

Open Data in the Humanities: Data Sharing and Publication for Triadic Co-Creation

Asanobu KITAMOTO

Center for Open Data in the Humanities (CODH)

Joint Support-Center for Data Science Research

Research Organization and Information and Systems

National Institute of Informatics

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

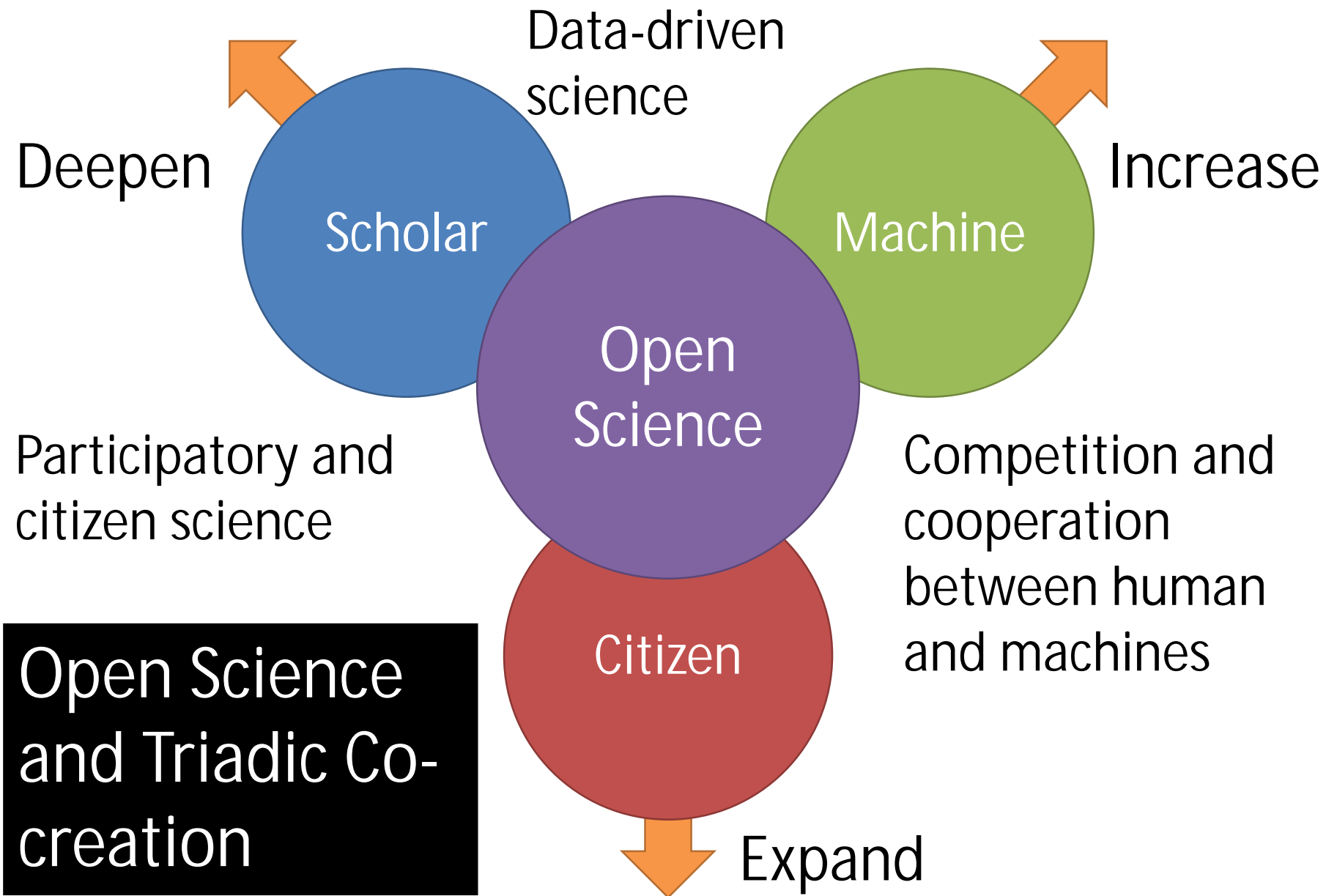
Twitter: @rois_codh



What is CODH?

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

- April 1, 2017: Officially launched. Faculty members consist of NII and ISM.
 - ROIS > Join Support-Center for Data Science Research > CODH
1. **Innovate humanities research by informatics and statistics technology.**
 2. Innovate informatics and statistics research by humanities (big) data.

