

Extending IIIF to Curation and Timeline: Case Studies in Cultural Heritage, Humanities, and Natural Sciences



Asanobu KITAMOTO

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

National Institute of Informatics

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

@rois_codh

Working with Humanities and Natural Sciences

Pre-modern Japanese Text



『画本虫撰』日本古典籍データセット（国文研所蔵）

<http://codh.rois.ac.jp/iiif/iiif-curation-viewer/index.html?pages=200014778&pos=24>

Old Japanese books before 1860s are
massively being digitized since 2014.



Himawari-8 Satellite

- The BEST geostationary satellite on earth.
- Frequency: every 10 minutes for full disk and 2.5 minutes for Japan and severe weather.
- Resolution: maximum 22,000x22,000 pixels.

Himawari-8 RGB: 1100 JST, July 7, 2015



Himawari-8 RGB: 1150 JST, September 29, 2015

Image Datasets

Dataset of Pre-modern Japanese Text

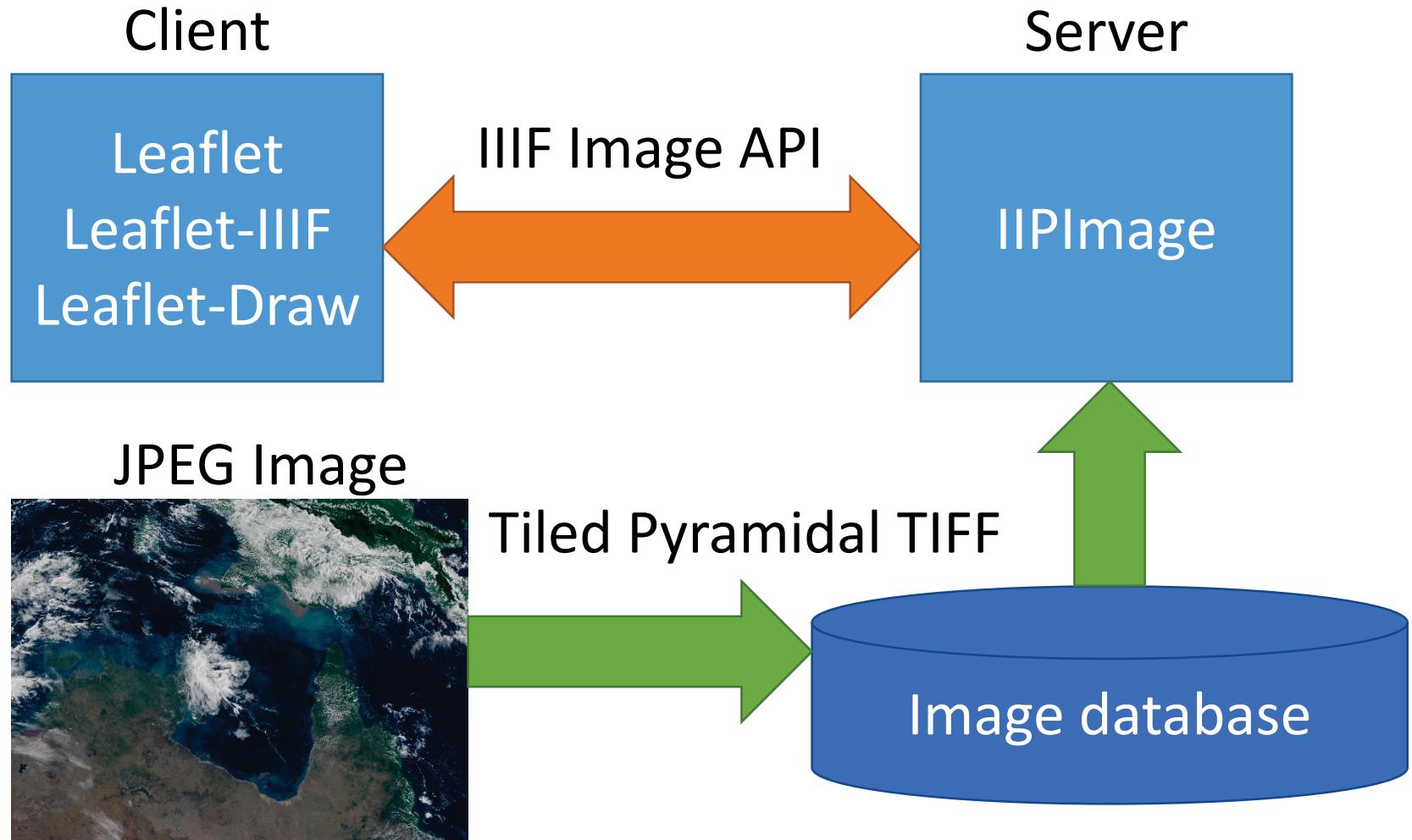
- <http://codh.rois.ac.jp/pmjt/>
- **701 manifests (158,553 images)**
- The final goal of **NIJL-NW project** is to provide 300,000 manifests with 68 million images (rough estimate).

Japanese Weather Satellite Images

- <http://agora.ex.nii.ac.jp/digital-typhoon/himawari-3g/clipping/>
- **2 manifests, with 100,000 and 400,000 images!**
- Older generation satellites (hour-level frequency) have about 300,000 images since 1979.

