

# FAIRness for Citizens: Workflow and Platform for Open Data with a Case Study on Edo Cooking Recipes



Asanobu KITAMOTO

Director, Center for Open Data in the Humanities (CODH) and  
National Institute of Informatics

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

@rois\_codh

# What is CODH?

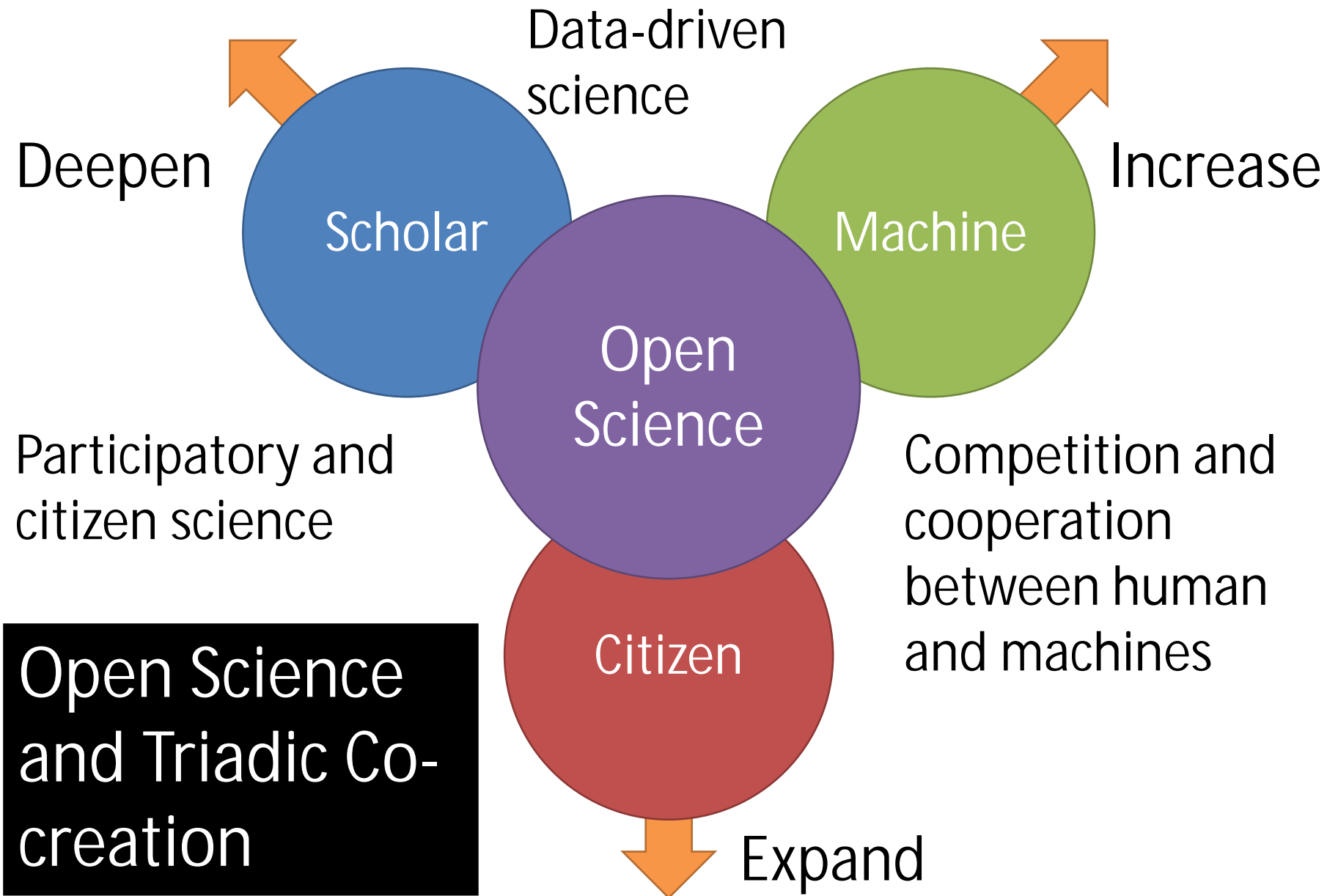
- April 1, 2017: Officially launched.
- One of the centers in Joint Support-Center for **Data Science Research**.
- 1. Humanities research using technology and methodology of informatics and statistics.**
- 2. Informatics and statistics research using humanities data.
- **NIJL-NW Project is one of our collaborators.**

# NIJL-NW Project



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

- Digitize 300,000 Japanese Premodern books.
- Digital images will be open data (CC BY-SA).
- The first “Big Data” on Japanese culture.