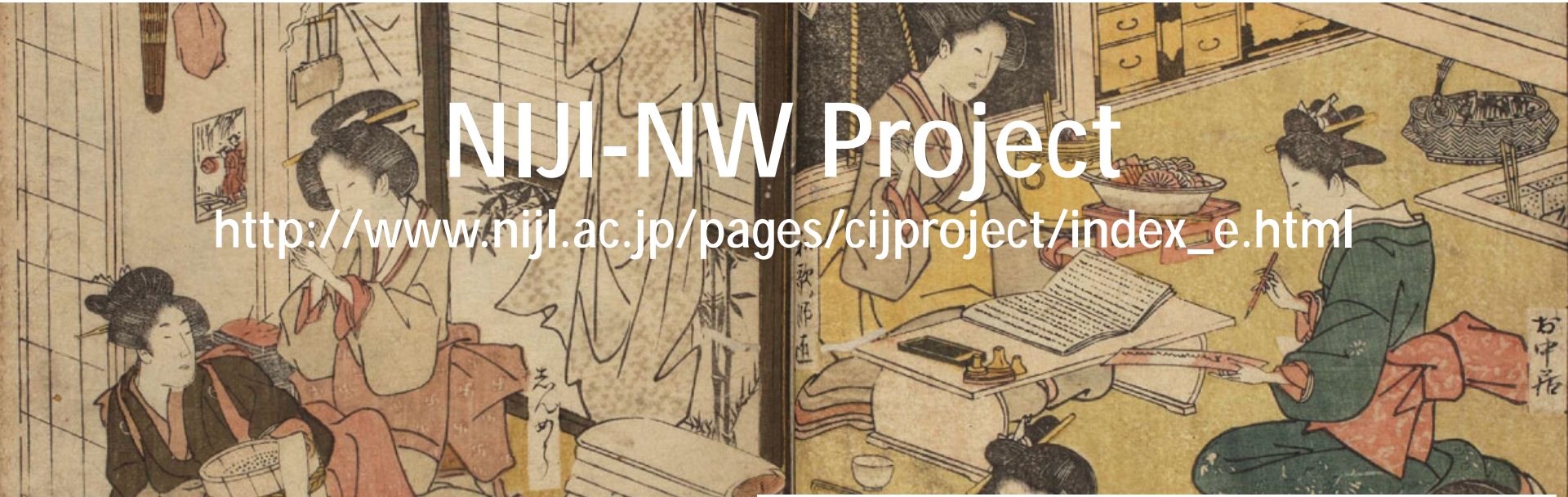


Data Sharing and Open Data for Japanese Old Books

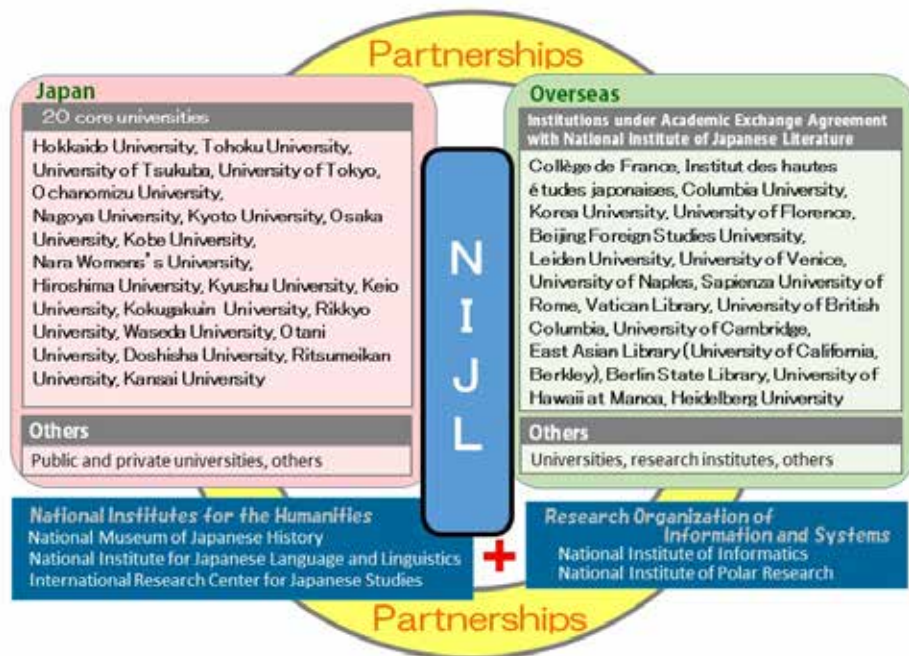
<http://codh.rois.ac.jp/>

NIJI-NW Project

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



It was decided to convert approximately 300 thousand “Pre-modern Japanese Books” into image data to be amalgamated with the bibliographic data base to produce the “Database of Pre-modern Japanese Books.”



