

デジタルアーカイブと研究データ：人文学における利活用の展望



北本 朝展（KITAMOTO Asanobu）

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（CODH）

国立情報学研究所 コンテンツ科学研究系

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

@rois_codh

DAと研
究データ
の出会い

Digital Archive Meets Research Data

Boy
Meets
Girl

1. Relating to a story, book, or film whose main subject is romantic relationships – **Cambridge Dictionary**
2. A typical romantic situation or story about two people falling in love – **Macmillan Dictionary**
3. Conventionally or trivially romantic – **Collins Dictionary**

“Meets”（出会い系）の流行

Machine Learning Meets Economics
12th Mar '14, 01:28 PM in Video

MIT Technology Review
Topics+ The Download

Discover the emerging technologies that will **CHANGE OUR WORLD**

View from the Marketplace
Customer Data Meets AI
A new day is dawning for the customer experience, driven by the application of artificial intelligence, machine learning, and automated technologies to CRM data. The potential exists to transform the customer's experience by providing service in a more predictive and intuitive way than ever before.
by MIT Technology Review Custom July 17, 2017

IBM
IBM Research

AI meets IoT [feedback](#)

Deep Genomics: Artificial Intelligence Meets the Human Genome
by VLAB

KDD2017 Workshop
Machine learning meets fashion
Data, algorithms and analytics for the fashion industry
14 August 2017, Halifax, Nova Scotia - Canada

「デジタルアーカイブ」と 「研究データ」の出会い

1. ロマンティックな出会いには、
「二人」の共感が重要？
2. 個性の違いを認めつつ、「同じ方向に歩む」ことで意気投合する。
3. ありきたりではない、意外な組み合わせに価値が生まれる。
4. 今日のシンポジウムで、良い出会いがあるといいですね！！

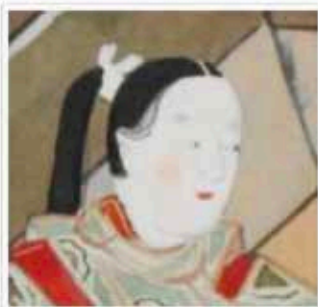
個性の違いと共感

1. 「デジタルアーカイブ」という言葉は、**出身ごとに考え方が大きく異なる。**
2. 「研究データ」という言葉は、**昔からあるものの現代的な意味が加わる。**
3. 両者は行動範囲が異なるため、これまで出会うチャンスは少なかった。
4. **人類知を創造し継承すること**については、共感できるところが多いはず。

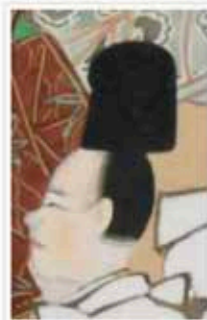
「デジタルアーカイブ」と「研究データ」



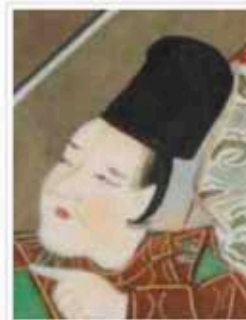
1



2



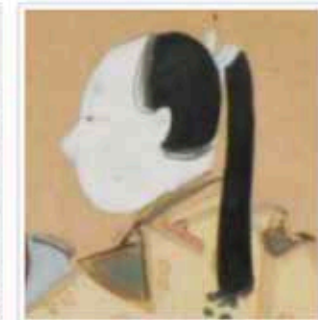
3



4



5



6



7



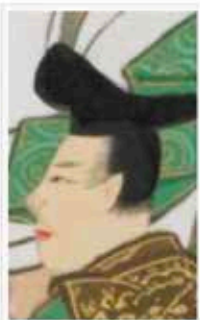
8



9



10



11



12



13



14



15



16



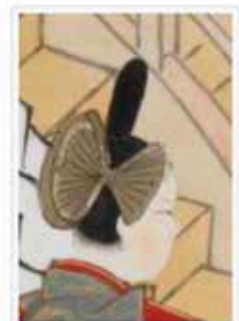
17



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

デジタルアーカイブとは？

1. **私の定義**：デジタル時代における人類知の創造と継承のプラットフォーム。
2. **アナログ時代との根本的な違い**：すべてのモノが「データ」に収束していく。
3. **新しい共通性**：データに関する問題は、組織横断的に共通性がある。
4. **職能の再定義**：モノの種類に基づく職能から、価値の種類に基づく職能へ。

