



Architecture of Japan Search

Design Principles, Technologies, and the Future of National Platform for Digital Resources

Takanori Kawashima
National Diet Library, Japan

Agenda

1. Overview of Japan Search
2. Requirements and design principles
3. Overall architecture
4. Details of components
 1. Metadata collecting
 2. Metadata search
 3. Image search
 4. Galleries and workspaces
 5. User interface
5. Future of national platform for digital resources

About myself

- Takanori Kawashima, PhD
 - Planning division at National Diet Library
 - 10 years of experience in library systems and digitization.
 - Also working at the Research and Development for Next-Generation Systems Office for 8 years.
 - Part-time lecturer at Tokyo University
- Research Interest
 - Text analysis for humanities
 - Tool development for huminites
 - Japanese local ordinance



The National Diet Library (NDL) is the parliamentary library of the National Diet of Japan and was founded in 1948

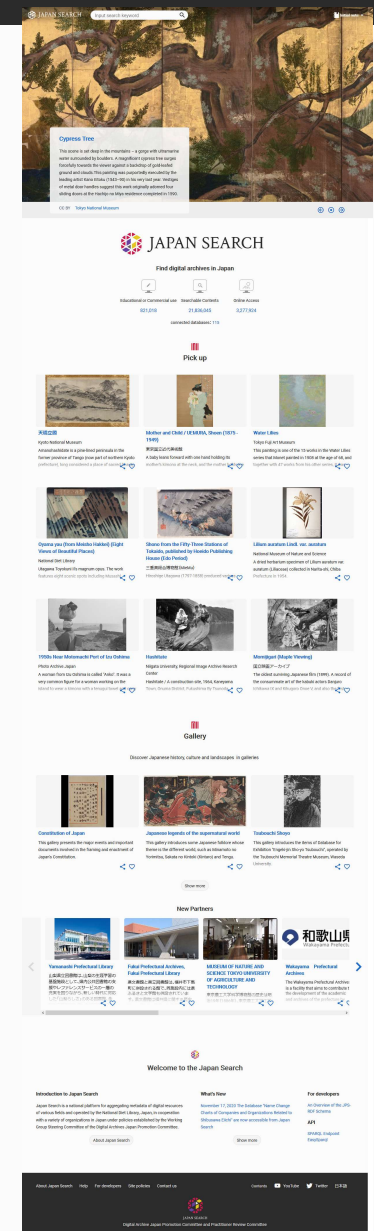


Overview of Japan Search

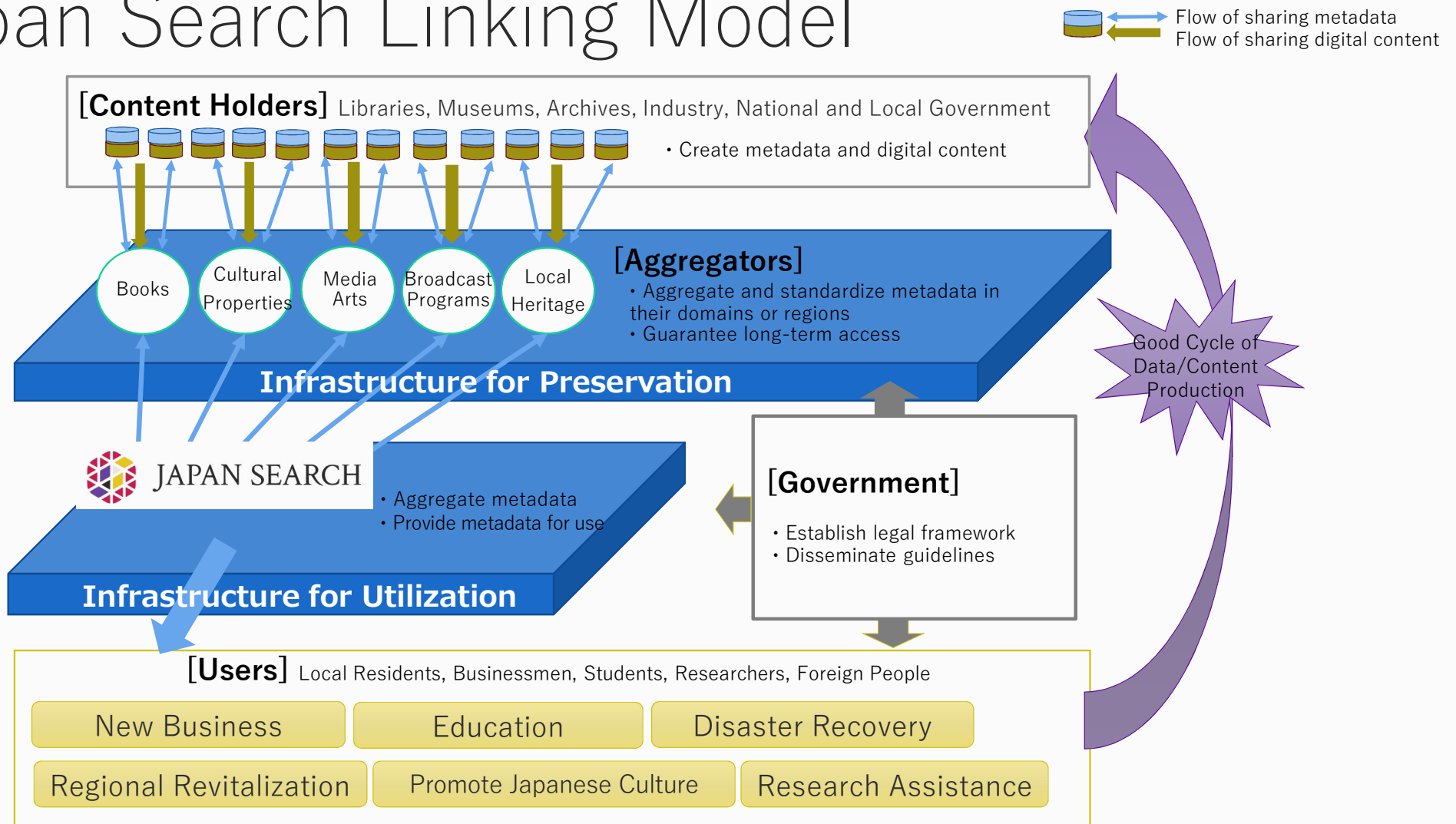
What is JAPAN SEARCH

- A search platform for metadata of Japanese content from a wide variety of fields
 - publications at libraries
 - cultural assets and art works at museums
 - documents at archives
 - academic resources in fields such as the humanities and natural history
 - broadcast programs and movies
- Our goals:
 - Enable one-stop discovery
 - Clarify secondary use conditions
 - Promote distribution of metadata→ To promote use and innovation of Japanese content