Open Data for Scholars

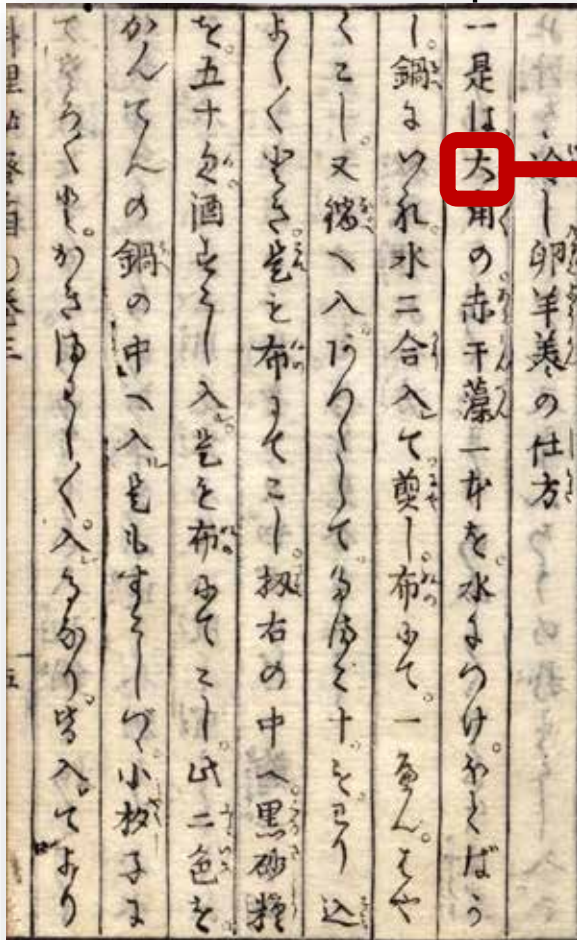
<http://codh.rois.ac.jp/pmjt/>



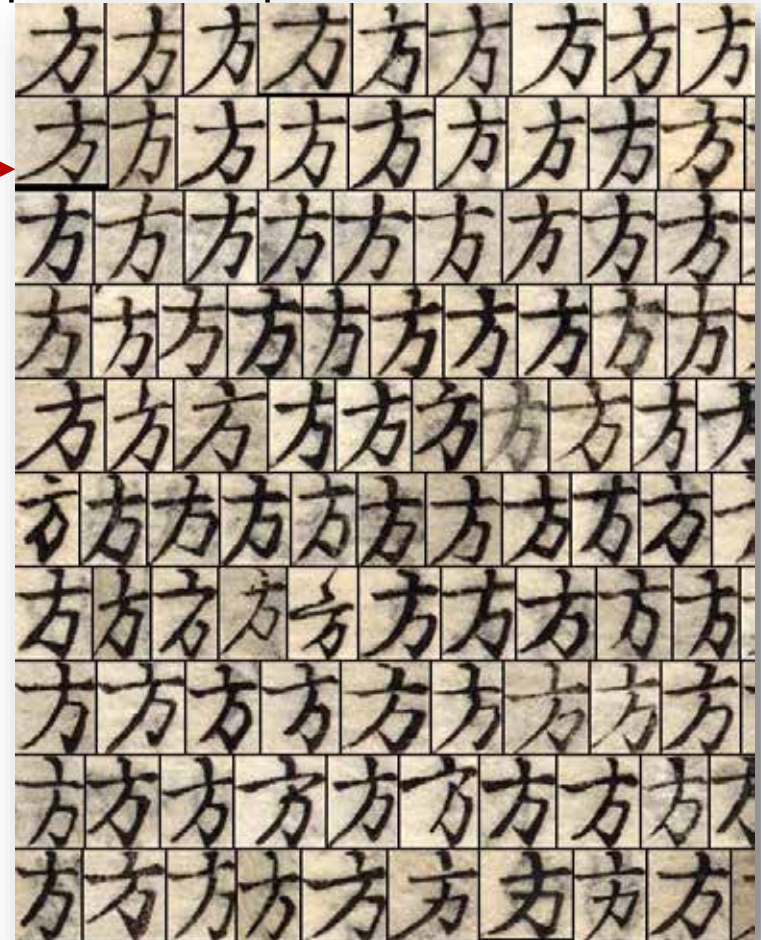
Pre-Modern Japanese Text Dataset (from NIJL)

Open Data for Machines

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



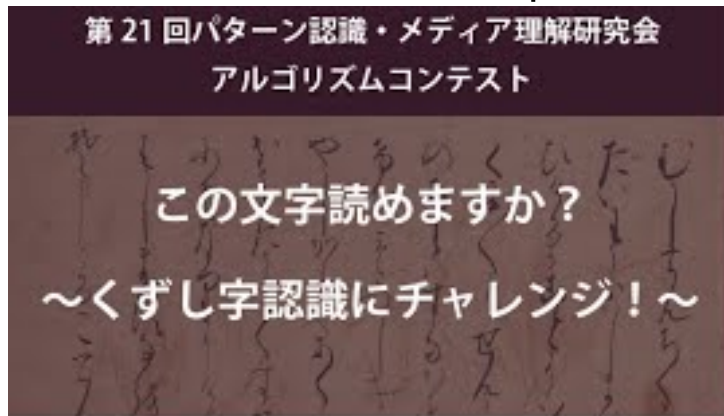
PMJT Dataset (from NIJL)



PMJT Character Shape Dataset
(from NIJL and processed by CODH)

Kuzushiji Challenge!

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



古典籍画像の指定領域に含まれるくずし字を認識して、各字の Unicode を出力する課題です。

使用言語：C++ / Python
開発・実行ができる
仮想環境を配布します



詳細はアルコン HP にて

サンプル公開：2017 年 4 月末
応募開始：2017 年 5 月 31 日
応募締切：2017 年 8 月 31 日



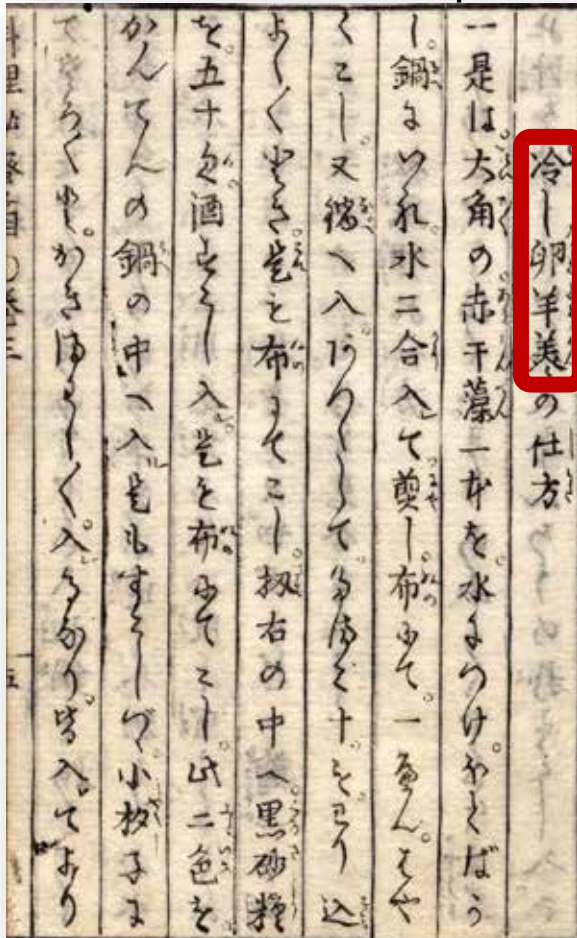
@alcon2017prmu

主催：電子情報通信学会 / パターン認識・メディア理解研究会
後援：国文学研究資料館
情報・システム研究機構 / データサイエンス共同利用基盤施設 / 人文学オープンデータ共同利用センター
情報処理学会 / 人文科学とコンピュータ研究会
データ提供：国文学研究資料館・人文学オープンデータ共同利用センター

- Optical Character Recognition (OCR) does not work.
- Can AI (artificial intelligence) read old characters?
- First competition is finished, and maybe the second next year?