デジタル空間での再統合

アルケイオン？

文字
複製物

文字
現物

モノ
現物

ライブラリ

アーカイブズ

ミュージアム

アナログ

アナログ

アナログ

デジタル

デジタル

デジタル

デジタルアーカイブ

DAで考えるべき3つの価値

1. **基礎価値**：時代に左右されず、安定的な価値を提供する規律。
 2. **付加価値**：現代を鋭く見抜き、価値の変化を捉える感性。
 3. **長期価値**：まだ見ぬ時代を想像し、将来の価値を見極める洞察。
- 3つの価値をすべて扱える人材は稀。

職能のマッピング

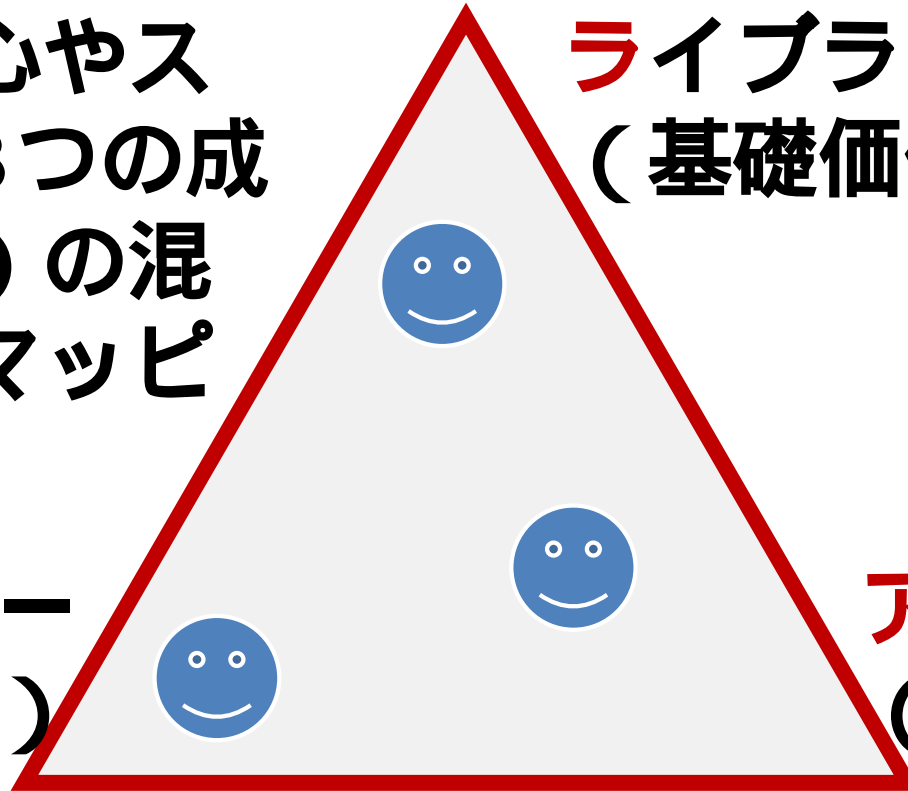
1. **基礎価値** = ライブラリアン
2. **付加価値** = キュレーター
3. **長期価値** = アーキビスト

- 3種類の価値の扱いに長けた人材が、**分業してデジタルアーカイブを担う。**
- 人材としては、**研究者**（開拓）や**ソフトウェアエンジニア**（構築）も必須。

ラ・キ・ア三角形

個人の関心やスキルを、3つの成分（属性）の混合としてマッピング。

キュレーター
（付加価値）



ライブラリアン
（基礎価値）

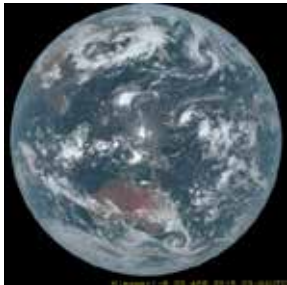
アーキビスト
（長期価値）

新しい専門家の誕生

1. **データライブラリアン**：データの整理のために、メタデータ付与やデータ連携などを安定的に行う。
2. **データキュレーター**：データの利活用のために、コラボレーションなども利用して潜在的な価値を顕在化する。
3. **データアーキビスト**：データの長期保存のために、データの管理・取得・廃棄などに関する計画を策定・実行する。
4. **+ サブジェクト**：内容に関する専門性が武器。

研究データの例 (DIAS)