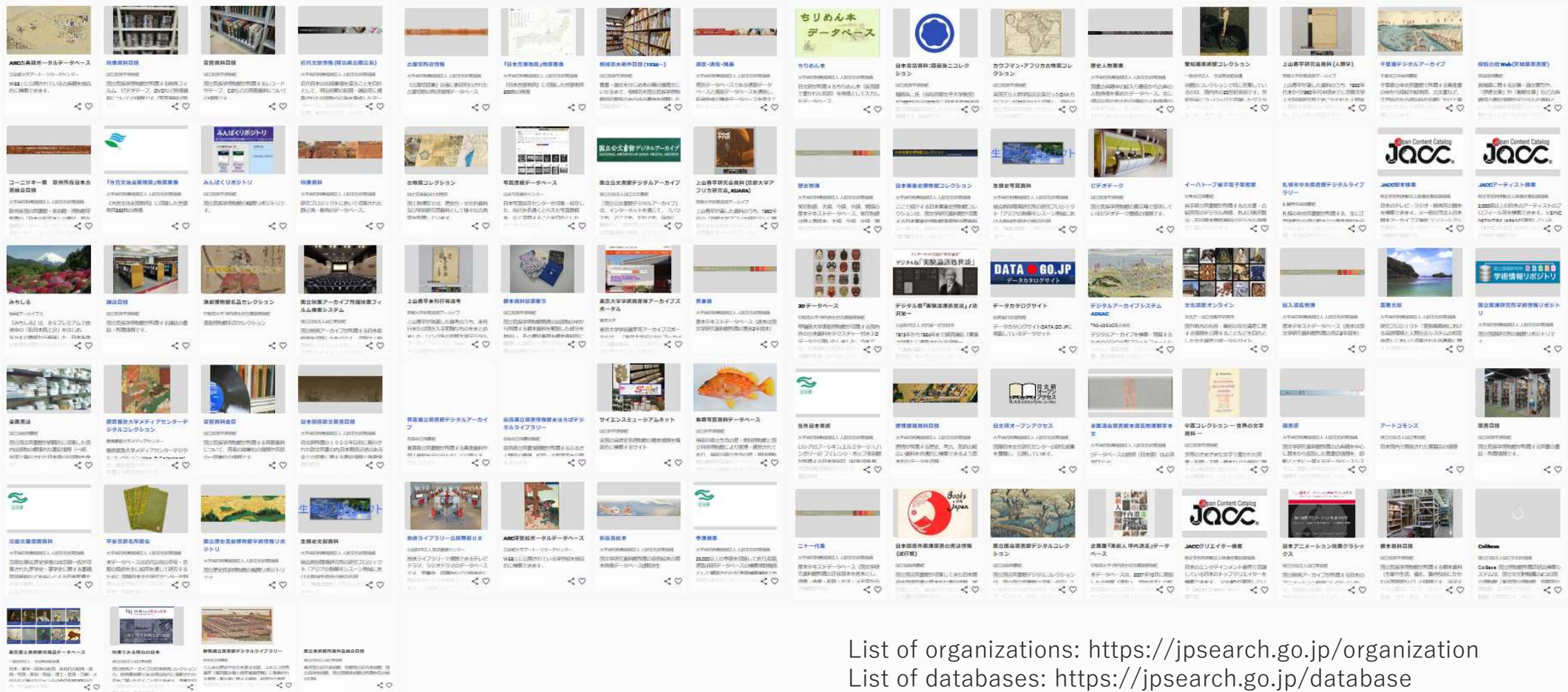
<https://jpsearch.go.jp>



Japan Search Linking Model



21.8 Million Metadata from 115 Databases



List of organizations: <https://jpsearch.go.jp/organization>
List of databases: <https://jpsearch.go.jp/database>



Requirements and design principles

Requirements

- Target users
 - Local Residents, Businessmen, Students, Researchers, Foreign People
- Usage
 - **Explore** : Users can reach contents they want to find.
 - **Utilize** : Users reached to the content can use it.



Requirements are ambiguous

Design principles

Flexible

- Flexible in format, search and gallery

Fast

- Faster is better

API

- Rich API

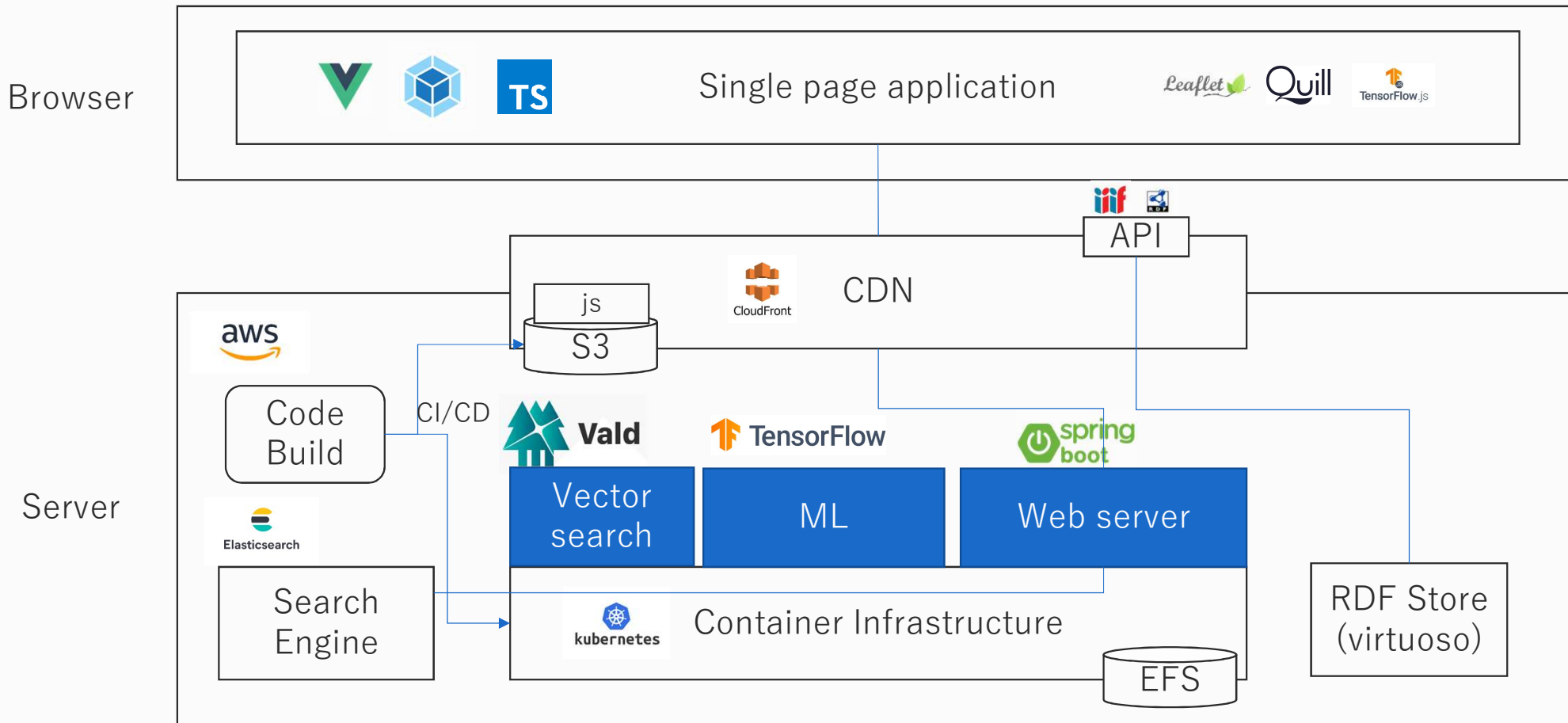
Core functionalities of Japan Search

- Collecting
 - Collect many format of metadata
 - Simple mapping functionality
- Explore
 - Search metadata
 - Able to create various search form
- Utilize
 - Create gallery page using metadata and contents
 - Provide SPARQL API, REST API