Open Data for Citizens

<http://codh.rois.ac.jp/edo-cooking/>




PMJT Dataset
(from NIJL)



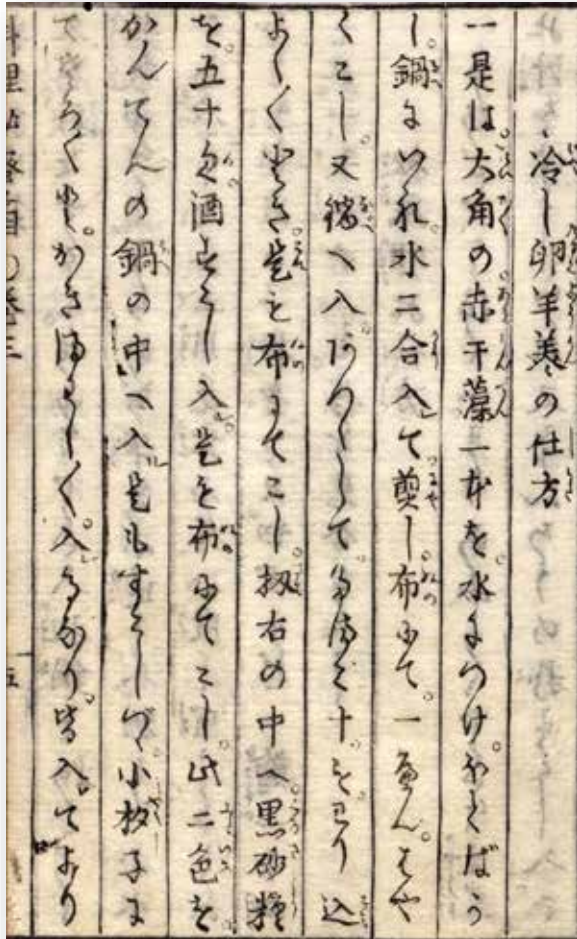
Edo Cooking Recipe Dataset
(Created by CODH)
Adapted Material on NIJL Dataset
(from NIJL)

Edo Cooking Recipe Dataset

- 
1. **Digitize** cooking recipe books.
 2. **Transcribe** old Japanese characters.
 3. **Translate** them into modern Japanese.
 4. **Adapt** translation into a recipe.
 5. **Release** the recipe at Cookpad.
 6. **Share** experiences at “Tsukurepo.”

Collaborated with AMANE LLC.

2. Transcription

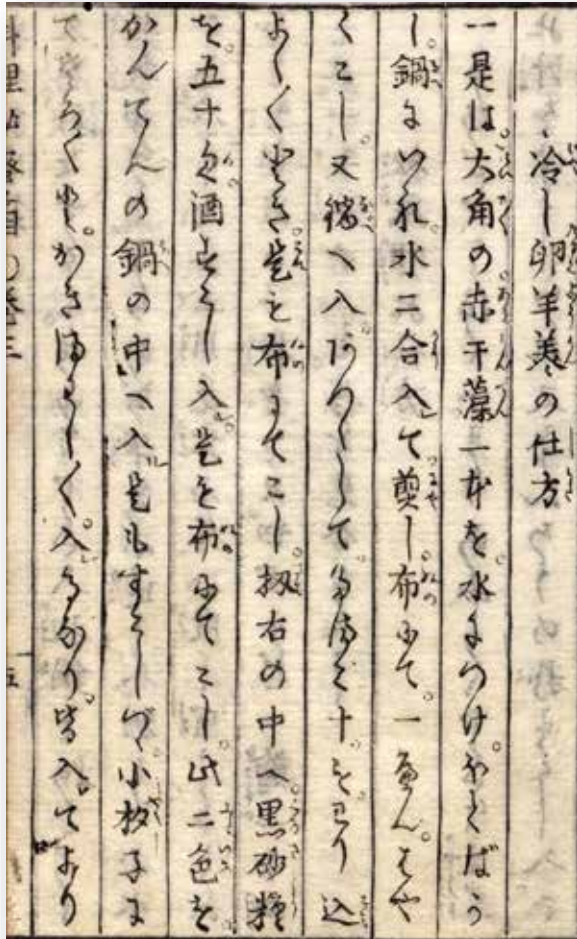


PMJT Dataset (from NIJL)

1	是は大角の赤干藻一本を水につけほとばかし
2	鍋にいれ水二合入れて煎し布にて一へんはやくこし又鍋へ入れあつくして
3	たまご十ウをわり込よくよくとき是も布にてこし
4	切右の中へ黒砂糖を五十匁酒すこし入ル是も布にてこし
5	此二色を かんてんの鍋の中へ入ル
6	是もすこしづつ小杓子にてそろそろと かきまわしかきまわし入れるなり
7	皆入れてより 又葛粉をすこし水にてとき入レ
8	切鍋をぬき 早く折敷にても うちあげ 平めに延し入レ物ともに 水に入レ 冷し遣ふ

Edo Cooking Recipe Dataset (Created by CODH)