# Organization of the Paper

1. Open data is a critical ingredient for data-driven research.
2. What is the best practice for promoting the usage of open data by citizens?
3. FAIR guiding principle may be a good starting point to identify relevant factors.
4. FAIR is developed for research data, but is it also applied to data for citizens?

# FAIR Data Principles

<https://www.force11.org/group/fairgroup/fairprinciples>

- **The FAIR Data Principle**: a set of guiding principles to make data **Findable, Accessible, Interoperable, and Re-usable**.
  1. **Findable**: identifiers, metadata, repository
  2. **Accessible**: open protocol, certificate
  3. **Interoperable**: common format, vocabulary
  4. **Re-usable**: license, provenance, standard

# Open Data for Scholars

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



Pre-Modern Japanese Text Dataset (from NIJL)

# FAIRness for Scholars

- **Findable**: DOI (digital object identifier) is assigned for each book.
- **Accessible**: all the data and metadata are available in a standard protocol and format.
- **Interoperable**: Image is delivered by IIIF, a widely adopted standard for image delivery.
- **Re-usable**: All data are provided by open license, namely CC BY-SA.



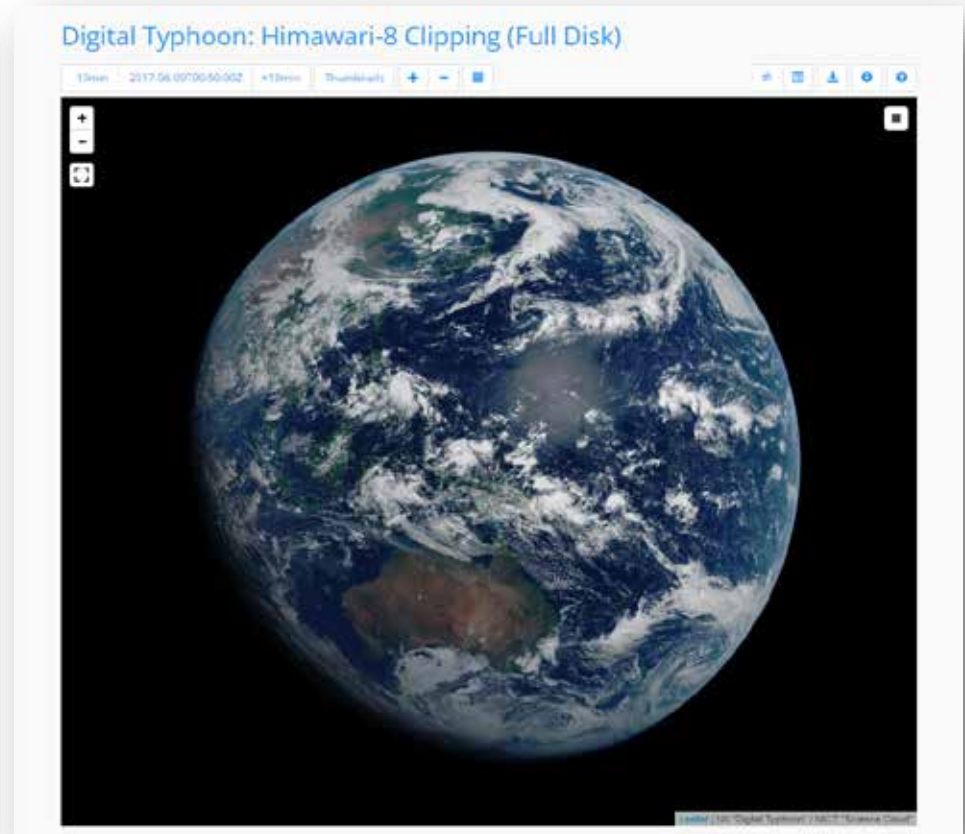
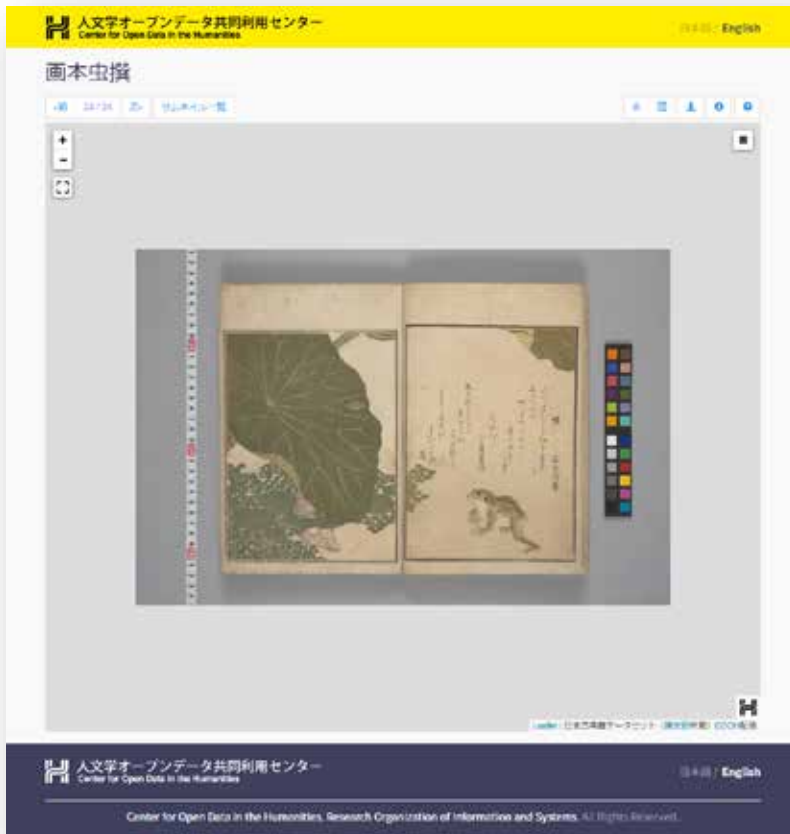


