

# データからアクションへ～専門的オープンデータに関するユーザとアプリケーションのモデル

人文学オープンデータ共同利用センター（CODH）  
データサイエンス共同利用基盤施設・情報・システム  
研究機構

国立情報学研究所（NII）

北本 朝展（Asanobu KITAMOTO）

<http://codh.rois.ac.jp/> , @rois\_codh



3rd GEO ●●●  
**DATA**  
PROVIDERS  
WORKSHOP

DATA PROVIDERS  
MEET USERS

FRASCATI, ITALY  
2-4 MAY 2018



UNIVERSITÉ  
DE GENÈVE



Third GEO Data Providers Workshop

Frascati, Italy, 2-4 May, 2018

[https://www.earthobservations.org/me\\_201805\\_dpw.php](https://www.earthobservations.org/me_201805_dpw.php)

# 専門的オープンデータ

- **オープンデータ**：行政が税金で構築したデータの公開（情報公開、透明性含む）。
- 研究・調査利用だけでなく、データに基づくビジネスの活性化にも期待。
- **専門的オープンデータ**：専門知識がないと処理や解釈が難しいデータの公開。
- クリーニング、品質管理、メタデータ付与など、データキュレーションが必要。

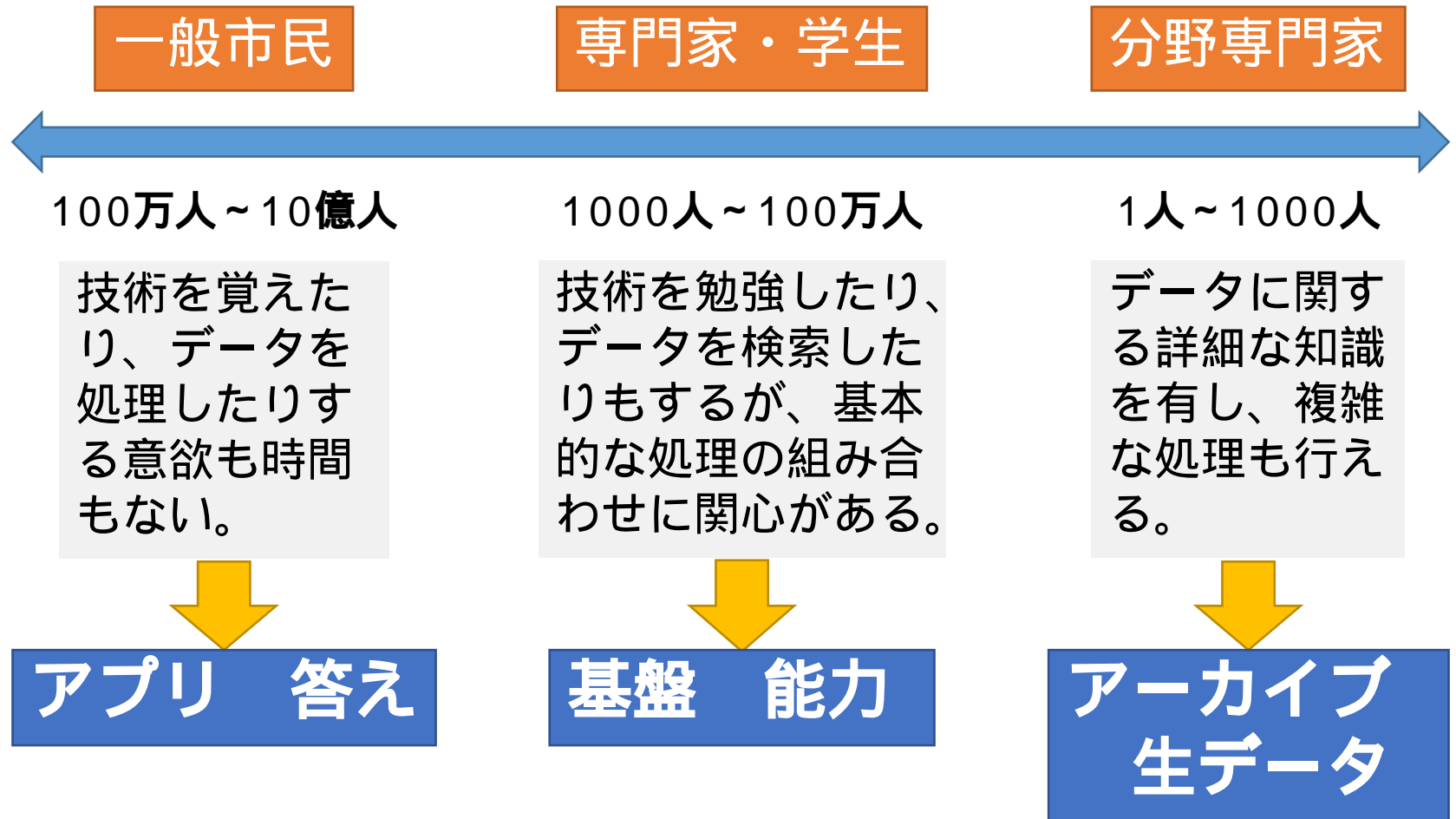
# 誰のためのデータ？

1. データを使ってもらうためにオープンにしても、実はあまり使われない？