3. Translation

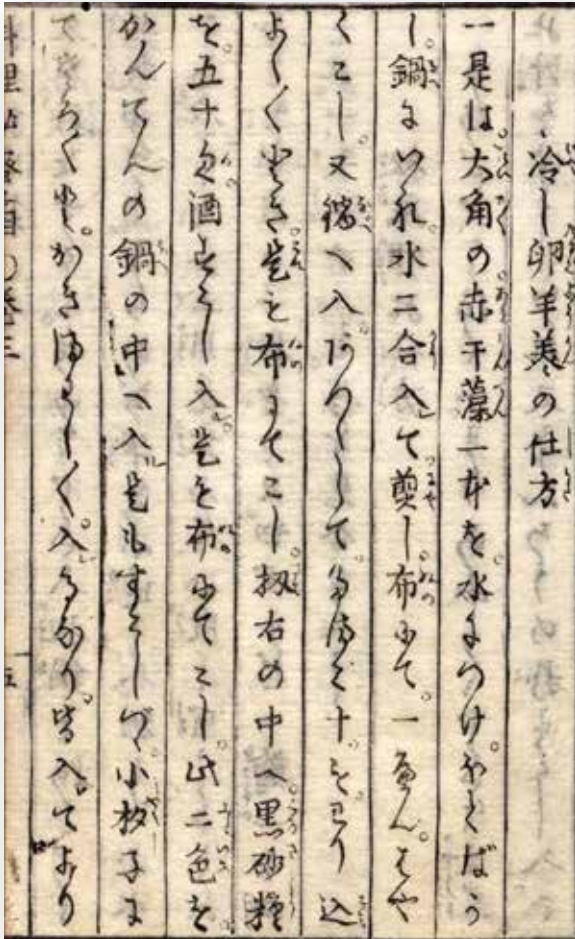


PMJT Dataset (from NIJL)

1	大きな赤寒天を1本水に付けてふやかす。
2	鍋に寒天と水2合 (360cc) を入れて煮溶かす。
3	を一度布で素早く漉し、再び鍋に入れて熱する。
4	生卵10個をよく溶き、布で漉す。
5	の中に黒砂糖50匁 (200g) と酒少しを入れ、布で漉す。
6	を寒天の鍋に入れる。小さな杓子で少しずつそろそろと混ぜながら入れる。
7	を全て鍋の中に入れたら、葛粉を水で溶き、鍋に入れる。
8	鍋を火から上げ、素早く中身を容器 (折敷) に広げ、平たく延ばし、容器ともに水で冷やす。

Edo Cooking Recipe Dataset (Created by CODH)

4. Adaptation



PMJT Dataset (from NIJL)

1	寒天を水につけて、ふやかします。
2	生卵をよく溶きます。
3	溶いた生卵を布でこします。
4	黒砂糖と酒を入れ、溶かします。
5	4を3に入れ、再びこします。
6	鍋に寒天と水（180cc）を入れて煮とかします。
7	6を布などでこし、再び鍋に入れて熱します。
8	7の熱した寒天の中に、5の卵液を少しずつ入れます。
9	全て入れ終わったら、水でといた片栗粉を鍋に入れてさっと混ぜ合わせます。
10	鍋を火からあげ、中身を容器に入れます。
11	冷蔵庫で、2時間程度冷やします。

Edo Cooking Recipe Dataset (Created by CODH)

Photographs by Cooking Experts



Dataset Release at 'Cookpad'

Joint work with Cookpad and The Japan Society of Home Economics, Division of Food Culture.



Deposit and release the data from a web service (app) where people are already well familiar with.

cookpad

江戸時代のスイーツ 甘さスツキリ冷卵羊羹

江戸の料亭本から見つけた和風スイーツです。プリン同様ですが、牛乳不使用でさらっとした菓種の甘さがやみつきになります。

材料 (約4名分)

卵	5個
寒天(赤)	1本(4g)
黒砂糖	100g
水	180cc
片栗粉	適量
酒	適量

カロリー: 276kcal/人 糖質: 0.3g/人

1. 寒天を水につけて、ふやかします。

2. 生卵をよく溶きまです。

3. 溶いた生卵を布でこします。

4. 3に入れ、再びこします。

5. 鍋に寒天と水(180cc)を入れて煮とがします。

6. 6を布などでこし、再び鍋に入れて熱します。

7. 7の熱した寒天の中に、5の卵液を少しずつ入れます。

<http://cookpad.com/recipe/4153357>

Big Impact from the Release

人文学オープンデータ共同利用センターさんがリツイート

うすら @caille2006 · 11月26日
このプロジェクトがすごいのは、古文書の情報をさらに現代の生きた情報にするために、クックパッドにアカウントを開設してレシピを公開し「つくれぽ」も受け付けていること。江戸ご飯とつくれぽというこの未来感バネい。 cookpad.com/kitchen/146046...



クックパッド江戸ご飯 のキッチン

プロフィール

トップ	レシピ 32	つくれぽ 0	献立 0
-----	-----------	-----------	---------

レシピを検索

7317 retweets

<https://twitter.com/caille2006/status/802575840819089409>

2017/12/06

人文学オープンデータ共同利用センターさんがリツイート

NII 国立情報学研究所(NII) @jouhouken · 11月24日
[プレスリリース]
江戸の文化を現代に取り込む「江戸料理レシピデータセット」を整備～江戸時代の料理本を「レシピ化」し、クックパッドでも公開～
nii.ac.jp/news/2016/1124



1 1,074 971

1052 retweets

<https://twitter.com/jouhouken/status/801693251052781568>

IIIF (International Image Interoperability Framework) for Data Sharing and Publication

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



IIF-based Image Delivery

- IIF (International Image Interoperability Framework) is now widely used in humanities-related communities.
 1. **Image API**: Delivery of single images.
 2. **Presentation API**: Delivery of a set of images (e.g. books) with metadata
- Interoperable APIs allow people to develop and use **digital tools that fit all**.