IIIF Adoption and Problems We Faced

Image Delivery System



Choice of Tools

IIPImage for the server side

- **Pros:** very fast and a large cache can be used.
- **Cons:** difficult to install and setup.

Leaflet IIIF for the client side

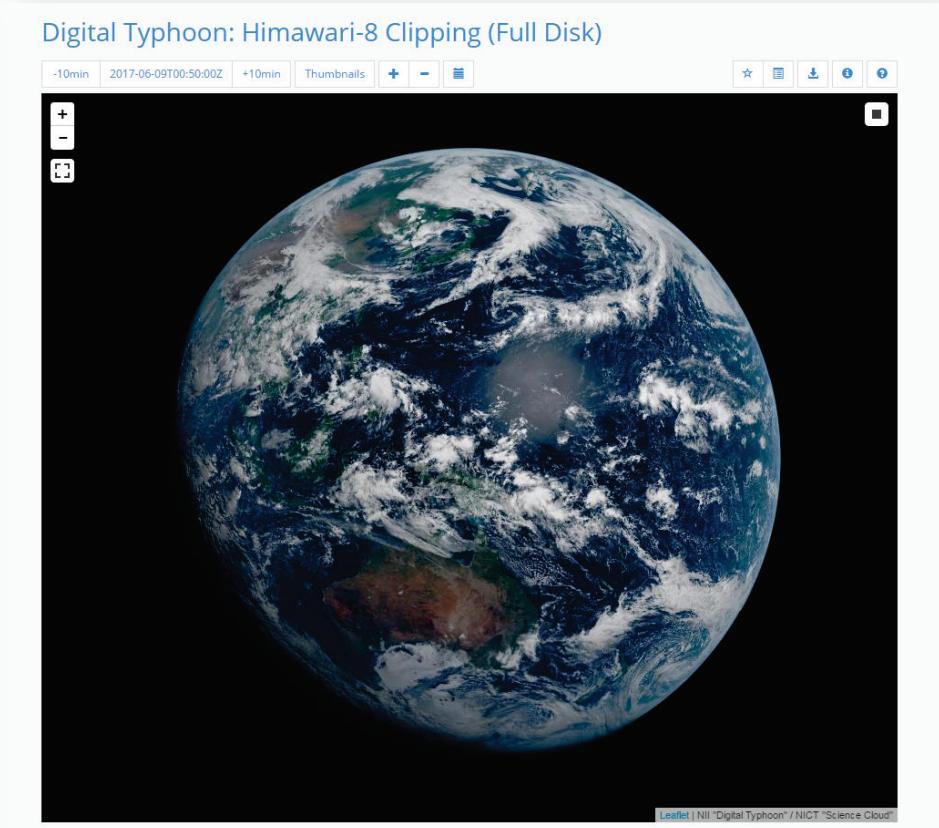
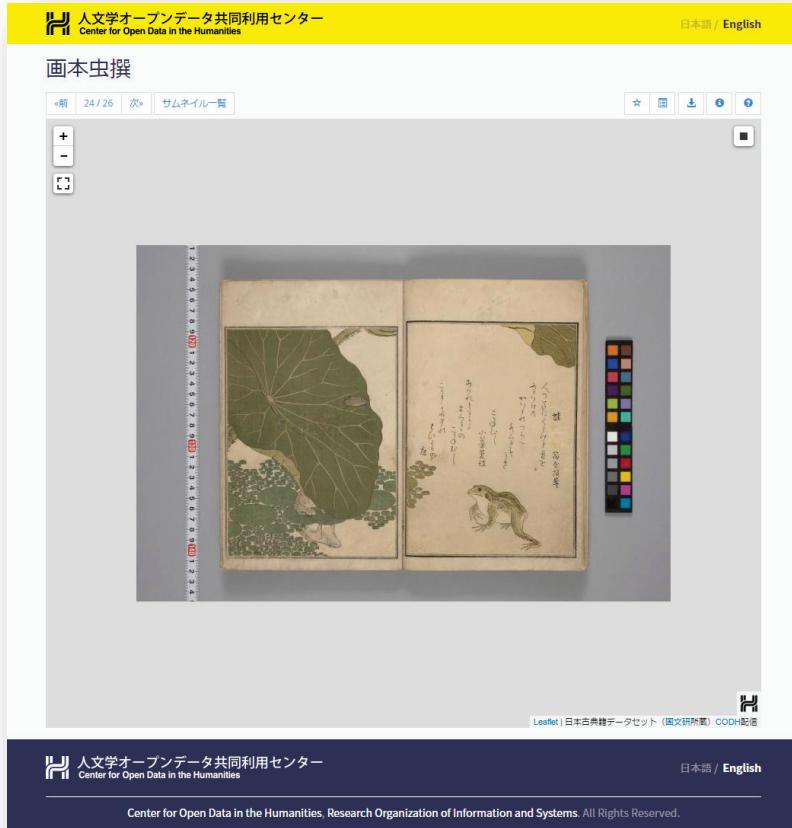
- **Pros:** Leaflet has an active community, and **plenty of plugins** have been developed on top of it.
- **Cons:** **No support for presentation API**. But this also means maximum freedom for development.