Overall architecture

System overview

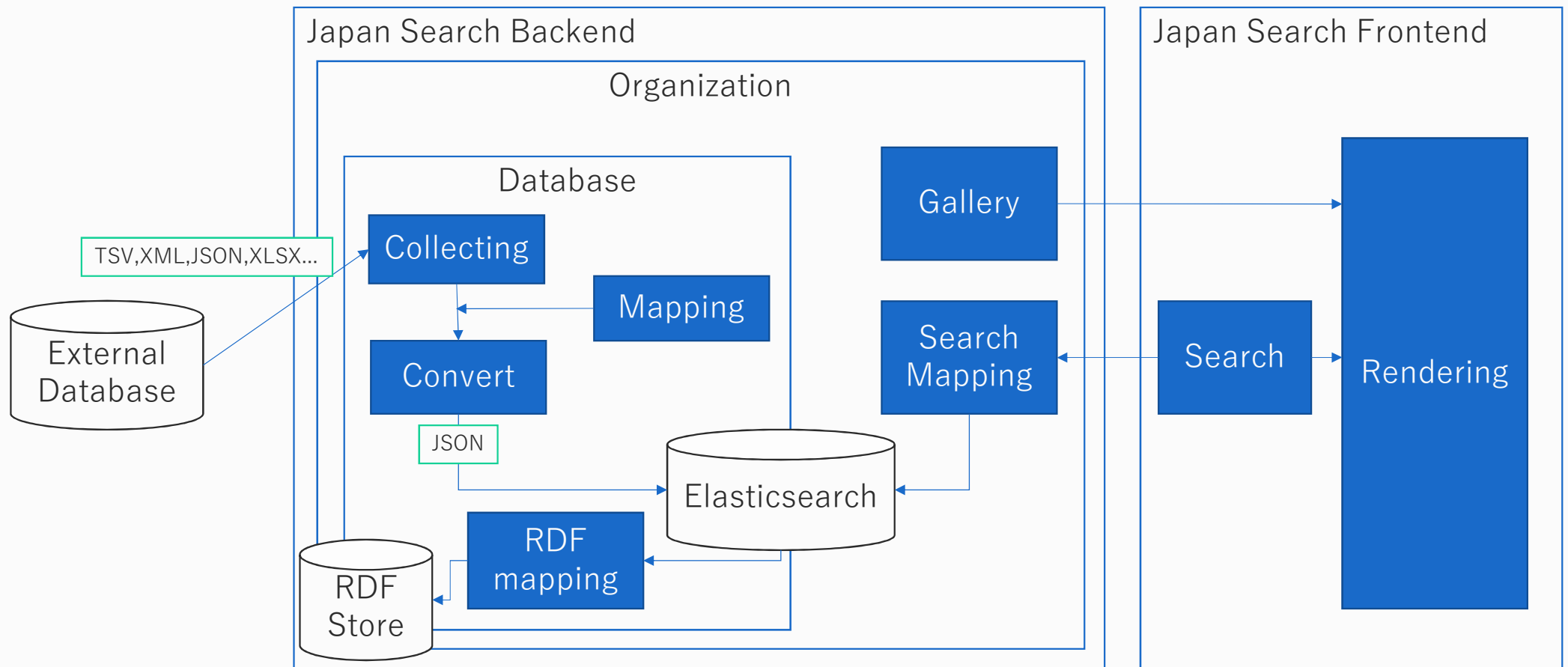


Key features

- Not-so-special-but-minimum-modern-architecture
 - Working on auto-scaling&auto-healing container infrastructure
 - Performance optimization with SPA and CDN
 - Some HTML headers are rendered on server side for OGP
 - bottle-neck is images from participating databases
 - CI/CD is implemented
 - About 50 releases in 3months.
 - No RDB
 - Most of data is stored in elasticsearch

Application overview

44k lines for backend(Java)
68k lines for
frontend(Typescript)





Details of components

Transition of Metadata in Japan Search

Data providers register metadata **in their original schema**

“Japan Search RDF Schema”

Data provider

Original metadata

名称	太刀 銘 三条 (名物 三日月宗近)
ID	159147
作者	宗近
分類	刀剣
時代世紀	平安時代・10世紀
所蔵者	東京国立博物館
URL	https://colbase.nich.go.jp/collectionItems/view/12f08f3c06a62af80737925634848303/179062

Registering/
Labeling



Providing common properties

Name/title	名称	太刀 銘 三条 (名物 三日月宗近)
Contributor	ID	159147
Time	作者	宗近
Provider	分類	刀剣
URL	時代世紀	平安時代・10世紀
	所蔵者	東京国立博物館
	URL	http://colbase.nich.go.jp/collectionItems/view/12f08f3c06a62af80737925634848303/179062

Info. of data-base

Database: ColBase
Type: Cultural Heritage
Content License: CC BY

Exhibiting

Searching

Mapping/
Converting

Japan Search

Generating RDF metadata

/data/cobas-159147	
rdf:type	type:刀剣
schema:name	太刀 銘 三条 (名物 三日月宗近)
schema:creator	chname:三条宗近
schema:dateCreated	0900-0999
jps:temporal/jps:era	time:平安時代
jps:accessInfo/schema:provider	chname:東京国立博物館
jps:sourceInfo/schema:provider	chname:ColBase
jps:sourceInfo/schema:url	https://colbase.nich.go.jp/collectionItems/view/12f08f3c06a62af80737925634848303/179062

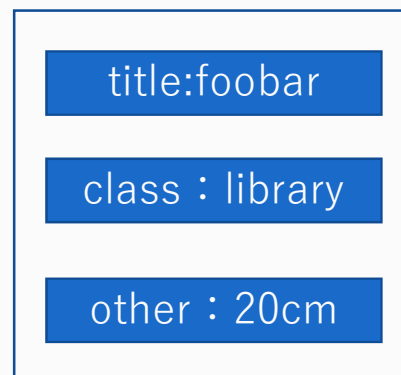
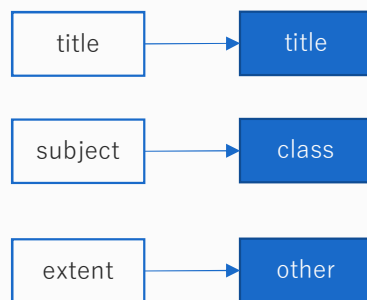
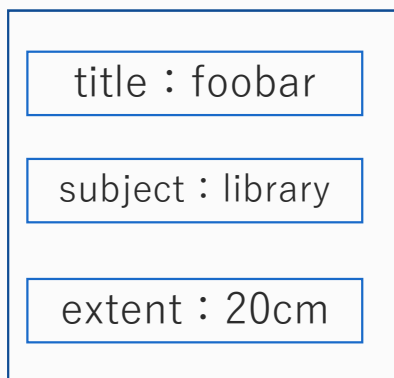
Original metadata is maintained

名称	太刀 銘 三条 (名物 三日月宗近)
ID	159147
作者	宗近
分類	刀剣
時代世紀	平安時代・10世紀
所蔵者	東京国立博物館
URL	http://colbase.nich.go.jp/collectionItems/view/12f08f3c06a62af80737925634848303/179062