Commercial & Open Source Communities

Image Apps

- Universal Viewer
- Mirador
- Internet Archive Book Reader
- LUNA
- divajs

Image Clients

- OpenSeadragon
- OpenLayers
- IIP Moo Viewer
- Leaflet** (circled)
- Leaflet JS
- FSI Viewer

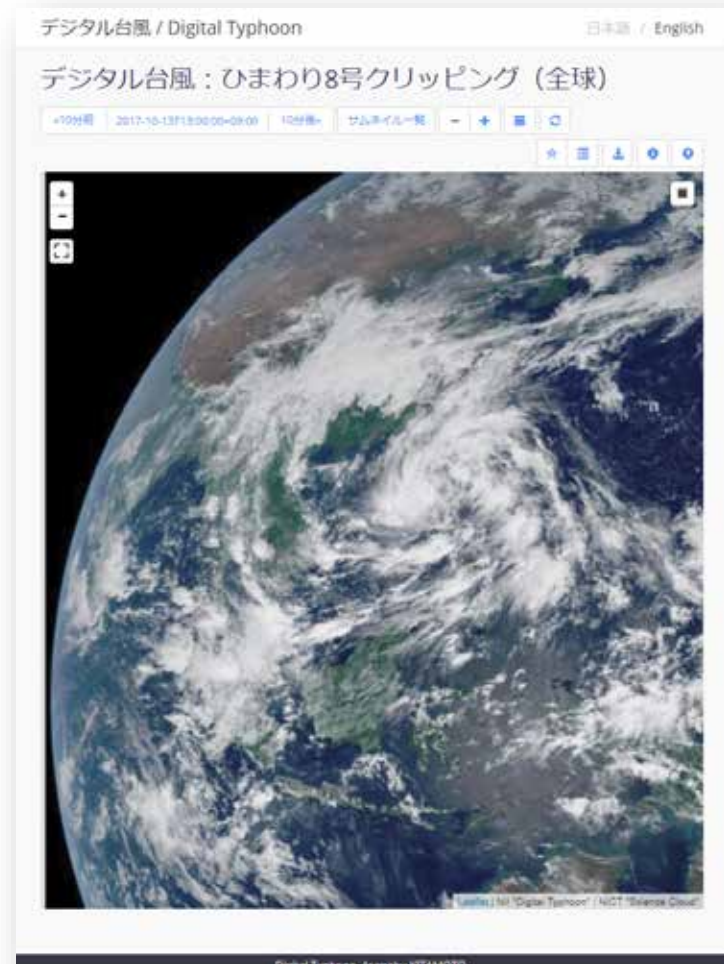
Image Servers

- IIP Image** (circled)
- FSI Server
- djatoka
- LuraTech
- CONTENTdm®
- LORIS
- digilib

Sheila Rabun, IIF Community Groups & Engagement, IIF Conference 2017.

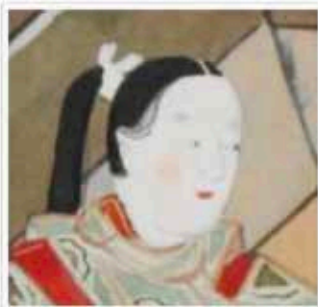
IIIF Curation Viewer (for Timeline)