Issues of Sequence

1. **Manifest** refers to a **natural sequence** derived from a physical book.
2. **Curation** refers to a **logical sequence** derived from the intention of a curator.
3. **Timeline** describes an **absolute sequence** derived from a conceptual model of time.
4. **The latter two cannot be realized within the current IIIF Presentation API.**

IIIF Curation Viewer (for Timeline)

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



Extension of Presentation API

Cursor API

- Use case: **time-series satellite images.**
- Make an on-demand request to a part of sequence.
- **How can we select a part of sequence in a single manifest?**

Curation API

- Use case: **theme-based exhibition or curation.**
- Make a list of (a part of) interesting canvases with metadata.
- **How can we collect a part of canvases across multiple manifests?**

Cursor API

Granularity of Manifests

- How to deal with a huge manifest of 338MB and 404,172 canvases?
- A simple solution is to divide a long sequence by calendar such as day and year.
- In Mother Nature, calendar-based division is artificial and makes unnecessary boundaries.
- In books, the granularity is intuitive because they are intentionally edited materials.

Cursor for On-Demand Access

- We decided to deal with the long sequence problem without dividing manifests.
- A solution: **cursor for on-demand access to a part of sequence** (e.g. Twitter API).
- **Cursor API** receives the start index, and returns a sequence for a specified range.
- **Time-based cursor** is based on “Unix time” to have an absolute coordinate in 1D.

Example of Time-Based Cursor

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

```
"viewingHint": "time",
"cursors": [
{
"@id": "http://agora.ex.nii.ac.jp/digital-typhoon/service/iiif/Hsfd/cursor",
"service": {
"@context": "http://codh.rois.ac.jp/iiif/cursor/1/context.json",
"@id": "http://agora.ex.nii.ac.jp/digital-typhoon/service/iiif/Hsfd/cursor",
"profile": "http://codh.rois.ac.jp/iiif/cursor/1/cursor"
},
"first": 1436234400,
"last": 1498867200,
"default": 1498867200,
"step": 600 }
```

Request and Response

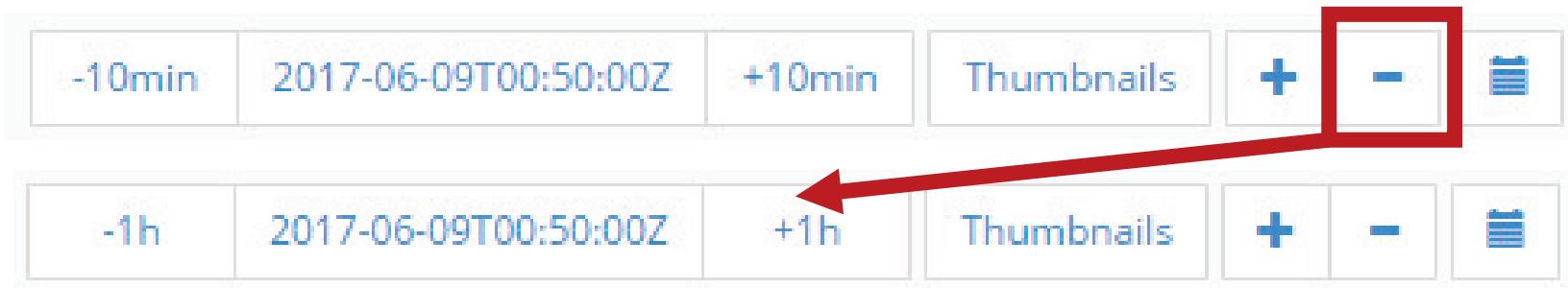
Request for Cursor API

<http://agora.ex.nii.ac.jp/digital-typhoon/service/iiif/Hsfd/cursor?cursorIndex=1493596800>