2. それはなぜなのか？という理由を、共通理解にしておきたい。
3. **想定するユーザは誰か？そのユーザは何の目的に使うのか？**
4. **想定ユーザに合わせたアプリケーション構築のためのインフラが必要。**

# 利用視点のデータ分類

1. **生データ (raw data)** : 観測したまま、分野に特化したフォーマットで保存され、品質管理や各種補正等も適用されていないデータ。
2. **解析即用データ (analysis ready data)** : 品質管理や座標系の統一などが適用済みであり、ライブラリ等があればすぐ使えるデータ。
3. **アプリ即用データ (application ready data)** : アプリの利用パターンに合ったAPIやウェブに適したフォーマットを用い、特別な作業なしにアクセスして使えるデータ。

# ユーザとアプリのモデル



# データからアクションへ

1. データが使える = ユーザにとって Actionableなデータである。
2. 一般市民にとってのActionable = 次にとるべき行動への答えが得られる。
3. 専門家・学生にとってのActionable = (専門外の) 研究や調査にすぐ使える。
4. 分野専門家にとってのActionable = 研究の材料となるデータが得られる。




# 江戸料理レシピデータ

協力：グルメクッキングファクトリー





# データセットのワークフロー

- 
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

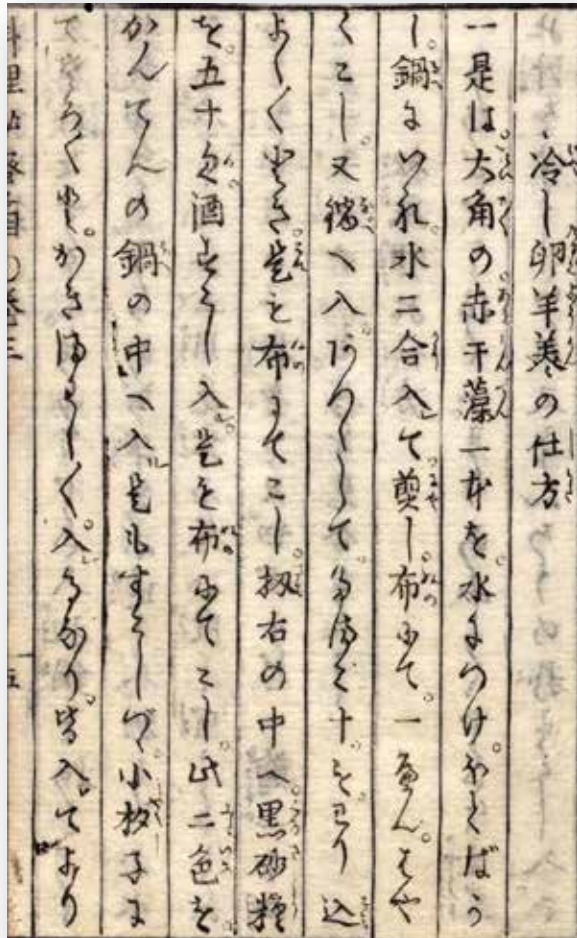
# 1. 江戸の料理本をデジタル化



日本古典籍データセット  
(国文研所蔵)