Advanced search with SPARQL

Metadata Mapping

Ordinary metadata
search system



title	foobar
class	library
other	20cm

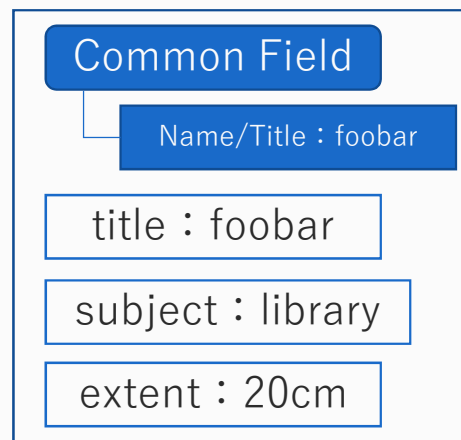
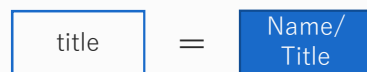
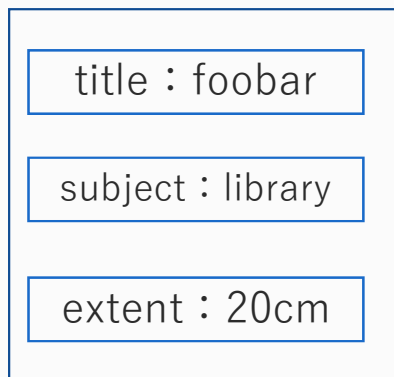
Original

Convert Rule

After Convert

Display

Japan Search



Common Field	
Name/Title foobar	
Original Field	
title	foobar
subject	library
extent	20cm

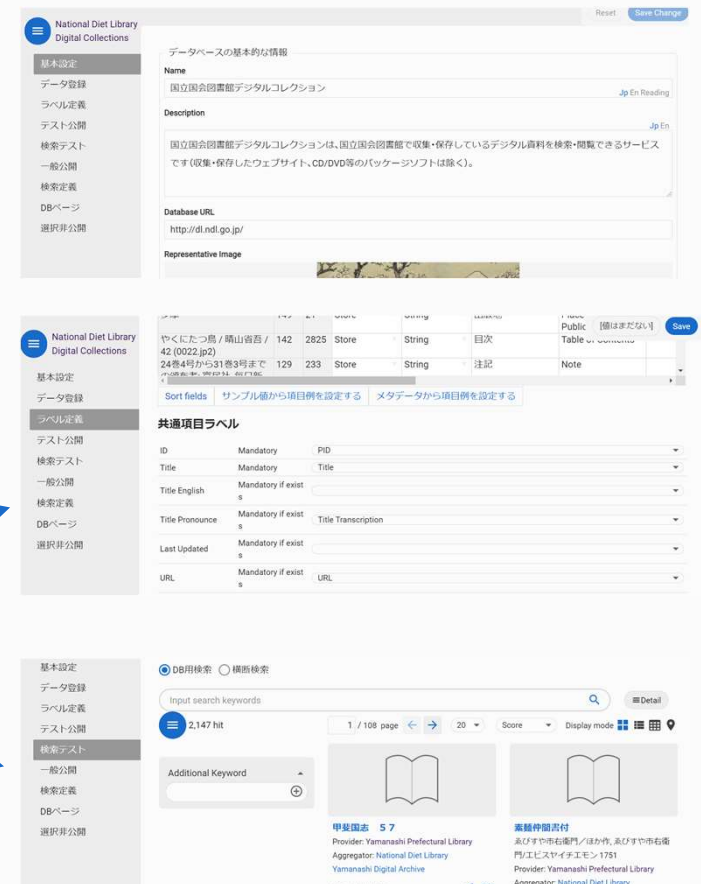
Metadata Mapping

- Retain all original metadata
 - No undistinguishable other field
- Minimum mapping required
 - Reduce the data provider's workload
 - Minimum mapping is ID, title, and URL to the original database
- Some fields can be come form database setting(e.g. license)
- RDF mapping is separated from basic mapping
 - The mapping is done by an experienced data engineer, but the process is optional(So, not every database has RDF).

Minimum 7 steps to connect a database

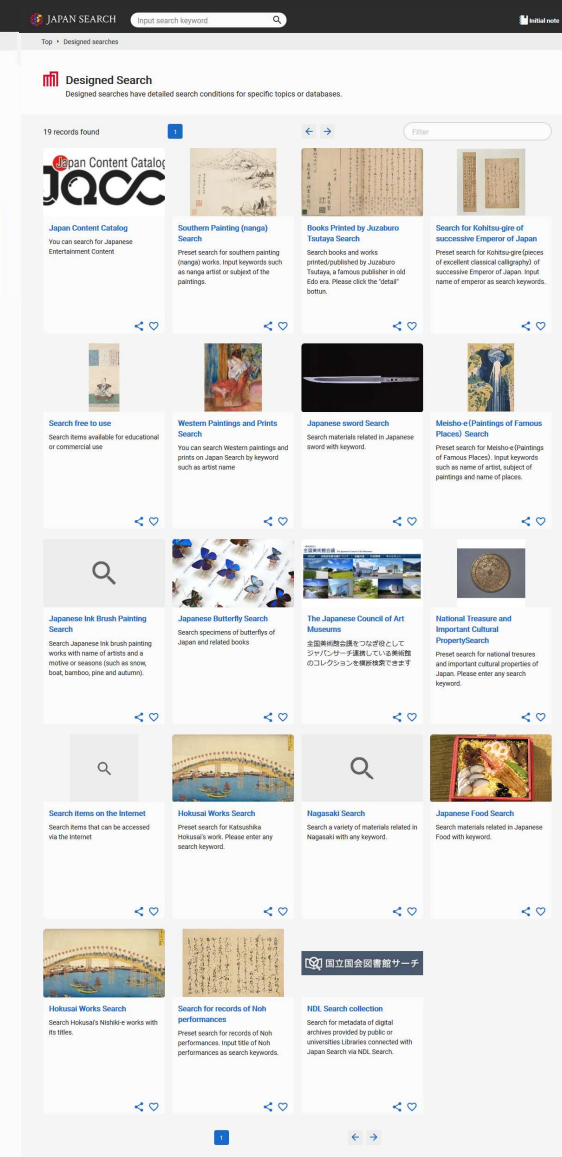
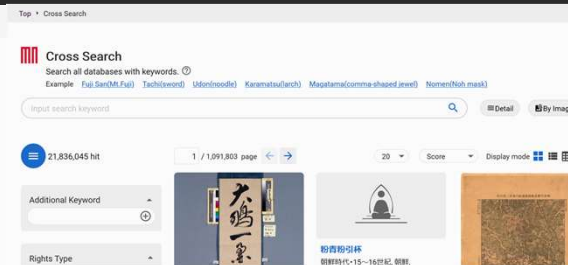
1. Register organization information
2. Create new database and input basic information
3. Extract metadata from original database with tsv.
4. Upload tsv or xlsx, with field names on first line
5. Designate id, title, url and other fields.
6. Test publish (Can test search)
7. Final publish

~30min, if everything goes well.