# IIF-based Image Delivery

- 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**.

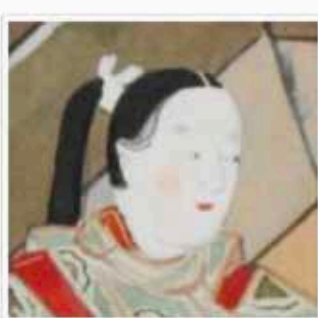
# IIIF Curation Viewer (for Timeline)

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





1



2



3



4



5



6



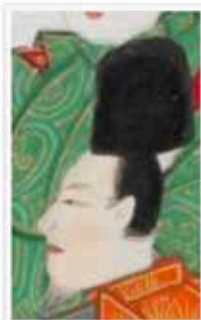
7



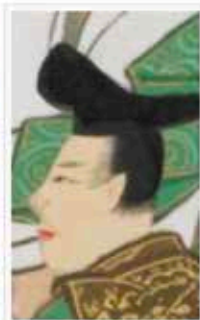
8



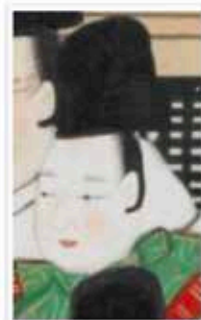
9



10



11



12



13



14



15



16



17



18



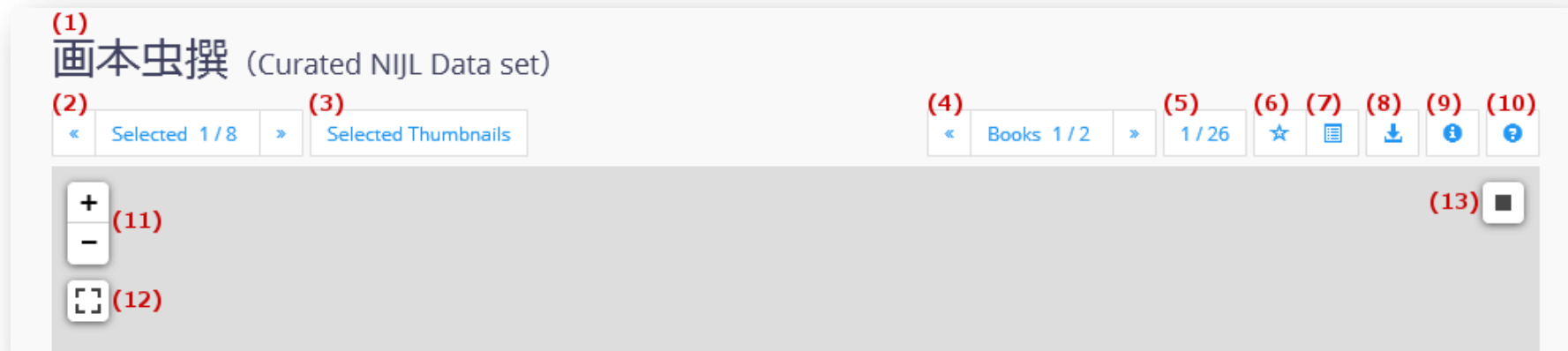
19



20

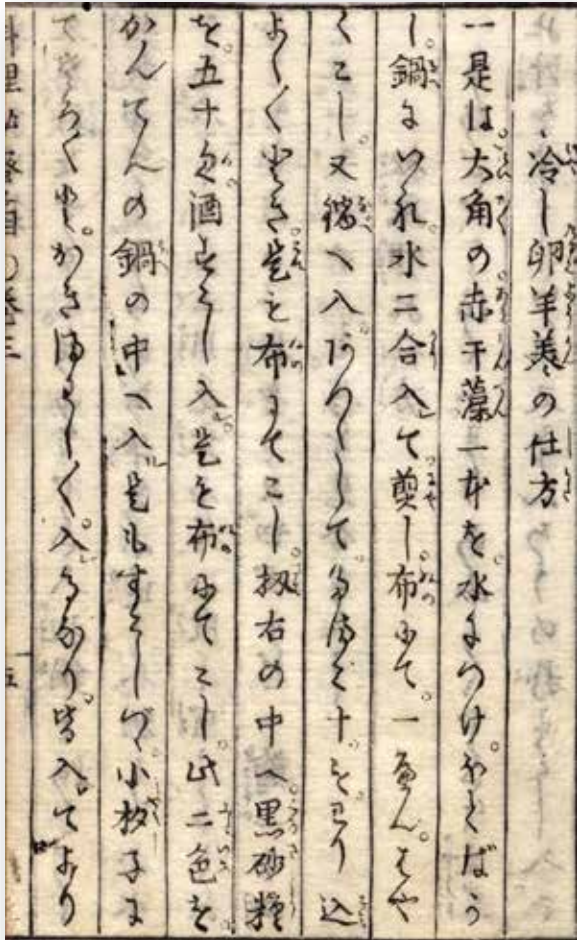
『宇津保物語』日本古典籍データセット（国文研所蔵）CODH配信

# 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 (a whole image or a region)

# Accessibility Problem

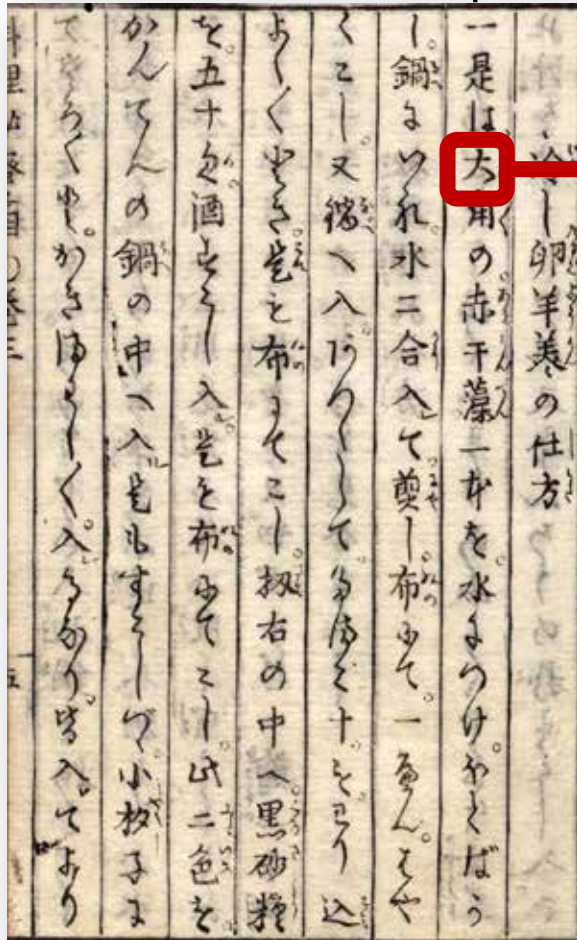


PMJT Dataset  
(from NIJL)

