world's problems.

Textual and Non-textual Digital Humanities

Images



Photographs



Maps



Characters



Digital humanities is not only about text.

Structured and unstructured data (visual and spatial sources) requires its own analysis and interpretation framework.

Open Datasets

<http://codh.rois.ac.jp/dataset/>



Pre-modern Japanese Text (3126 books, 609631 pages)

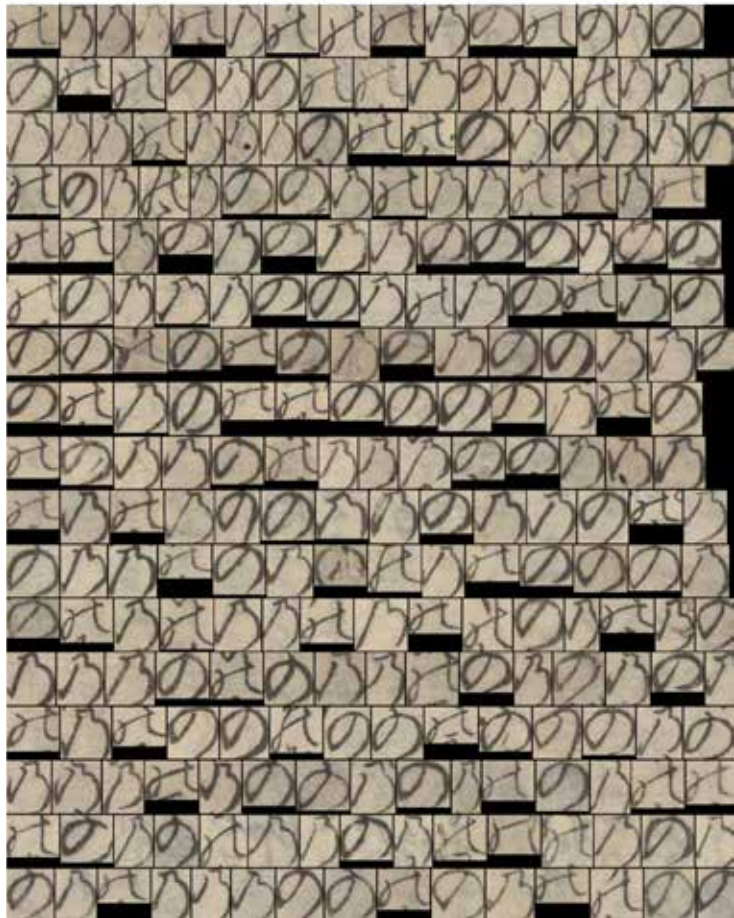


Edo Cooking Recipes (103 egg dishes)