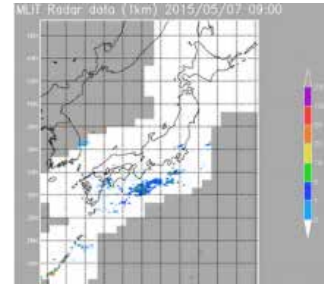
Data Dissemination



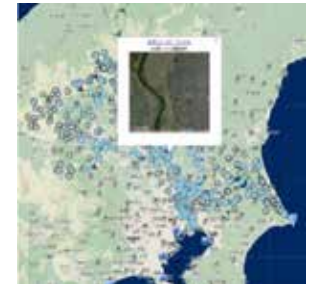
Himawari-8 Satellite



Weather Forecast GPV

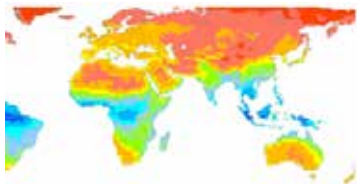


Radar Data

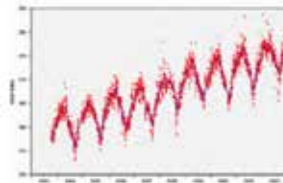


River Telemeters

1. Climate



CMIP5 Model



Visualization Tools

2. Water



Dam Control



Water Management

3. Agriculture

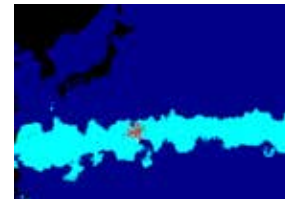


Potential of Rice Crops after Climate Change

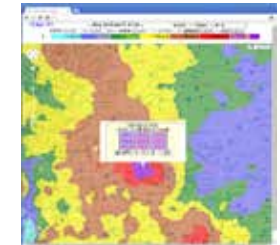
4. Biodiversity



Citizen science-based observations



Fish eggs and growth distribution



Accumulated radar rainfall

研究データの3分類

1. **研究資源データ**：研究の入力として利活用するデータ。原則オープン。
 2. **研究過程データ**：研究の過程で生み出されるデータ。原則クローズ。
 3. **研究成果データ**：研究の出力（成果の証拠）となるデータ。クローズから原則オープンへ。
- **ビッグデータ / AI時代**：利活用が重要。
 - **オープンサイエンス時代**：利活用のための公開に加え、成果の証拠の公開も。

研究データの3つの価値

1. **基礎価値**：時代に左右されず、安定的な価値を提供する規律。
 2. **付加価値**：時代を鋭く見抜き、価値の変化を捉える感性。
 3. **長期価値**：まだ見ぬ時代を想像し、将来の価値を見極める洞察力。
- デジタルアーカイブと基本的には同じ？
 - **人類知を創造する現場に出て、価値増大プロセスに参与できる余地がある。**

現場に出る専門家

1. アーティストと協働するキュレーター、作家と協働する編集者のように、**現場に出て創造者と協働する専門家。**
2. 研究者と現場で議論しながらデータ作成・管理方法を調整するなど、**流動的・進化的なプロセスを回せる専門家。**
3. 価値が定まらない、または価値が後に大きく変化するなど、**混沌とした現場を俯瞰的・長期的に分析できる専門家。**