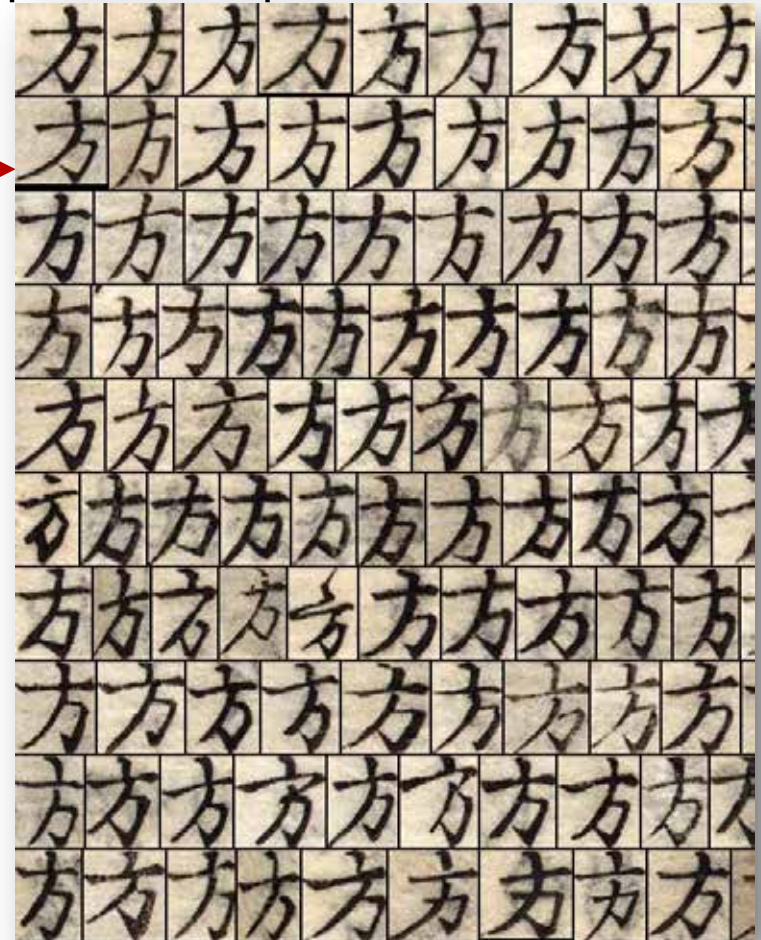
- It is not machine-readable data, because optical character recognition (OCR) software is not available for “kuzushi-ji.”
- It is literary not citizen-readable data, because even native Japanese speakers cannot read “kuzushi-ji.”

# Open Data for Machines

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



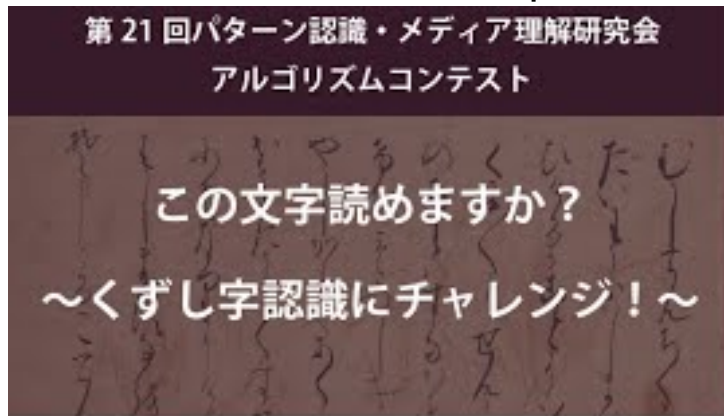
PMJT Dataset (from NIJL)



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

# Kuzushi-ji 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?
- Hold a competition to invite computer scientists to participate.

# Case Study on a Recipe Book



PMJT Dataset  
(from NIJL)

- Target of the case study: *Mambo Ryori Himitsubako*, a cooking book published in 1785 (Edo Period).
- It consists of more than 100 cooking recipes of egg dishes.
- The book indicates culinary culture toward enjoying the variation of cooking.