Response from Cursor API

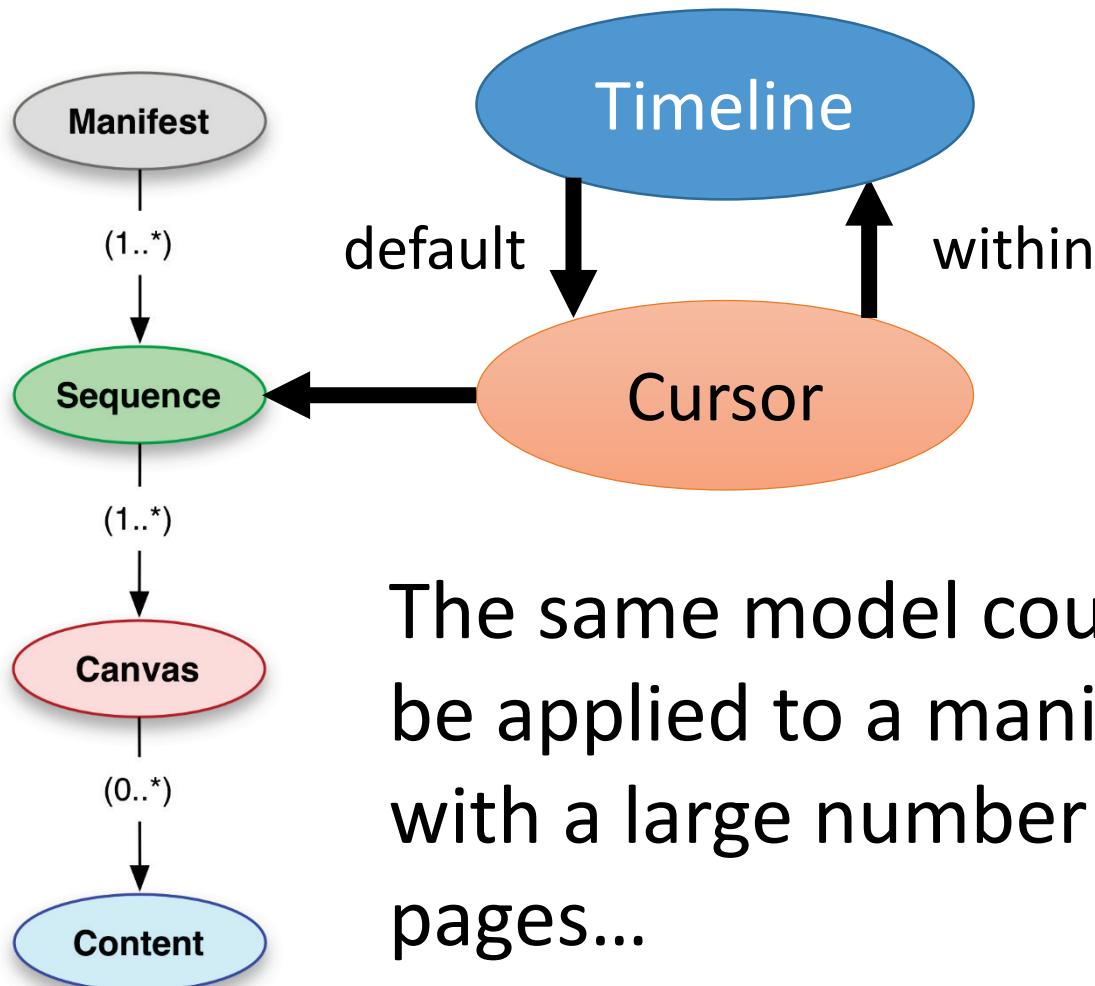
```
"within": "http://codh.rois.ac.jp/iiif/iiif-curation-viewer/sample/himawari/timeline.json",
"sequence": {
  "@id": "http://agora.ex.nii.ac.jp/digital-typhoon/service/iiif/Hsfd/cursor?cursorIndex=1493596800",
  "@type": "sc:Sequence",
  "label": "2017-05-01T00:00:00Z ~ 2017-05-01T23:50:00Z",
  "canvases": [ {
    "@id": "http://agora.ex.nii.ac.jp/digital-typhoon/iiif/2017/05/01/20170501000000", "@type": "codh:Canvas",
    "label": "2017-05-01T00:00:00Z",
    "height": 11000,
    "width": 11000,
    "cursorIndex": 1493596800,
    "images": [
```

Temporal Zoom



- Like spatial zoom, the viewer offers **temporal zoom buttons** to change temporal resolution.
- For cursor, it means “how many canvases” one request skips.
- For time-based cursor, it means “how much time” one request skips based on “step.”

Extension to Presentation API



The same model could
be applied to a manifest
with a large number of
pages...

Thickest Books

4,032 pages

The screenshot shows a web page from the Guinness World Records website. At the top, there are social media sharing icons. Below them, the title "Thickest book published" is displayed. To the left of the text, there is a grid of nine small images related to the record, including a person reading a very thick book, a dog on a surfboard, and the Guinness World Records logo. On the right side of the page, there is detailed information about the record:

- Who:** HARPERCOLLINS
- What:** 322 MILLIMETRE(S)
- Where:** UNITED KINGDOM, LONDON, FOYLES BOOK SHOP
- When:** 20 MAY 2009

A large paragraph below the details describes the book's dimensions and contents. At the bottom of the page, there are two buttons: "APPLY NOW" and a trophy icon.

The thickest published book measures 322 mm (12.67 in) in width and was unveiled by HarperCollins in London, UK, on 20 May 2009. All Agatha Christie's Miss Marple stories - 12 novels and 20 short stories - are collected and published in this volume. The book contains 4,032 pages and weighs 8.04 kg. There are 68 crimes committed; 11 philandering lovers; 68 secrets and lies; 22 false accusations; 59 red-herrings and 21 romances. In all, 43 murders are solved: 12 poisonings; 6 strangulations; 2 drownings; 2 stabblings; a burning; one blow to the head; one death by an arrow and 2 people pushed. 143 cups of tea are drunk in the massive volume, there are 66 maids and 47 garments are knitted. 500 copies of this limited edition record-breaking book will be produced and they will retail at £1,000 each.