- 『**万宝料理秘密箱** (まんぼ  
うりょうりひみつばこ)』  
初版は1785 (天明5) 年。
- 「**卵百珍**」：素材としての  
卵一種で100の料理のバリ  
エーションを楽しむ。
- **専門的データ**：くずし字や  
江戸時代の文化に関する知  
識がないと活用は難しい。

## 2. くずし字を翻刻

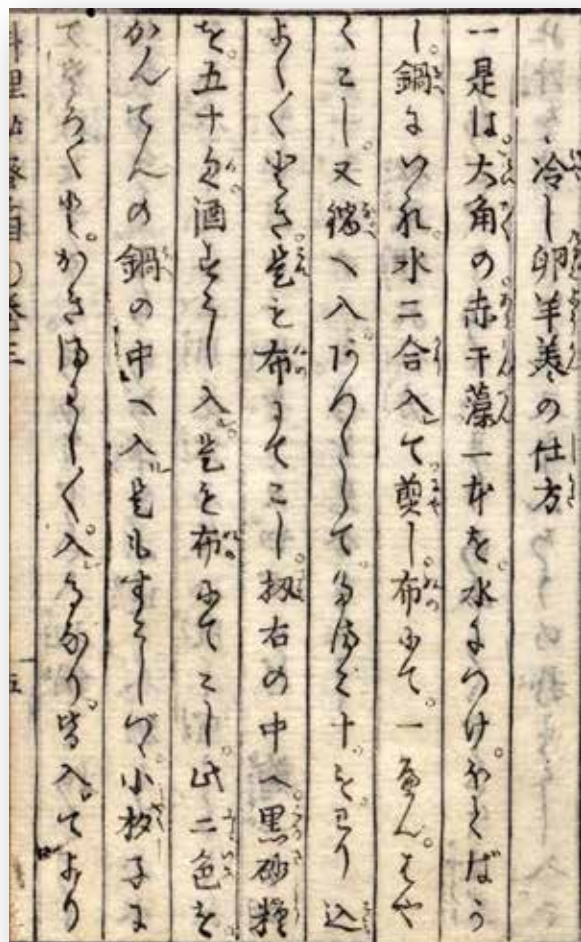


日本古典籍データセット  
(国文研所蔵)

1	是は大角の赤干藻一本を水につけほとばかし
2	鍋にいれ水二合入して煎し布にて一へんはやくこし又鍋へ入レあつくして
3	たまご十ウをわり込よくよくとき是も布にてこし
4	扱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レるなり
7	皆入レてより又葛粉をすこし水にてとき入レ
8	扱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

江戸料理レシピデータセット (CODH制作)