Kuzushiji Dataset

<http://codh.rois.ac.jp/char-shape/>

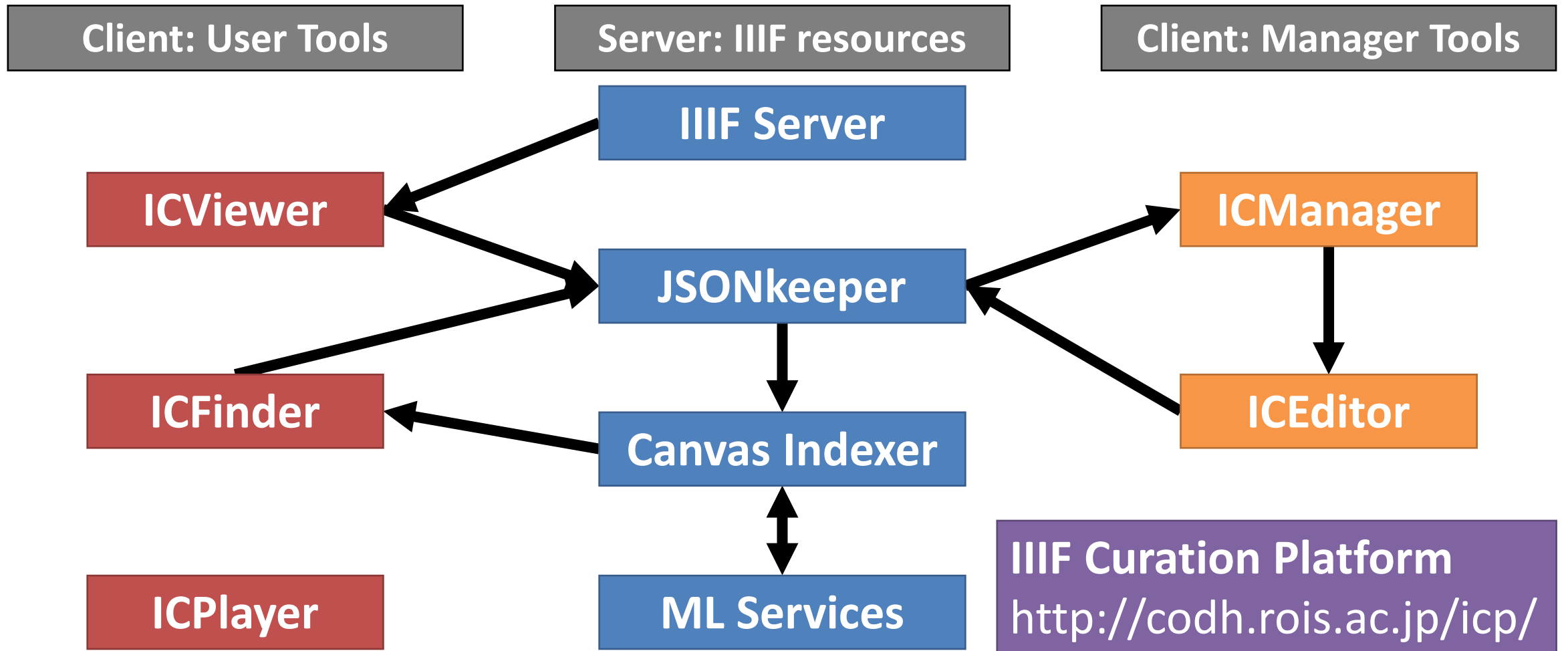
雨月物語 (1890)



- National Institute of Japanese Literature created and CODH curated.
- The open data consists of
 - Character types: 4,328
 - Character shapes: 1,086,326
- Download the Zip file and use it as training data for machine learning.
- The release of dataset stimulated research on AI kuzushiji recognition.

Open Software

<http://codh.rois.ac.jp/software/>



IIIF Curation Viewer

Developed by CODH
since 2016

<http://codh.rois.ac.jp/software/iiif-curation-viewer/>



1. ■ is the "crop" button →
Selects a rectangular region
2. ☆ is the "favorite" button →
Collects regions you need