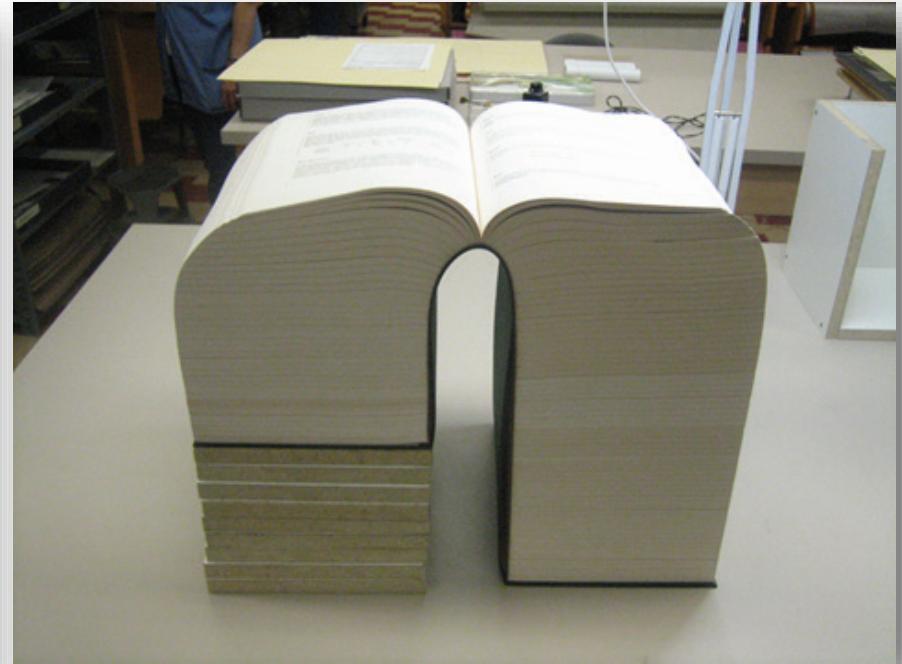
All records listed on our website are current and up-to-date. For a full list of record titles, please use our Record Application Search. (You will need to register / login for access)

Comments below may relate to previous holders of this record.

<http://www.guinnessworldrecords.com/world-records/thickest-book-published>

2017/6/9

10,119 pages



10,000 page book bound at Conservation Lab:

<https://blog.lib.uiowa.edu/preservation/2011/01/07/10000-page-book-bound-at-conservation-lab/>

2017 IIIF Conference

20

Curation API

Manifest vs. Curation

1. **Manifest** refers to a **natural sequence** derived from physical books.
 2. **Curation** refers to a **logical sequence** derived from the intention of a curator.
- Manifest is an authoritative information from a data provider.
 - How to make a user-driven sequence without copying or modifying manifests?

Selection of Objects

From IIIF Presentation API 2.1:

The following are **not** within scope:

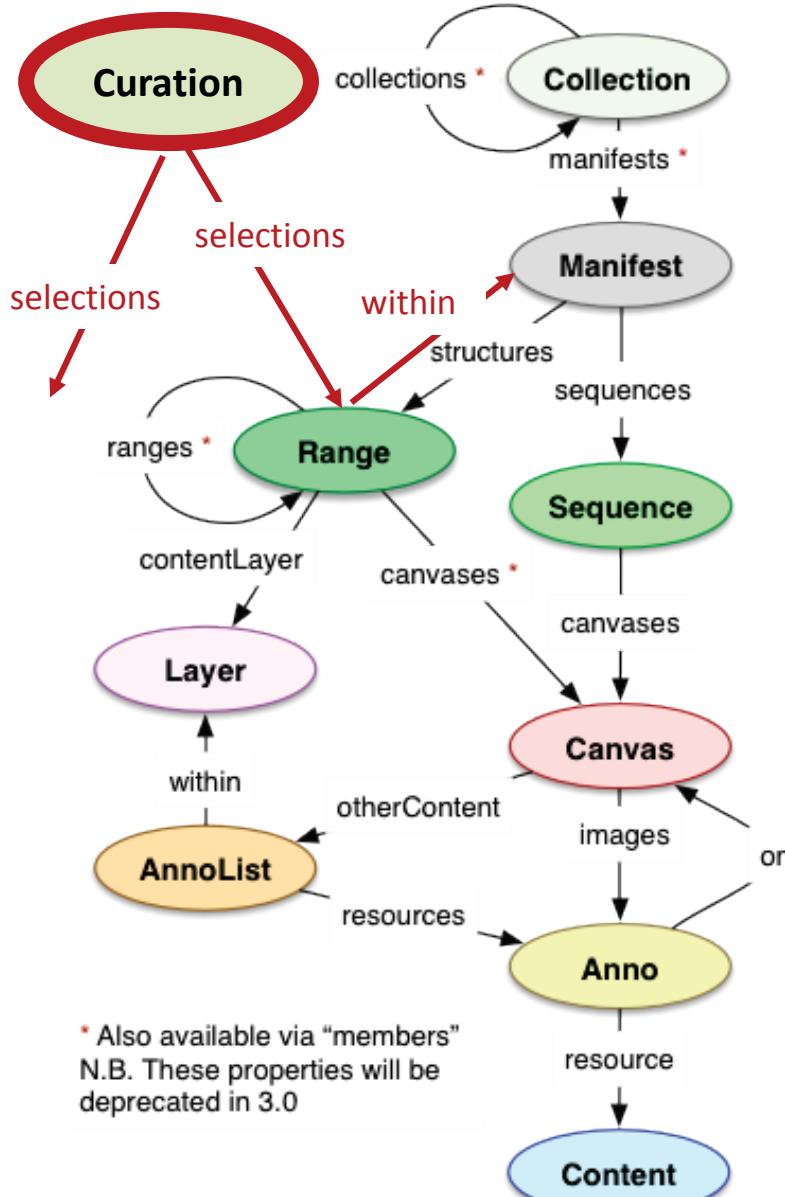
The discovery or selection of interesting digitized objects is not directly supported; however hooks to reference further resources are available.