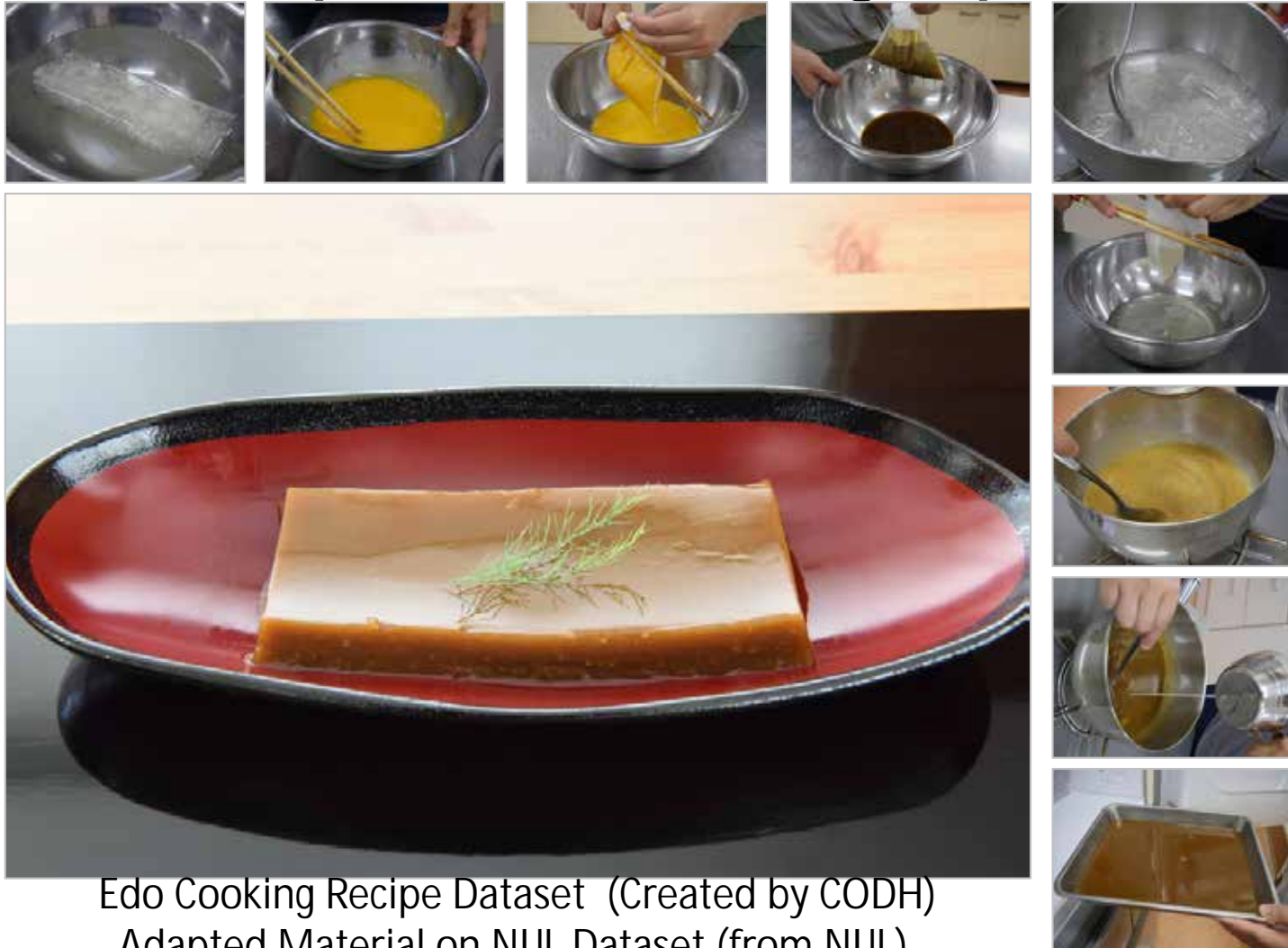
# Increase Accessibility for Citizens

- We designed a workflow to reduce the barrier for citizens.
  1. **Organize** digital images in the Dataset of PMJT.
  2. **Transcribe** text from digital images.
  3. **Translate** old Japanese text to modern Japanese text.
  4. **Transform** modern Japanese text into the modern format of cooking recipes.
- The content of books is more human readable and accessible in the later stage of the workflow.

# Transformation Process

- **Quantity**: time and amount is estimated to reduce the risk of failure and make the recipe actionable.
- **Ingredients**: Some of local ingredients are changed to similar ones that are easier to obtain now.
- **Tools**: Old cooking tools are changed to modern ones, such as refrigerator and food processor, so that we can take advantage of today's civilization.
- **Photographs**: stepwise photographs and an attractive photograph for the final presentation convince people that the recipe is reproducible.