Open Services and Archives

<http://codh.rois.ac.jp/project/>



edomi – Data Portal for the
Historical Edo



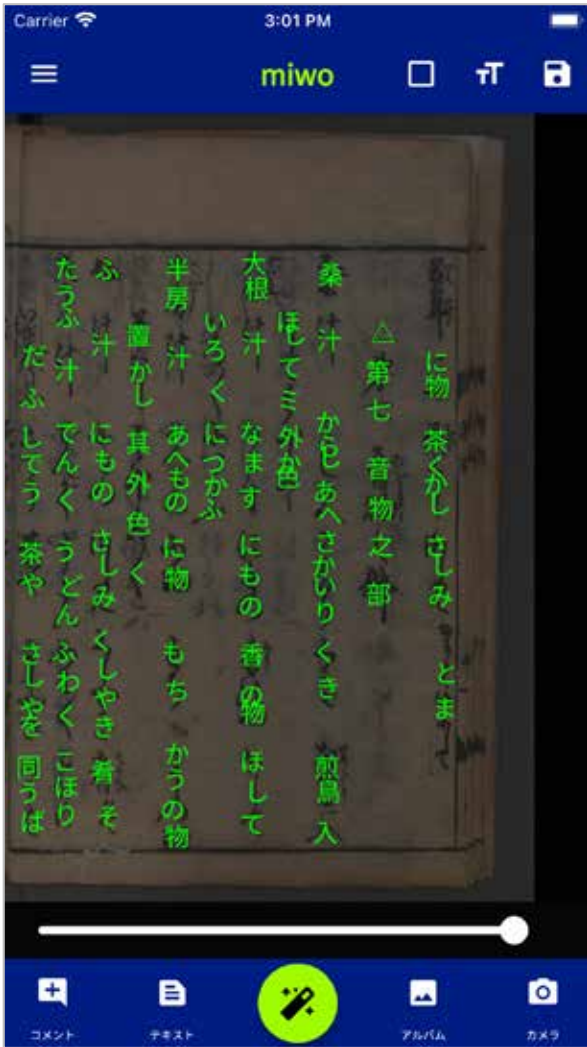
North China Railway Archive



miwo – App for AI Kuzushiji Recognition

<http://codh.rois.ac.jp/miwo/>

1. The name comes from the 14th chapter of **The Tale of Genji** “miwotsukushi,” referring to waterway signs.
2. Just as the miwotsukushi is a guide for boats in the sea, **we aim to make our "miwo" app as a guide for traveling the ocean of historical documents.**



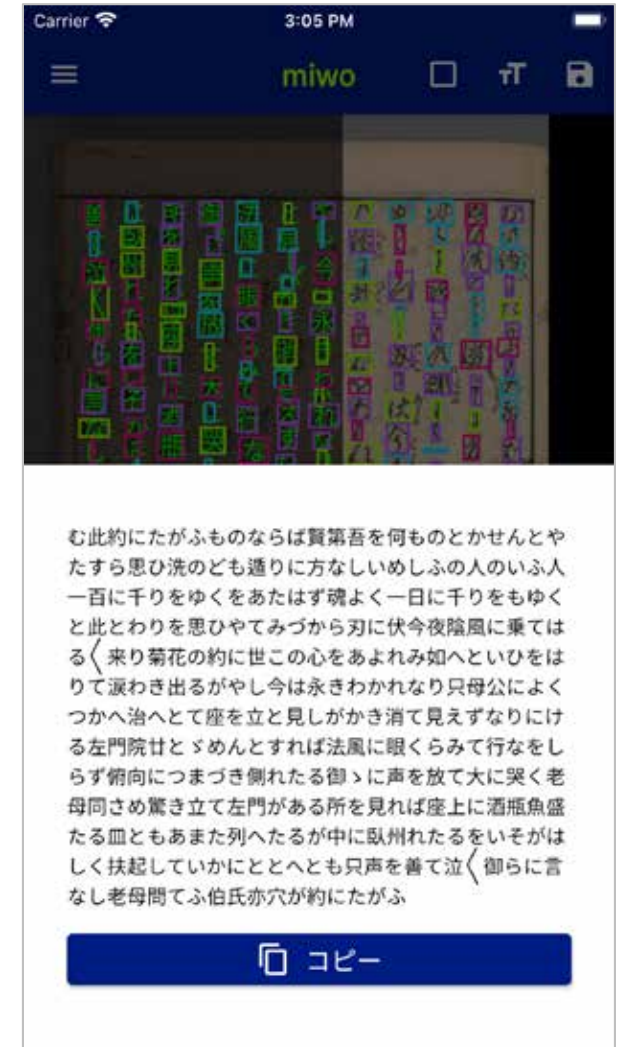
Show a recognition result in characters



Show a recognition result with bounding boxes



Modify the error with reference to root characters.



Generate the text output from the recognition result