「デジタルアーカイブ」と「研究データ」 の統合

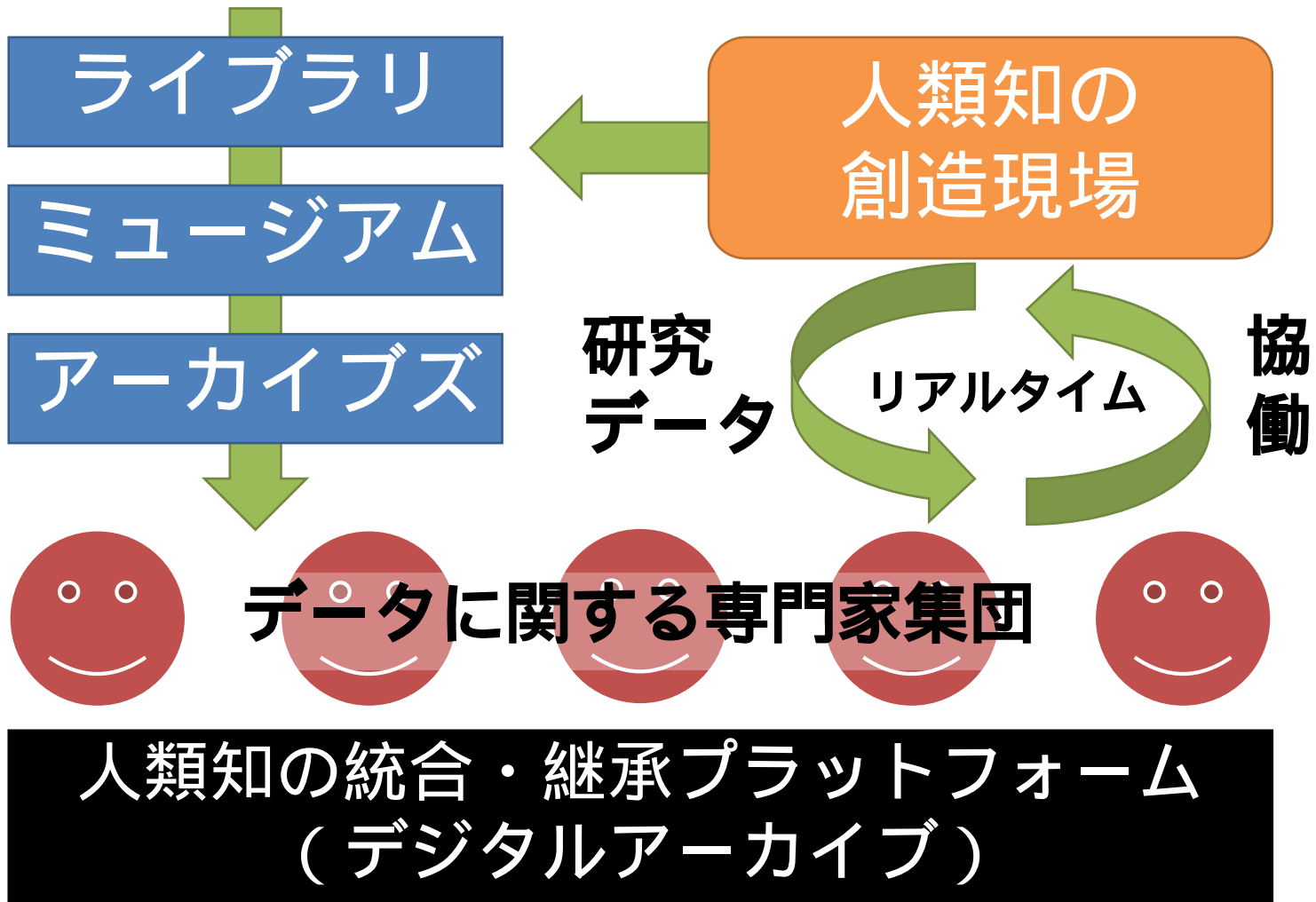
DA第1期→DA第2期→OS

1. **第1期（2000年以前）**：価値が定まったお宝（国宝等）だからDAに入れるという、**優品主義**が支配的だった。
2. **第2期（2011年以降）**：**東日本大震災**を受けて、後世に継承すべきデータの範囲が大きく広がった。
3. **2014年頃～**：**オープンサイエンス**の流れで「研究データ」の議論が深まり、DAと共通する課題も浮上してきた。

DAと研究データの比較

1. デジタル人類知の利活用と継承の点では**技術や人材の問題など共通点が多い。**
2. **知の創造プロセスの現場に出て行くことは、DAよりも重要性が高い。**
3. **透明性・研究公正のための長期保存など、DAにはなかった要請もある。**
4. **研究データは競争力に直結するため、DAよりも商業化の動きが加速している。**

DAプラットフォーム



デジタル時代の専門性と人材

1. アナログ時代は、異なる種類の「モノ」を扱う専門性が必要だった。
2. デジタル時代は、異なる種類の「価値」を扱う専門性が必要となる。
3. プラットフォームの主体がモノから価値に転換→専門性も再定義が必要。
4. こうした専門性を備えた人材は稀少ないので、育成支援の仕組みが必要。

「出会い」の事例



人文学オープンデータ共同 利用センター (CODH)

人文学のデジタルアーカイブと研究データの両方を扱う

人文学オープンデータ共同利用センター
Center for Open Data in the Humanities

日本語 English メニュー

『源氏物語』日本古書群データセット (国文学研究館)

人文学オープンデータ共同利用センター (Center for Open Data in the Humanities / CODH) は、情報学・統計学の最新技術を用いて人文学データへのアクセスを改善する研究開発を進めるとともに、オープンサイエンスの考え方に基き多くの人が参加できるデータプラットフォームを構築することで、データ駆動型の人文学研究や学際的な人文学研究など、情報技術を用いた新しい人文学の方法論を開拓します。[もっと詳しく.]