- “Selection of interesting objects” is exactly what we want to do, **but not within scope**.
- Why not **extending Presentation API** to satisfy our needs based on the concept of Curation?

Selection of (a part of) Canvases

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

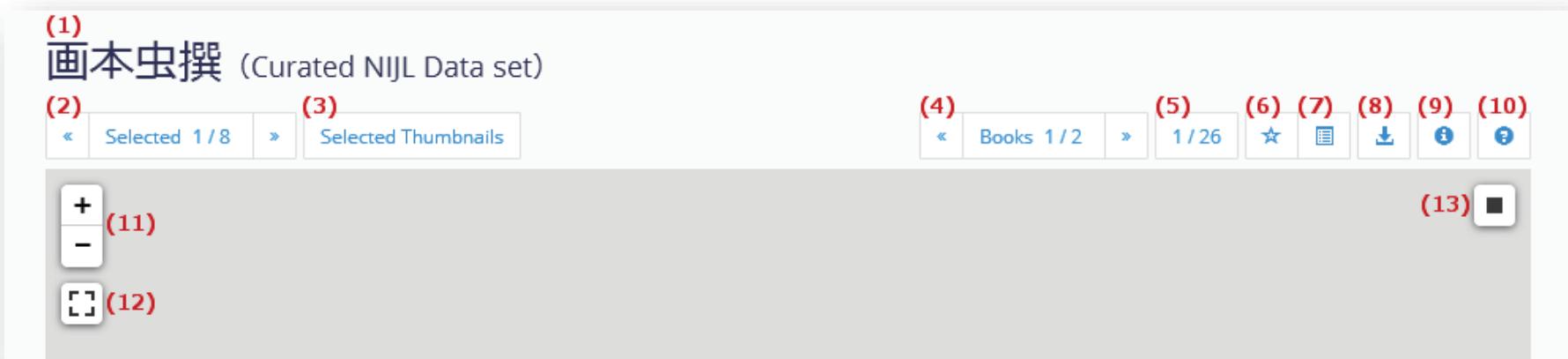
```
"selections": [ {  
    "@id": "http://codh.rois.ac.jp/pmjt/book/200014778/range/r1",  
    "@type": "sc:Range",  
    "label": "Curated contents from 『画本虫撰』",  
    "canvases": [  
        "http://codh.rois.ac.jp/pmjt/iiif/200014778/canvas/00000",  
        "http://codh.rois.ac.jp/pmjt/iiif/200014778/canvas/00023",  
        "http://codh.rois.ac.jp/pmjt/iiif/200014778/canvas/00023#xywh=3435,2487,840,750"],  
        "within": "http://codh.rois.ac.jp/pmjt/book/200014778/manifest.json" }, {  
            "@id": "http://codh.rois.ac.jp/pmjt/book/200003067/range/r1",  
            "@type": "sc:Range",  
            "label": "Curated contents from 『唐糸草紙』",  
            "members": [  
                { "@id": "http://codh.rois.ac.jp/pmjt/iiif/200003067/canvas/00000", "@type": "sc:Canvas", "label": "p.1"},  
                { "@id": "http://codh.rois.ac.jp/pmjt/iiif/200003067/canvas/00010", "@type": "sc:Canvas", "label": "p.11"},  
                { "@id": "http://codh.rois.ac.jp/pmjt/iiif/200003067/canvas/00010#xywh=2850,1000,730,680",  
                    "@type": "sc:Canvas", "label": "p.11" },  
                "within": { "@id": "http://codh.rois.ac.jp/pmjt/book/200003067/manifest.json", "@type": "sc:Manifest",  
                    "label": "唐糸草紙" } } ]
```



Extension to Presentation API

- **Curation** is another upper element of Range.
- **Range** points to **Manifest** using “within” property.
- **Curation** is an array of ranges across multiple manifests.