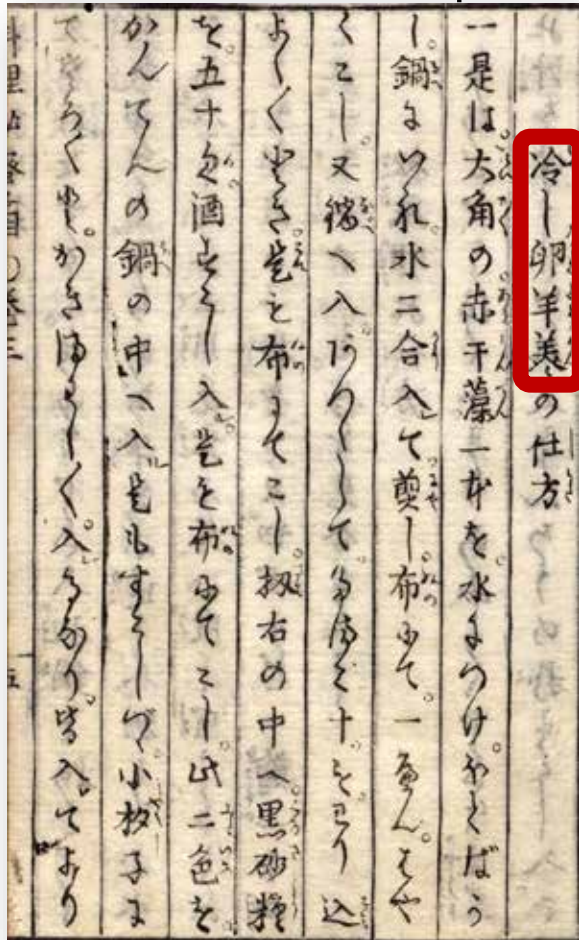
# Stepwise Photographs



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

# 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)

# Where to Deposit Data?

- Cookpad is the largest cooking recipe service in Japan.
- We decided to deposit recipe data to “**Cookpad Edo Dishes,**” run by Cookpad and The Japan Society of Home Economics.



<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

うずら @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...](https://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>

2017/9/11

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](https://nii.ac.jp/news/2016/1124)



1 1,074 971

**1052 retweets**

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

JADH2017

22

# Try to Cook the Recipe

The image shows a screenshot of the news24 website. At the top, there is a navigation bar with various categories like 'ニュース情報', 'ドラマ', 'バラエティ音楽', 'スポーツ', '映画', 'アニメ', 'ミニ番組', 'イベント', '通販', '動画', and 'プレゼント'. Below this is the '0TEL NEWS 24' logo and a date indicator for December 5th, 2016. A large green banner for 'UR賃貸住宅' (UR Rental Housing) is prominent, with text '人は、ふれあって育つ。' and '社宅も UR'. To the right of the banner are four red boxes with white text: '礼金 ¥0', '手数料 ¥0', '更新料 ¥0', and '保証人 不要'. Below the banner is a row of smaller logos for 'Soulberry', 'Sレポニュース', 'every.', 'NEWS ZERO', 'ガソリン', 'NINドキュメント', '深層NEWS', 'Daily Planet', and '0TEL NEWS 24'. A navigation bar with categories like 'トップ', '社会', 'エンタメ', '国際', '政治', '経済', 'スポーツ', '投稿', 'ブログ', '特集', '地方', '天気', '地震', and '防災' is visible. The main article title is '「江戸ご飯レシピ」再現、作って食べてみた' (Reproduction of Edo Rice Recipe, Made and Tasted). The article includes a video player showing a person preparing a dish of dark rice with a hard-boiled egg. A 'Hot Word' label is next to the video. Below the video is a '全文' (Full Text) button. To the right of the article is a sidebar with a clothing advertisement for 'GAUCHIO PANTS' by 'soulberry' priced at 2,149 yen. Below the ad is a search bar and a 'アクセスランキング' (Access Ranking) section with a list of five items.