お知らせ

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

文化遺産のデジタルアーカイブ
デジタル・シルクロード
国文学研究館

Digital Silk Road

Digital Archives of Cultural Heritage
Digital Silk Road Project
Nagasaki University of International Studies

検索 出典

デジタル・シルクロード・プロジェクト Digital Silk Road Projectは、情報学と人文学の協働に基づき大規模なデジタルアーカイブを構築する、デジタル・ヒューマニティーズの研究プロジェクトです。[もっと詳しく >>>] [シルクロード世界遺産 >>>]

プロジェクト

- 『源氏物語』国文学研究館 デジタルアーカイブ
- シルクロード遺跡データベース
- 地図で見るシルクロード
- スタイン地名データベース
- 高野山デジタルマップ
- 中国石窟データベース
- ペリオドック辞書辞源辞典

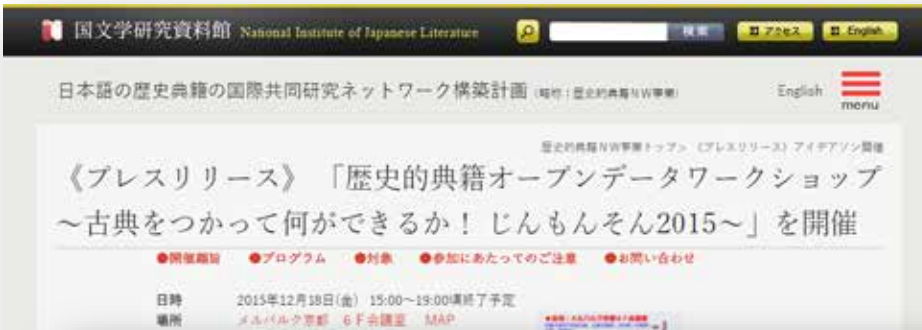
新着情報

- 『東洋文庫西蔵』西蔵デジタルアーカイブ
10冊目による高野山遺跡の公開+最終への1000の件目を公開しました。
2017年07月27日
- 新たに4冊のものを追加。合計で34冊、1135ページとなりました。
2017年07月13日
- データベースの公開を開始しました。文献版1冊、ページ数162ページです。
2014年04月04日
[ニュースアーカイブ]

<http://dsr.nii.ac.jp/>

人文学における利活用の展望

1. 人文情報学でも（大規模）データを活用した研究が広まりつつある。
2. 歴史的典籍NW事業など、大規模な研究資源データの公開が進行している。
3. データの意外な利用方法などのアイデアは、分野の専門家だけでは出ない。
4. 「出会い」を生み出す場としての「アイデアソン」の事例を紹介する。



じんもんそん 2015

グループNo.6の発表

【江戸レシピの“つくれば”作成】

江戸時代の料理本に書かれているものを復元する試み。江戸時代の料理本は、正確な分量等を記載していないことも多い。だからこそ、その記述から、現代においていろいろな料理が出来上がる可能性もある。一般の人に江戸料理本をもとにした料理を実際に作ってもらい、そのレシピや作ってみた感想などのレポートをウェブにあげてもらおう。そこから、様々な料理の可能性を追求ができればと考えた。これは、古典の料理を元の通りに復元するのではなく、現代に活かす取り組みと言える。



もちろん元となった古典の資料やその翻刻文、さらには現代語訳や挿絵などもウェブにあげ、一つの料理から多彩な情報や知識が広がる仕掛けを凝らし、SNS経由で「いいね！」を押してもらおうなどの工夫も必要。また、料理は地域ごとに違いがあるものなので（たとえばお雑煮の多彩さなど）、そうした情報をウェブにあげて比較できるようにすることも面白いと考えた。更に、古典を介した上で、日本料理の情報を発信することで、