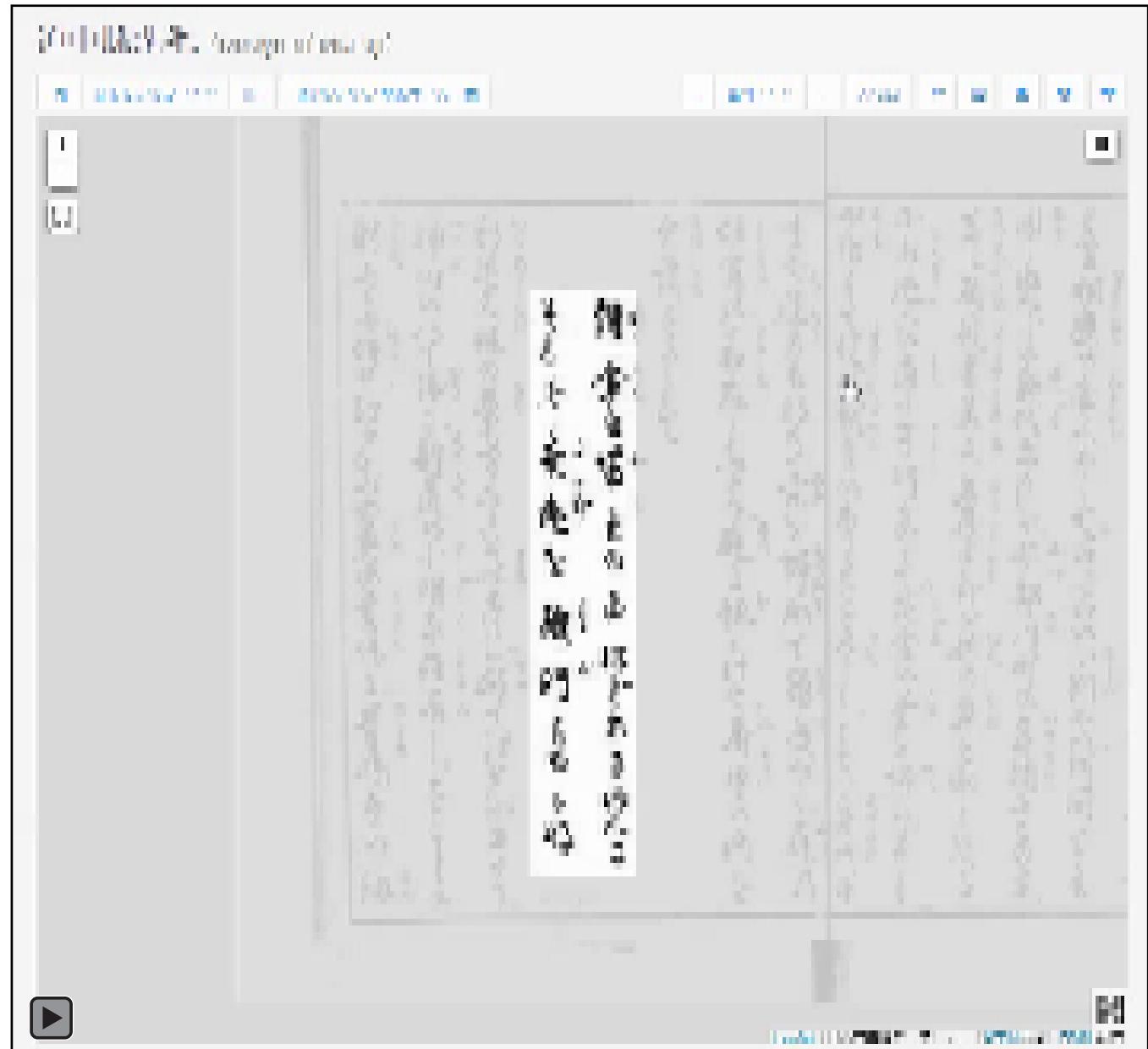
User Interface of Curation Viewer



- ‘☆’ (6) is a “favorite” button to collect interesting objects.
- ‘■’ (13) is a tool to draw a rectangle on a canvas to select the region of interest.
- The viewer should deal with two sequences.