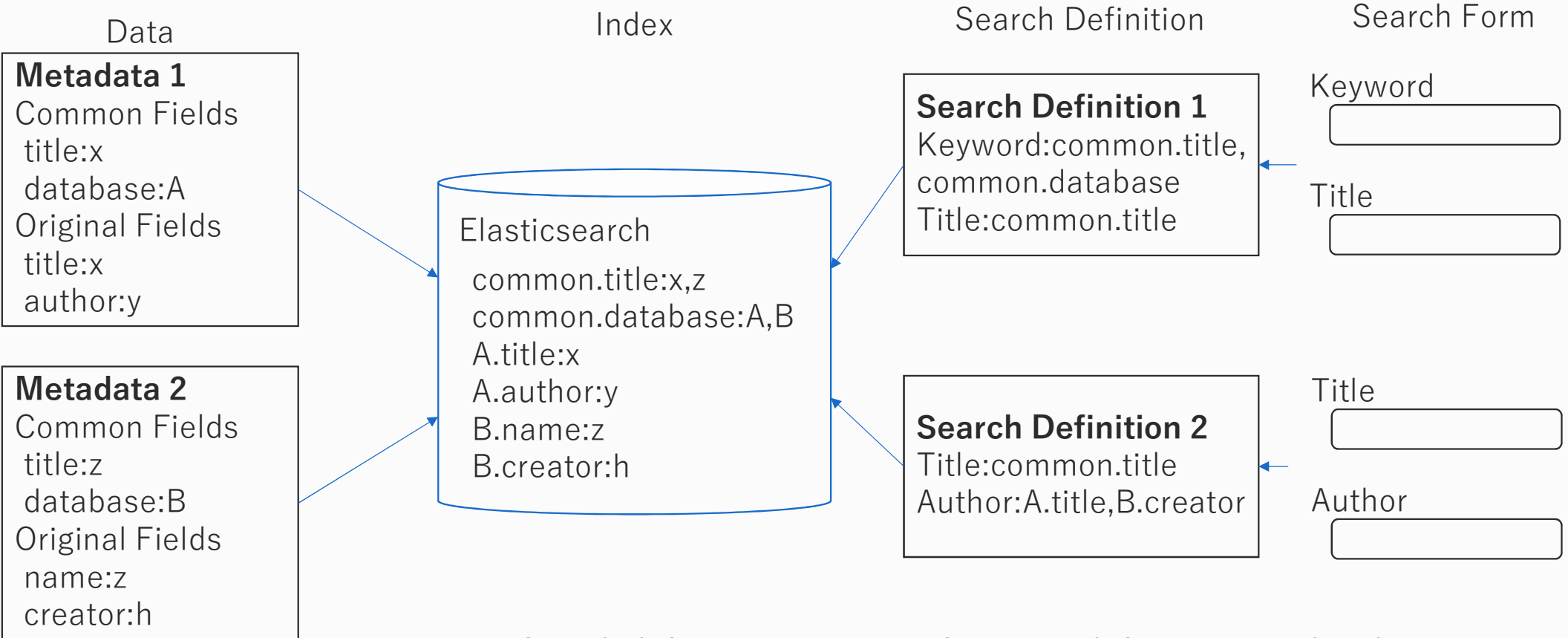


Metadata search

- Japan Search have multiple search forms.
 - Cross search
 - Search all fields(including original one) of all databases
 - Designed search
 - Search form for specific purpose
 - e.g. search Japanese Sword
 - Database search
 - Search form for specific database



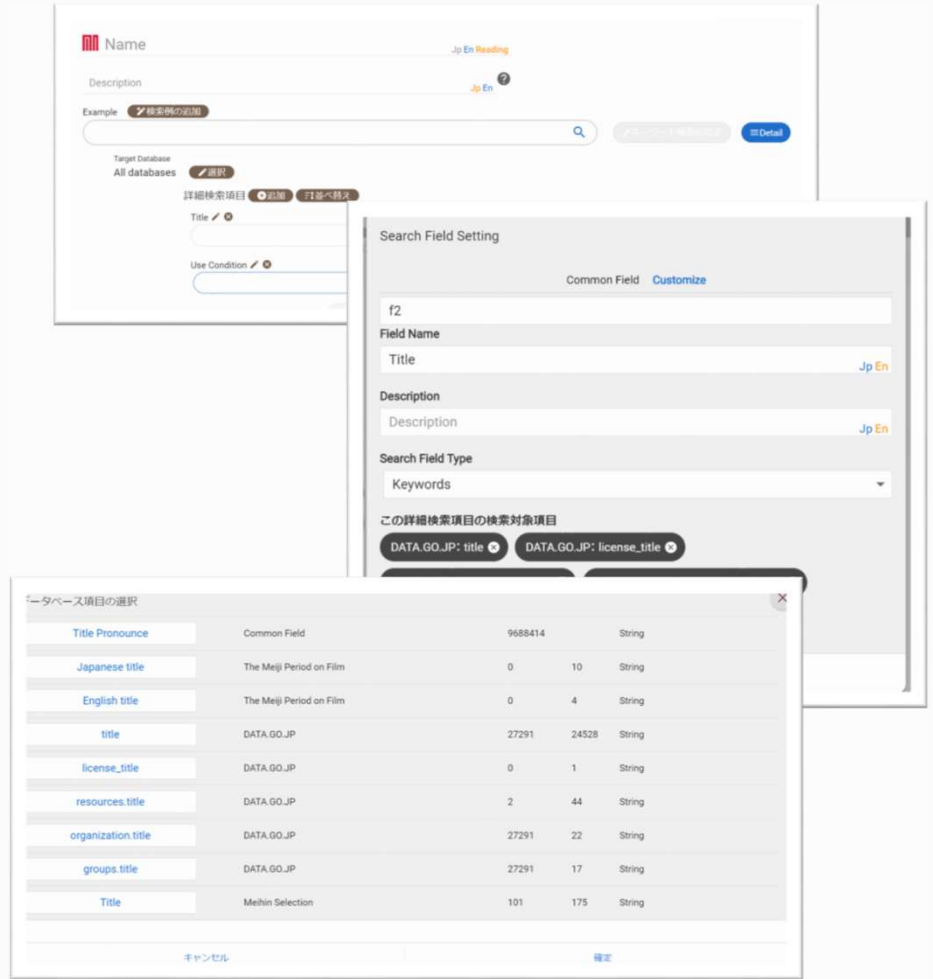
Metadata search: behind the scene



Search definition is mapping from search form to search index.

Metadata search: search designer

- Japan Search includes WYSWYG search designer.
- You can have
 - Simple keyword field which target specific fields
 - Multiple detailed search fields
 - Search restriction, which restricts search results with specific conditions (like database or use conditions)
- You can search fields of databases.