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





1



2



3



4



5



6



7



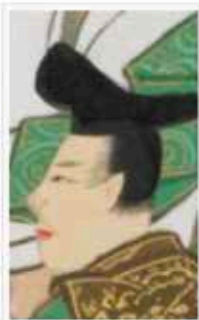
8



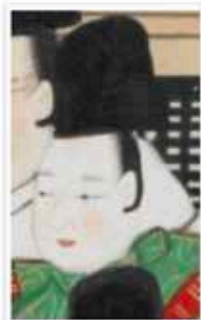
9



10



11



12



13



14



15



16



17



18



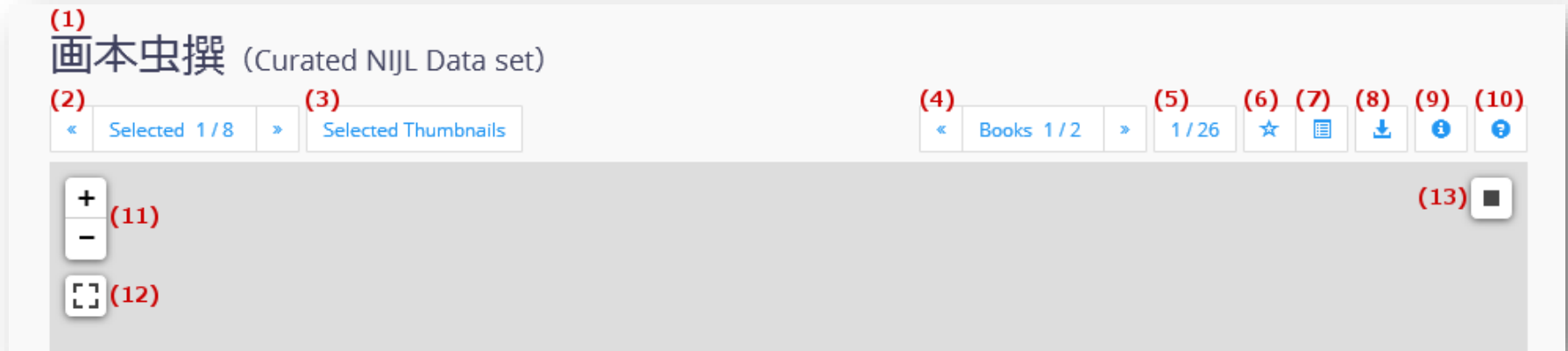
19



20

『宇津保物語』日本古典籍データセット（国文研所蔵）CODH配信

Curation on the Viewer



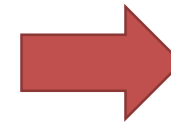
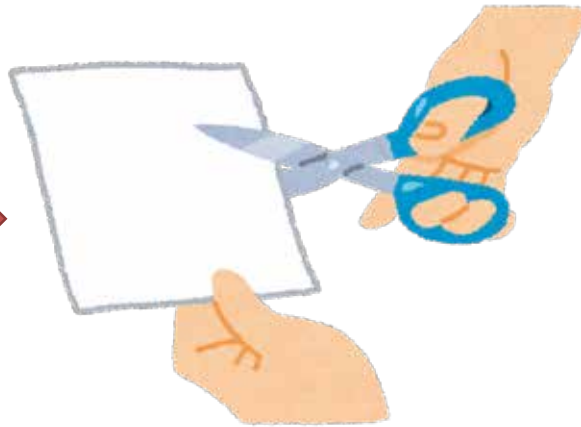
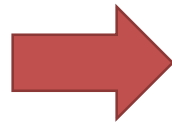
- We define curation as selection and ordering of interesting objects from the collection.
- ' ' (13) is a tool to draw a rectangle on a canvas to select the region of interest.
- ' ' (6) is a "favorite" button to keep interesting objects (the entire image or a region)

Good Old Analogue World



1 Scissors

2 Paste



Source: いらすとや, <http://www.irasutoya.com/>

神図7 粟田口隆光作品の官貌の比較

粟田口隆光作品	『石山寺縁起絵巻』	粟田口隆光作品	『石山寺縁起絵巻』
			
豊田宗廟 上巻第一段	第二段	豊田宗廟 上巻第二段	第一段
			
豊田宗廟 上巻第二段	第一段	織田念仏 上巻第十段	第三段
			
豊田宗廟 上巻第一段	第三段	織田念仏 上巻第十段	第三段
			
豊田宗廟 上巻第四段	第一段	豊田宗廟 上巻第二段	第一段
			
豊田宗廟 上巻第二段	第一段	豊田宗廟 上巻第一段	第一段



神図5 『織田念仏縁起絵巻』 清涼寺



神図6 『豊田宗廟縁起絵巻』 豊田八幡宮



『石山寺縁起絵巻』



『織田念仏縁起』

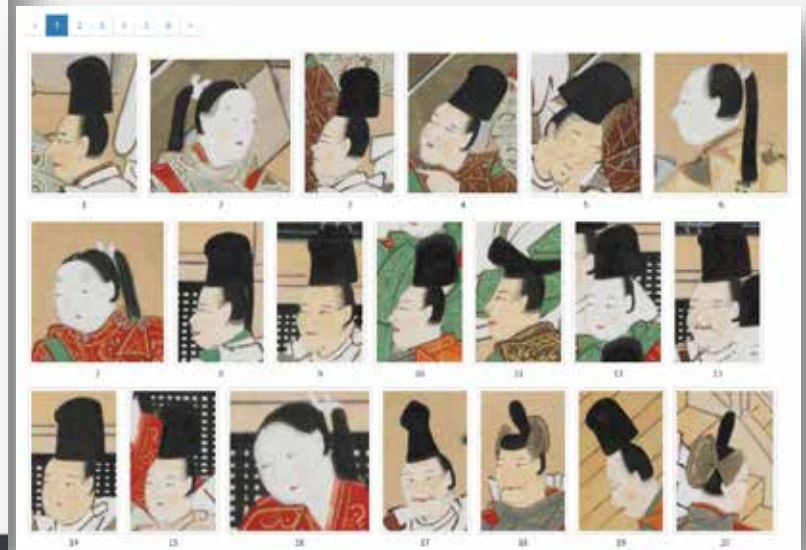
神図7 相沢正彦氏の複製の神図『織田念仏縁起』と『石山寺縁起絵巻』

相沢正彦『石山寺縁起絵巻集成 論考・資料編』中央公論美術出版（2016年）P.20

Frictionless Digital World



1. Draw a box, and 2. Add to favorites – very simple.





2017-07-26T14:30:00+09:00



2017-07-26T14:40:00+09:00



2017-07-26T14:50:00+09:00



2017-07-26T15:00:00+09:00



2017-07-26T15:10:00+09:00



2017-07-26T15:20:00+09:00



2017-07-26T15:30:00+09:00



2017-07-26T15:40:00+09:00



2017-07-26T15:50:00+09:00

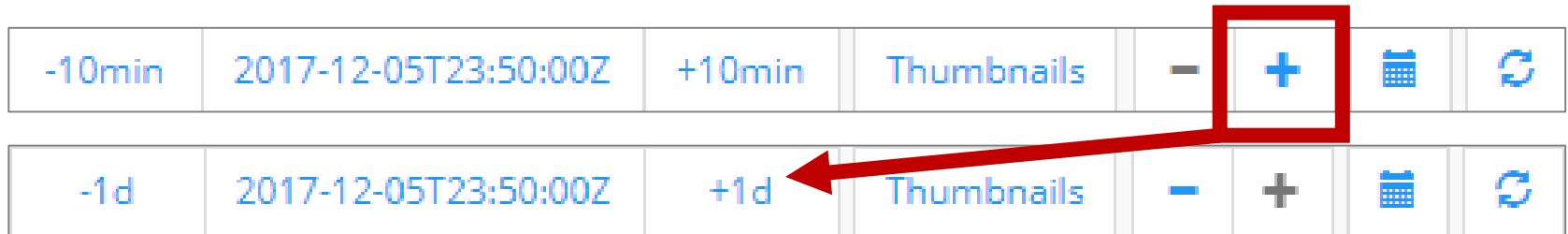


2017-07-26T16:00:00+09:00



ひまわり8号クリッピング : <http://agora.ex.nii.ac.jp/digital-typhoon/himawari-3g/clipping/>

Navigation of Page or Time



1. **Generalization of a book:** for scientific time-series data, "next page" should be generalized to "next observation time."
2. Time interval can be changed by the button, which is pre-defined from 10 minutes (min) to 1 day (max).