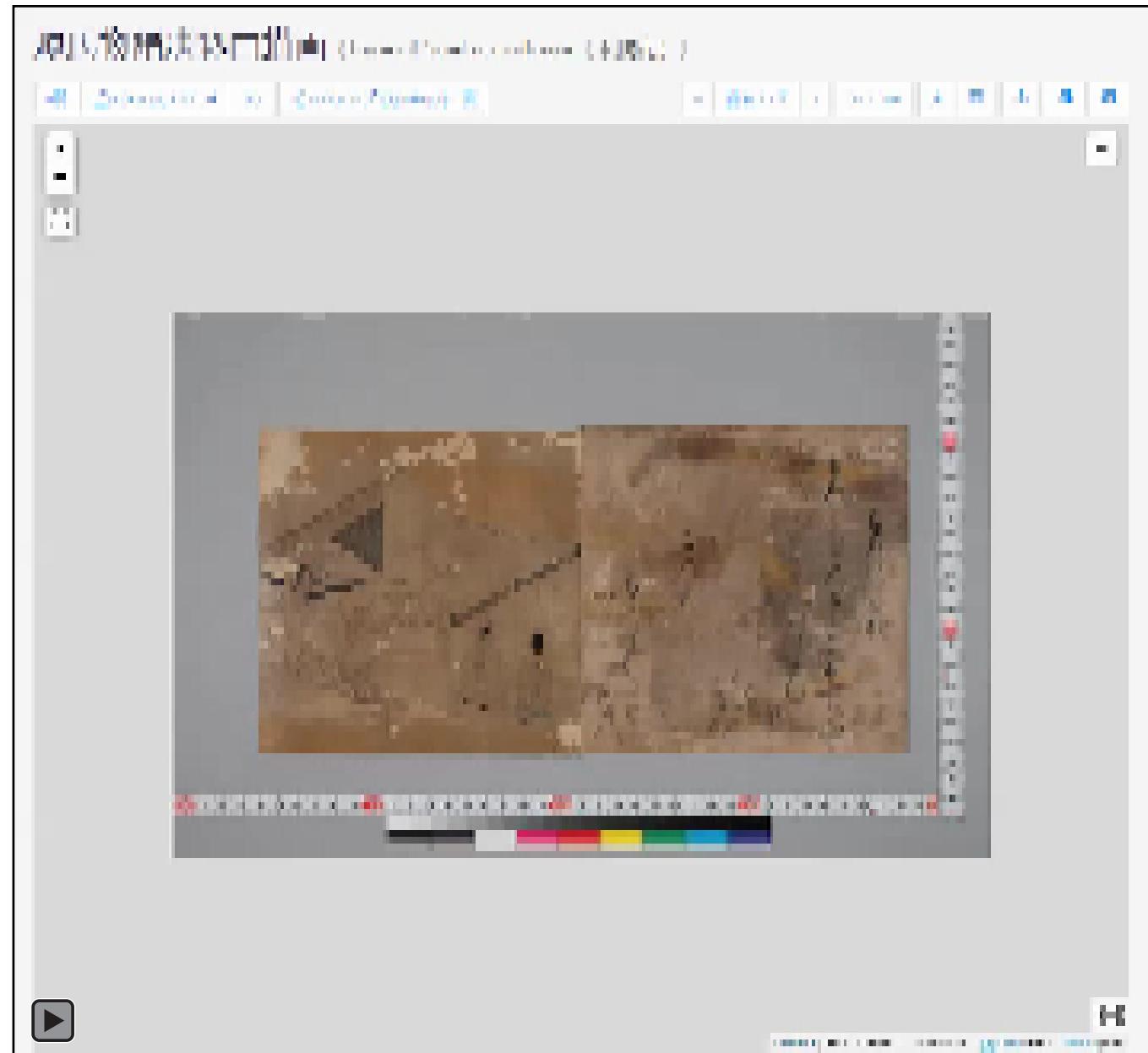
Design of Mt. Fuji

[http://codh.rois.ac.jp/
iiif/iiif-curation-
viewer/index.html?cu
ration=/software/iiif-
curation-
viewer/demo/fuji.json](http://codh.rois.ac.jp/iiif/iiif-curation-viewer/index.html?curation=/software/iiif-curation-viewer/demo/fuji.json)



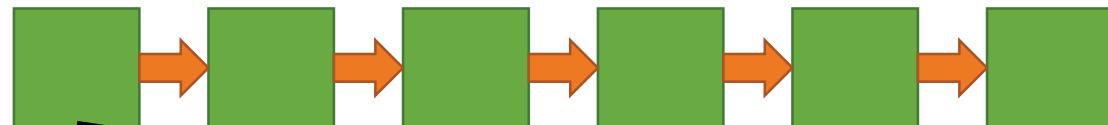
Face of Suetsumuhana

[http://codh.rois.ac.jp/
iiif/iiif-curation-
viewer/index.html?cu
ration=/software/iiif-
curation-
viewer/demo/suetsu
mu.json](http://codh.rois.ac.jp/iiif/iiif-curation-viewer/index.html?curation=/software/iiif-curation-viewer/demo/suetsumu.json)

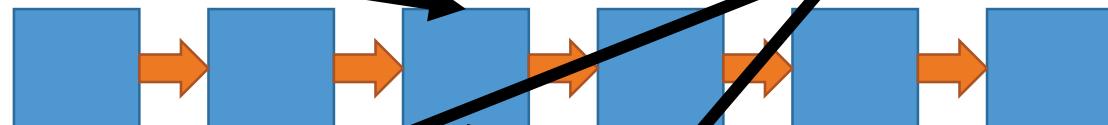


Fragmentation + Recombination

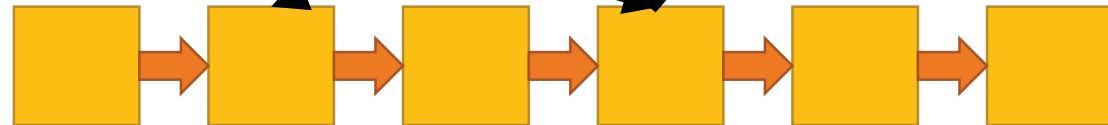
Manifest



Manifest



Manifest



- A manifest is **fragmented** into canvases, and then **recombined** by a curator.
- **Manifest is about content, while curation is about context.**

Evolution of Curations

- The result of curation can be downloaded as a JSON file and shared across viewers.
- Combining with Timeline, you can also make a curation of objects for a long sequence.
- Registry of curations could be used for discovering relevant images and regions.
- A curation could be a base for derivative curations to accumulate multiple viewpoints.

Summary

Re-Thinking Presentation API

1. All the extensions were proposed for the upper structure than canvas.
2. Presentation API should realize a good balance between data modeling and usage.
3. Curation focuses on user-driven discovery of stories in image archives.
4. User-driven content can link to manifests, but how to link back is a big issue.

Future Plan

- 1. Digital Archive of Toyo Bunko Rare Books**
 - <http://dsr.nii.ac.jp/toyobunko/>
 - 245 manifests, 72,591 images.
 - To become IIIF-compliant in the next month?
- 2. Registration of images across manifests, which is important for an IIIF-based research tool.**
- 3. Registry of curations, which may be useful for the discovery of interesting objects.**
- 4. Inclusion of Cursor API and Curation API into a new Presentation API?**

Acknowledgment

- Leaflet IIIF and Mr. Jack Reed for the foundation of our viewer development.
- Mr. Jun Homma (@2SC1815J) as a core contributor for IIIF Curation Viewer.
- Dr. Chikahiko Suzuki (CODH) for creating two curations.
- IIIF Curation Viewer is copyright of Center for Open Data in the Humanities (CODH), and MIT licensed.
- Note that the purpose of IIIF Curation Viewer is a reference implementation of the proposed APIs, and not meant to be the complete implementation of IIIF Presentation API.