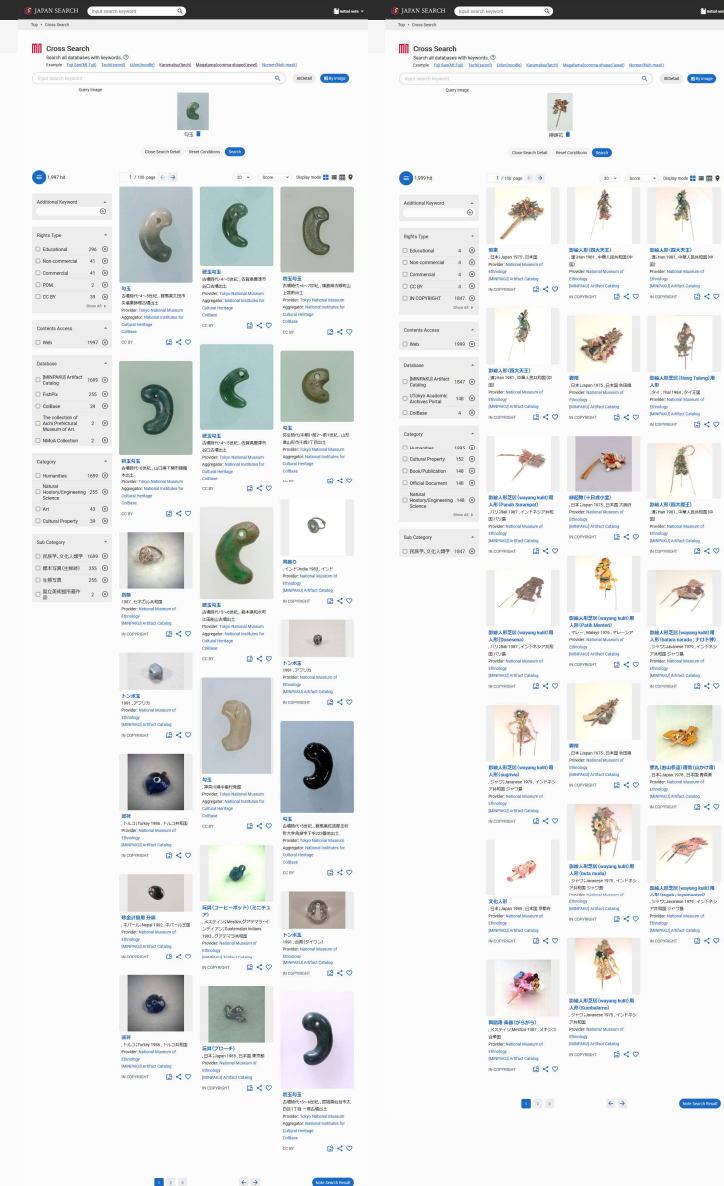


Points of metadata search

- Requirements for search is vary.
 - Metadata-level mapping restricts search possibilities.
 - Search-level mapping is superior in flexibility.
 - (But limited number of users for detailed search)
- Search designer can make searcher's knowledge sharable
 - Not teaching how to use search system, but creating search form easy to understand.
- Search is manipulatable object, not embedded to the system.
 - You can include designed search to gallery.

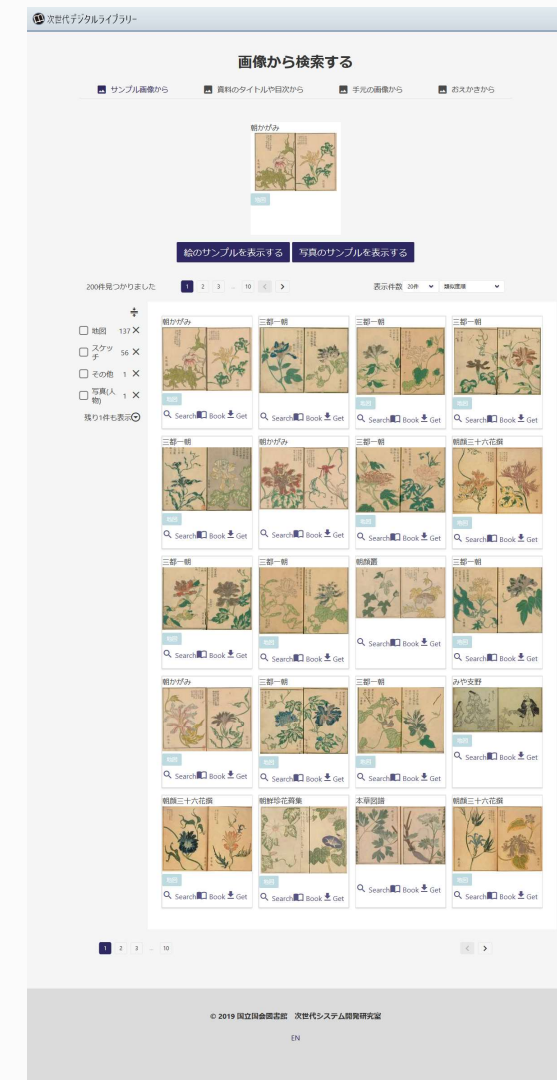
Image Search

- Select the image and find thumbnails which is similar to it.
- Using ImageNet and Mean Average Pooling layer from MobileNetV2 as feature vector.
- Using NGT/Vald for vector search engine.
- After retrieving similar image IDs, we search elasticsearch with them to enable faceting.



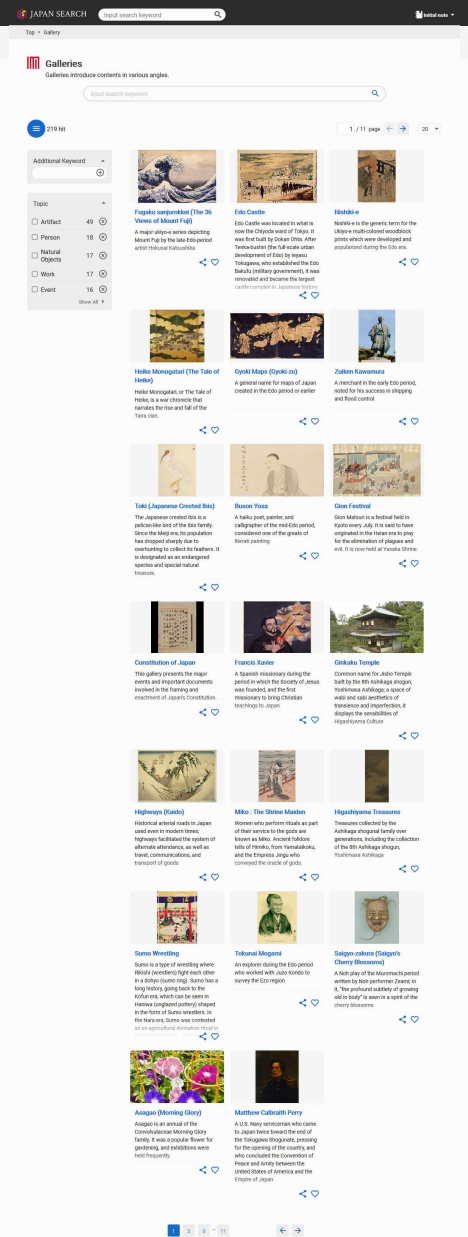
Next-gen Digital Library

- Originally, we implemented image search on books.
- The service can search images from 335,426 public-domain books.
- Search functionality is almost same, but in this case we extract images from books before indexing.
 - using ESPNetV2(Semantig segmentation model)