Sharing Interesting Scenes

<http://agora.ex.nii.ac.jp/digital-typhoon/himawari-3g/gallery/>



Data Publication

<https://codh.repo.nii.ac.jp/>

インデックス

IIIIE ▶ キュレーション

<http://doi.org/10.20676/00000321>

Permalink : <http://doi.org/10.20676/00000321>

「羅生門」顔貌キュレーション

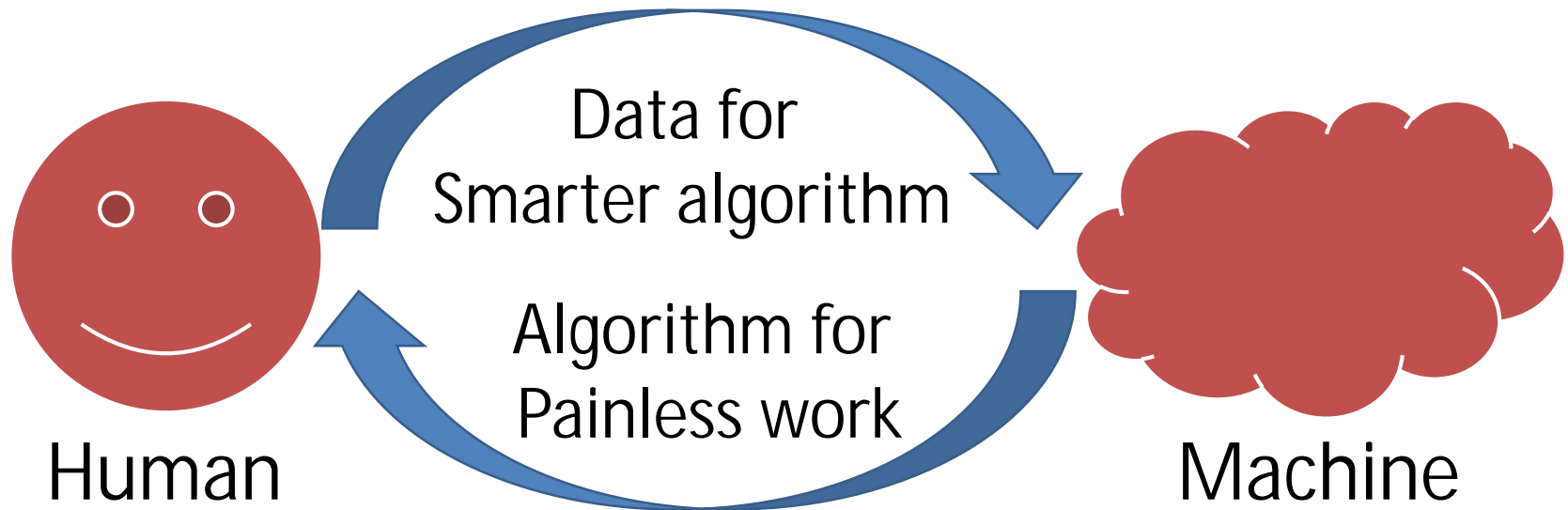
利用統計を見る

JALC DOI	info:doi/10.20676/00000321
アイテムタイプ	Data or Dataset (Curation Json)
言語	日本語
作成者	鈴木 義彦 / スズキ テカヒコ (CiNii ID: 9000245823692)
作成者(其他言語)	Suzuki Chikahiko (CiNii ID: 9000245823692)
公開者	人文学オープンデータ共創利用センター / ジャンプンガクオープン
概要	国文学研究資料館「羅生門物語」 (DOI 10.20730/200003096)
キュレーション対象 (キュレーションコマ数)	登場人物の顔貌 (41コマ)
フォーマット	curation.json
資源識別子URI	「羅生門」顔貌キュレーション閲覧
本文フルテキストへのリンク	http://codh.rois.ac.jp/software/iiif-curation-viewer/demo/index/curation=http://codh.rois.ac.jp/curation/exhibition/3/json/ros
権利	CC : BY-SA
出典情報	国文学研究資料館「羅生門物語」
出典画像	国文学研究資料館「羅生門物語」



@ JAIRO Cloud Repository

Human-Machine Co-Evolution



1. Curation = **annotation about interesting regions with simple metadata (tagging)**.
2. Curation = **training data for machine learning** (e.g. face recognition).

Summary

1. **Triadic co-creation**: scholars, machines and citizens collaborate each other to promote data-driven science.
2. **Japanese old Books**: Open data should be designed to increase the potential of usage.
3. **IIIF**: interoperable technology realizes frictionless infrastructure for data sharing and publication.

Related Websites

- **Center for Open Data in the Humanities (CODH)**
 - <http://codh.rois.ac.jp/>
- **IIIF**
 - <http://codh.rois.ac.jp/>
- **Himawari-8 Clipping**
 - <http://agora.ex.nii.ac.jp/digital-typhoon/himawari-3g/clipping/>
- **Open Science**
 - <http://agora.ex.nii.ac.jp/~kitamoto/research/open-science/>