海外の人へも日本文化の良さをPRすることが出来ると考えた。

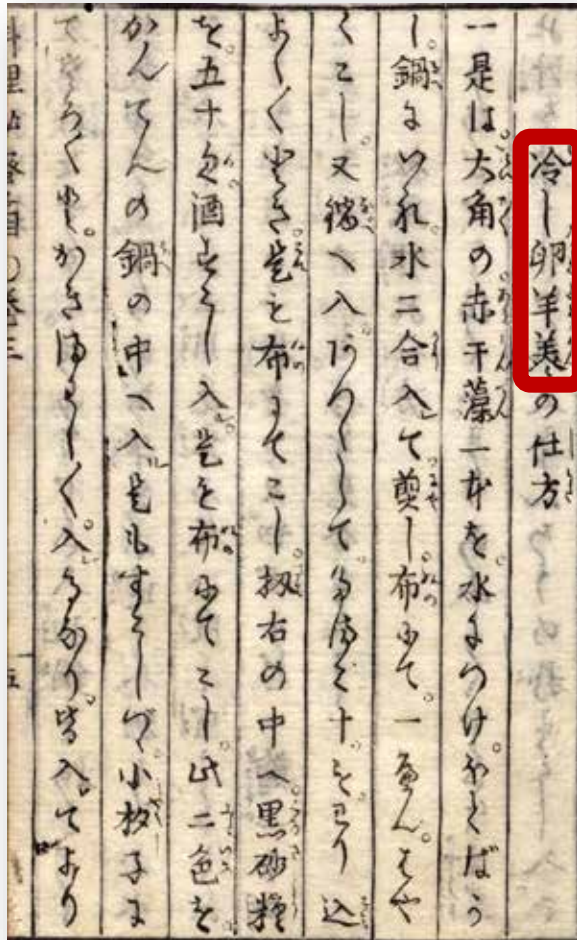
*当日は8グループ(5～6名)に分かれてディスカッションしていただきます。

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

http://www.nijl.ac.jp/pages/cijproject/report_20151218.html

江戸料理レシピデータセット

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




日本古典籍データセット
(国文研所蔵)




江戸料理レシピデータセット
(CODH制作)
日本古典籍データセット
(国文研所蔵)を翻案

アイデアの現実化

- 
1. 2015年12月 アイデアソン「じんもんそん」で料理本の存在を初めて知り、今回のアイデアを発案する。
 2. 2016年1月 クックパッドを訪問し、協力を依頼。
 3. 2016年11月24日（和食の日）江戸料理レシピデータセット公開 + クックパッドでもレシピ公開。
 4. 2017年6月 レシピを増やす。

データの作成プロセス

- 
1. 江戸の料理本をデジタル化
 2. くずし字を翻刻
 3. 翻刻を現代語訳
 4. 現代語訳をレシピ化・公開
 5. クックパッドでもレシピ公開
 6. つくれぽで個人の経験を共有



料理研究家の写真

協力：グルメクッキングファクトリー



クックパッドでもレシピ公開

クックパッドと日本家政学会 食文化研究部会が運営する「クックパッド江戸ご飯」に参加。



レシピは写真つき、現代語訳は写真なし。元文献の記述やCODHページへのリンクも加えて公開。

The screenshot shows a recipe page on Cookpad.com. The title is '江戸時代のスイーツ 甘さスツキリ冷卵羊羹'. The ingredients list includes: 卵 (Egg) 5個, 寒天(赤) (Red Agar) 1本(4g), 黒砂糖 (Black Sugar) 100g, 水 (Water) 180cc, 片栗粉 (Cassava Starch) 適量, and 酒 (Sake) 適量. The page includes a photo of the finished dish and a series of numbered steps with photos and instructions for preparation.

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

極めて大きな反応

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「つくれぼ」も受け付けていること。江戸ご飯とつくれぼというこの未来感パネい。 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	つくれぼ 0	献立 0
-----	-----------	-----------	---------

レシピを検索

7317 リツイート

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

2017/11/06

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 リツイート

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

基礎価値と付加価値

1. どのデータを誰に向けて選択するか？
 2. どこにデータを持っていくか？
 3. どのようにデータを変換するか？
- デジタルデータ公開を「基礎価値」とすれば、クックパッド公開は「付加価値」。
 - 自分が保有するコンテンツを素材に、どんな展示を組み立てれば新たな価値を見出せるか → データキュレーターのセンス。

まとめ

おわりに

1. デジタルアーカイブを、デジタル時代における人類知の創造と継承のプラットフォームと定義したい。
2. 研究という創造の現場から、人類知をリアルタイムで汲み上げ、広め、継承できるデータ専門家が必要である。
3. データの集積から新しい価値を生み出すために、新しい職能を身につけた専門家をどんどん増やしたい。

より詳しく

- **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 (CODH)**
 - <http://codh.rois.ac.jp/>
- **デジタル・シルクロード**
 - <http://dsr.nii.ac.jp/>
- **オープンサイエンス**
 - <http://agora.ex.nii.ac.jp/~kitamoto/research/open-science/>
- **Researchmap**
 - <http://researchmap.jp/kitamoto/>