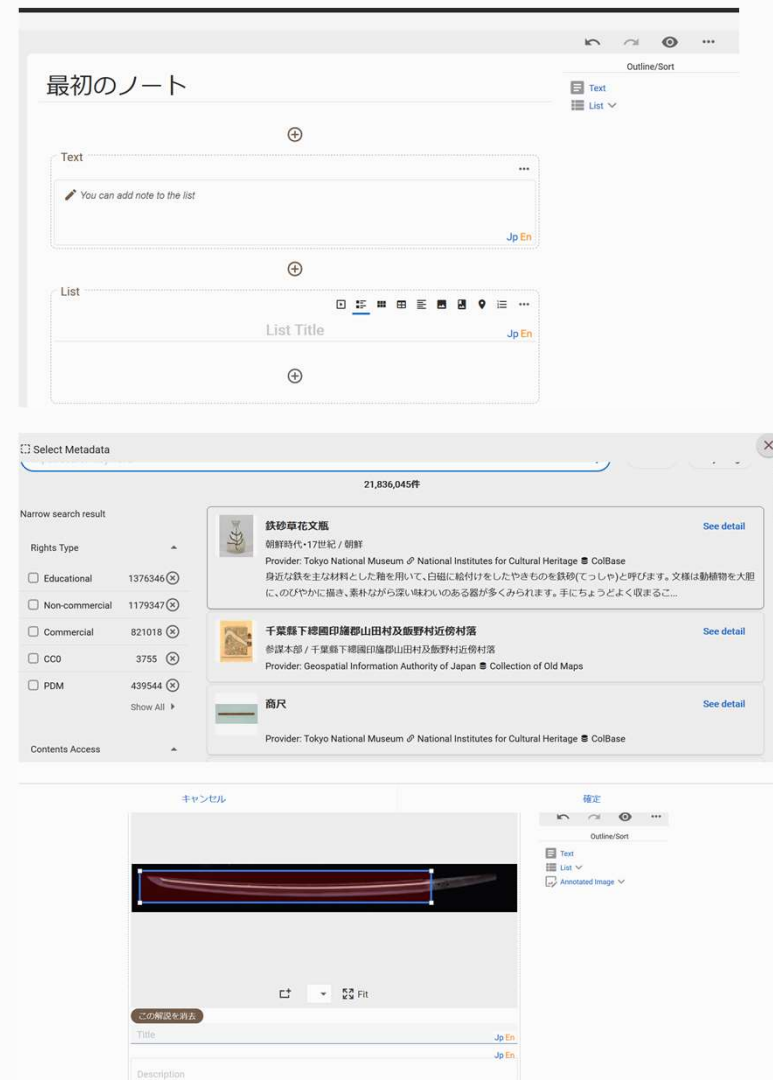
Galleries

- Galleries are manually created pages introducing contents of Japan Search.
 - We have currently 219 galleries
- Galleries behave has landing-page from search engine, and “Pre-defined search result” for beginners.



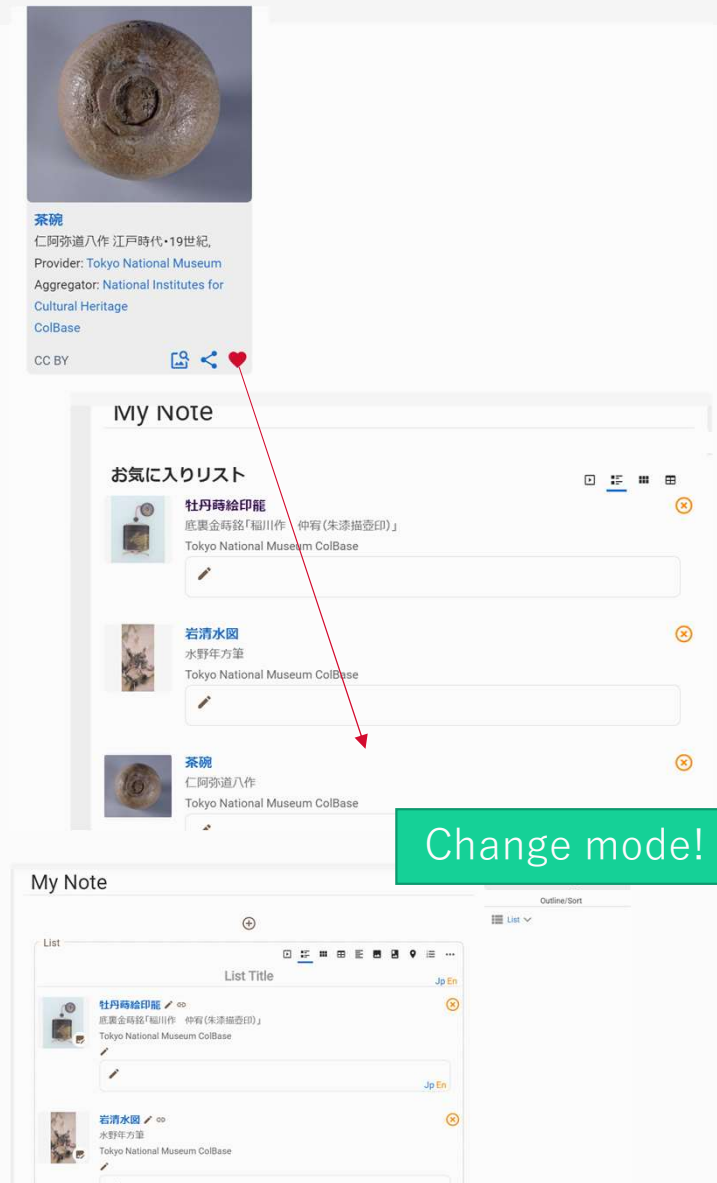
Gallery Editor

- Japan Search has built-in WYSYG gallery editor.
- Users can select pre-defined parts to create any page they like.
 - Parts available
 - Section, Text, Image, List of Metadata (Which can be shown as list, table, map, etc...), Search box, Annotated Image, Timeline
- Actually, non-gallery pages are created with this editor
 - Japan Search is Content Management System specialized for metadata in its core.