0TEL NEWS 24  
2016年(平成28年) 12月5日(火) 今日天気 仙台

人は、ふれあって育つ。  
UR賃貸住宅  
社宅も UR  
礼金 ¥0 手数料 ¥0 更新料 ¥0 保証人 不要

Soulberry Sレポニュース every. NEWS ZERO ガソリン NINドキュメント 深層NEWS Daily Planet 0TEL NEWS 24

トップ 社会 エンタメ 国際 政治 経済 スポーツ 投稿 ブログ 特集 地方 天気 地震 防災

トップ > 社会 > 「江戸ご飯レシピ」再現、作って食べてみた

0TEL NEWS 24 「江戸ご飯レシピ」再現、作って食べてみた  
ツイートする シェアする 2016年11月30日 21:05

江戶ご飯レシピ Hot Word

全文

11月28日、「江戸ご飯レシピ」という言葉を含んだツイートが大きく伸びました。

「いま見るべき」「いま知るべき」厳選ニュースをメルマガでお届け

ニュース検索 検索

カレンダー検索

★ アクセスランキング 社会

- 1 豊中市強請刑給「殺意はなく事故」と否認
- 2 電車で女性の髪切り「オークションで売る」
- 3 千葉大医学生3人の英名公表 医師も逮捕
- 4 女性の裸画像「ばらまくぞ」航空自衛官逮捕
- 5 電車の乗降スタッフ降りない社ご乗降

<http://www.news24.jp/articles/2016/11/30/07347892.html>

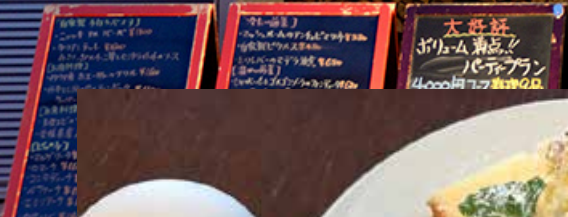
# Photographs by Cooking Experts







Collaboration: Pizzeria Zio Pippo





# Chain of Creations

1. “Tsuku-repo” is a service to submit derivative works by users.
2. Recipes show that they take the recipe into their daily lives.
3. Actionable data can stimulate users to make new actions.

<https://cookpad.com/recipe/3867392/tsukurepos>

2017/9/11

JADH2017

26

# Typical Responses from Citizens



Me: "I will release Edo cooking recipes on our CODH website."  
They: "That sounds nice" (without interests).



Me: "I will release Edo cooking recipes also on Cookpad."  
They: "That's a good idea!" (with a brilliant smile).

# Impacts of Data Release

1. Release of digitized books on CODH website → **no response**.
2. Release of recipe data on CODH website → **a little response, but not so much**.
3. Release of the same recipe data on Cookpad → **unexpectedly large impact**.
4. This result suggests that **the value of release is not in data release itself, but in the choice of deposit website (Cookpad)**.

# Why Citizens were Excited?

1. Hearing the word “Cookpad” they can **instantly imagine how to use the data.**
2. Recipe with photographs convince people that the **recipe is reproducible.**
3. To see data in their familiar platform makes them feel that **it is our data.**
4. A gap between modern platform and old content **attracts curiosity of people.**

# FAIRness for Citizens

- **Findable**: deposit data into a popular platform in the daily lives of citizens.
- **Accessible**: transform the content into modern format with additional information.
- **Interoperable**: interface and description does not require new training.
- **Re-usable**: convince people reproducibility and allow them link their derivative work.

# POCiNP Approach

- **New wine must be put into new wineskins:** Don't try to go along with the old tradition when you want to adapt to a new way of thinking.
- **Putting Old Content into New Platforms (POCiNP):** old content should be transformed to new format so that it is compatible with a new platform such as Cookpad.
- **Inversion of freshness:** citizens perceive old content as *fresh* content which was never seen on a new platform.

# Generalize POCiNP Approach

- **Wikipedian in residence**: Putting museum content to Wikipedia.
- **Participatory transcription**: Putting manual transcription activity to live video streaming.
- **Social networking service**: a very popular variation of this approach.
- Think about **creative combination** between your content and modern platform.



# Summary

- **FAIR guiding principle** is a good starting point for identifying relevant factors.
- **Transformation process** is necessary for increasing accessibility for citizens.
- **Where to deposit data** has a large impact on FAIRness for citizens.
- **POCiNP approach** suggests that old content can revive on a modern platform.

Please visit our Website  
<http://codh.rois.ac.jp/>



Asanobu KITAMOTO

Director, Center for Open Data in the  
Humanities (CODH) and  
National Institute of Informatics

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

@rois\_codh