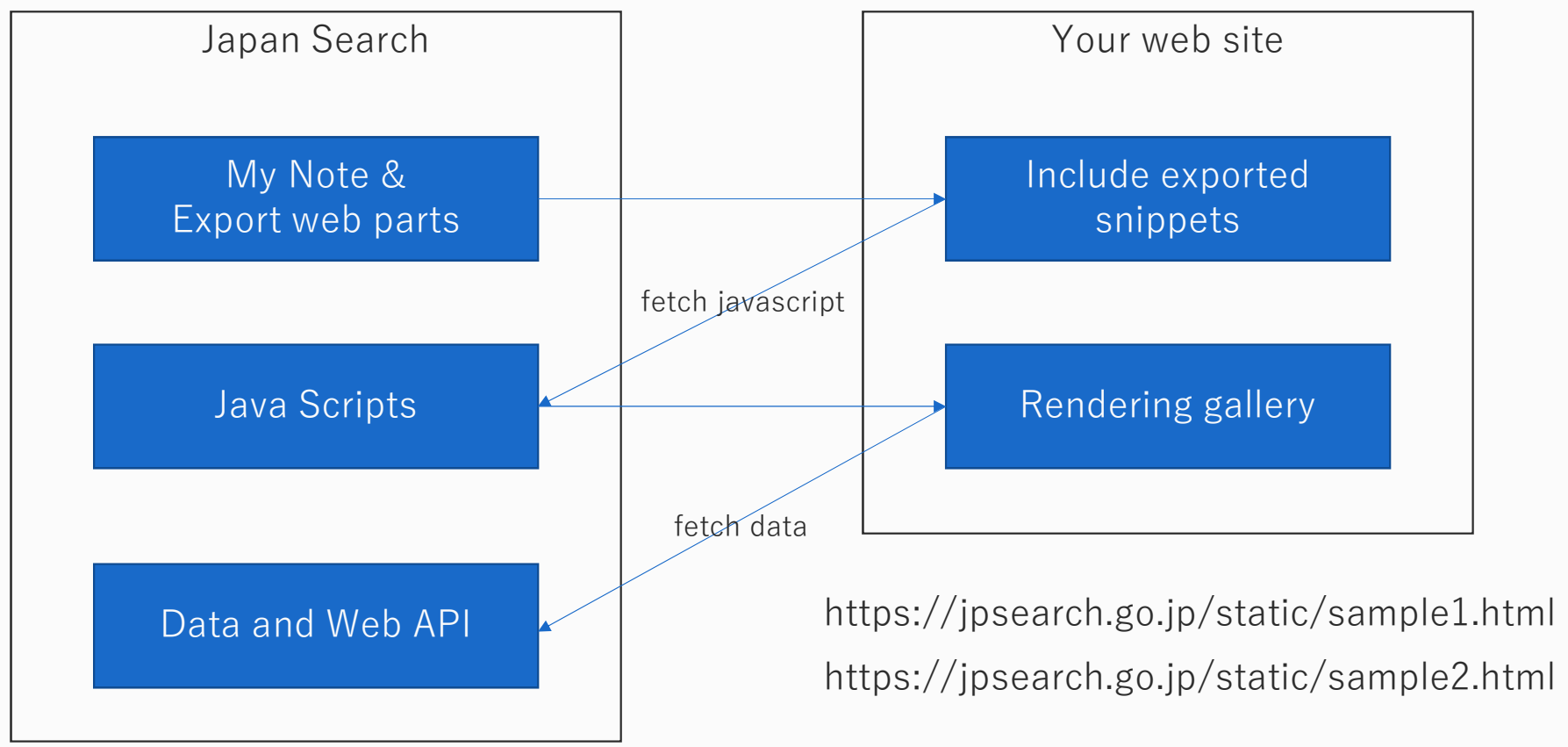


MyNote and Web Parts

- MyNote is a tool that enables users to save lists of their favorite items... at first glance.
- Actually, its simplified gallery editor. Users can create their own gallery with editor.
- My Note is saved on browser, but you can export them as JavaScript snippet so that you can include them with any website you own.
 - JPS is go.jp domain, so we must be very careful about user participation...



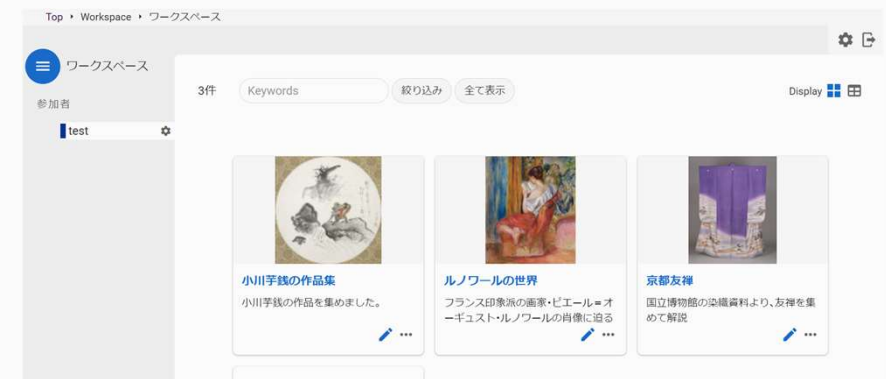
Create your own digital library site



With search designer and gallery editor, you can create any DL site based on data in Japan Search.

Workspace

- Workspace is collaborative gallery editor.
- Multiple users can edit gallery at the same time.
- Currently workspace is used for events, and teaching at schools.



UI Concepts

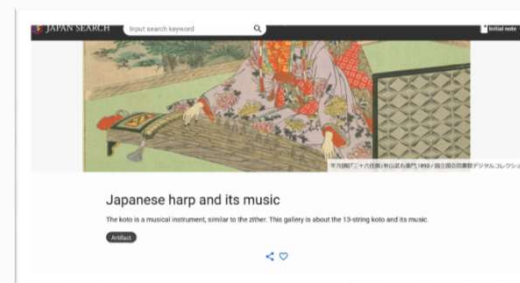
- No knowledge to use
- 4 primary pages has similar layouts.
 - Organization, Database, Metadata, Gallery
- They are interlinked so that user can “berry-picking” various contents on Japan Search.



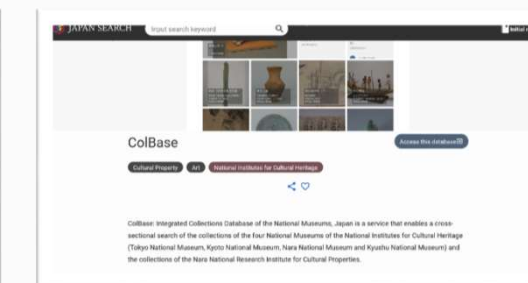
Metadata



Organization



Gallery



Database

CAN I USE IT?

- The most important UI for secondary use is license.
- Japan Search offer 15 types of license for database, such as CC-BY or PDM.
- We show not only license name, but it's detail, Reference display, and simplified chart.

CAN I USE IT?
Attribution

CC BY

EDUCATIONAL NON-COMMERCIAL COMMERCIAL

Specific Conditions
<https://colbase.nich.go.jp/pages/term?locale=en>

This license with the highest degree of freedom, with the main condition being to credit (name, title of the work, etc.) the original author; It also allows modification of the work, as well as secondary use for profit.

Details

Condition
<https://colbase.nich.go.jp/pages/term?locale=en>

Reference
歌川広重筆"東都遠足名所・羽根田辨天詣" owned by Tokyo National Museum, retrieved from ColBase (<https://jpsearch.go.jp/item/cobas-51531>)

Copy

Provider: Iokyo National Museum
Aggregator: National Institutes for Cultural Heritage
ColBase
CC BY

神奈川縣武蔵國橋樹郡神奈川青木町及羽澤村
參謀本部 神奈川縣武蔵國橋樹郡神奈川青木町及羽澤村
Provider: Geospatial Information Authority of Japan
Collection of Old Maps



Future of national platform for digital resources

What is digital archive?

- Digital library, especially DL of articles and books for researchers has established non-denial role.
 - They are replacement of existing paper (academic) library.
 - But digital archives, which focuses on cultural assets?
 - They are not replacement because museum can not switch to digital.
 - They are not replacement because user want to go to see cultural assets in person.
 - COVID-19 may impact, though.
- ⇒ Digital archives need additional cost & additional focus

Importance of national platform

- National platform is under control of a nation, compared to search engines.
 - But, apparently, we can not create superior search system...
- One-stop & permanent URL/API is fundamental in long run.
- Preservation is important. Many digital archives disappear when they run out of funds.
- Japan Search is generic digital archive platform in it's heart. An organization needs only TSV and image files to publish database on Japan Search.
 - But sometimes one want to create “yet another not so easy to use their own database”.

Chances

- Education
 - Not every classroom can go to museum. Instead, they can search and find things on Japan Search.
 - Japan Search does not contain “Fake” at least.
 - VR might have potency, if someone can pay costs.
- Designs
 - Japan Search is explicitly show license, so designers can easily use materials for their creation.

Future plans

- Familiarize the public with JPS
 - Empower users and the partner community
 - Promote use of JPS content in various fields
 - education, business, disaster recovery, regional revitalization, etc.
 - Raise awareness of open distribution
- Expand the scope of cooperation for local digital resources
 - local museum, local gallery, local government etc.
- Continuous development
 - It's just started.
 - The system/service must be updated to match the web. No more “scrap & build”
 - The world is going for digital even more.

For further inquiries:

<https://jpsearch.go.jp/contact>



JAPAN SEARCH OFFICIAL (@jpsearch_go)