# 3. 翻刻を現代語訳

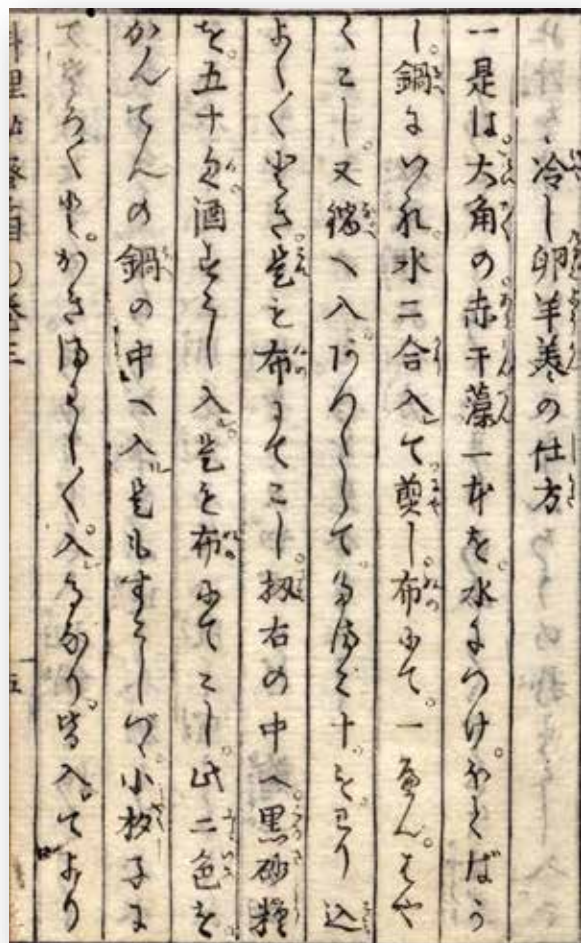


日本古典籍データセット  
(国文研所蔵)

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。

江戸料理レシピデータセット(CODH制作)

# 4. 現代語訳をレシピ化



日本古典籍データセット  
(国文研所蔵)

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。

江戸料理レシピデータセット (CODH制作)

# 文字と写真による調理手順



江戸料理レシピデータセット (CODH制作)

# 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。

クックパッド サービス一覧 お弁当におすすめのどんぶり料理 プレミアム

毎日の料理を楽しみに 253+ レシピ

検索 料理名・食材名 X 目的・用途 レン

ほうれん草 干き椎茸 親子丼 ダイコンの葉 ちつぽ

クックパッド江戸ご飯の新レシピ

### 江戸時代のスイーツ 甘さスツキリ冷卵羊羹

江戸の料理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糖の甘さがやみつきになります。

クックパッド江戸ご飯

材料 (2~4名分)

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

カロリー:276kcal/人 糖質:0.3g/人

栄養価 自動中 加工

1. 寒天を水につけてふやかします。

2. 卵をよく混ぜます。

3. 溶いた卵液を布でこします。

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、再び鍋に入れて熱めます。

7. 7の熱した寒天の中に、5の卵液を少しずつ入れます。

写真のせる 菓砂糖と酒を入れ、溶かします。

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

# 生データからアクションへ

- 生データ（江戸時代の料理本）を、調理可能なレシピに変換するワークフローに必要なスキル。
  1. 江戸時代の文字が読める。
  2. 江戸時代の単語が理解できる。
  3. 江戸時代と現代との違いを埋められる。
  4. 調理の細かい手順を推測できる。
  5. レシピの時間や分量を推測できる。
  6. 料理を美しく盛り付けできる。
  7. レシピを馴染みのサイトにアップロードできる。
- レシピから生データへ、逆方向に遡及可能とすることで、データの検証可能性も担保。



# デジタル台風

<http://agora.ex.nii.ac.jp/digital-typhoon/>

デジタル台風：台風画像と台風情報

次世代気象衛星「ひまわり8号」がいよいよ登場！ (2015-07-07)

リアルタイム台風速報

台風201724号

台風発生数

台風データベース

静止気象衛星画像 (地球)

- ひまわり画像やアメダスデータ40年分など、**データの網羅性が大きな売り**。
- 生データの提供ではなく、**処理済みデータの多角的な調査能力**に大きな価値を提供する情報基盤。
- 専門家・学生などを主な対象と想定し、利用状況約2000万PV / 年。

# 気象データの場合

一般市民

専門家・学生

分野専門家



# 専門的データの役割分担

- **分野専門家**：高品質の前処理とメタデータ付与を通して、**解析即用データ**を提供。
- **情報専門家**：解析即用データにAPIと標準形式を与え、**アプリ即用データ**を提供。
- **専門家・学生**：基盤が提供する能力を活用し、一般市民向けに**付加価値**を提供。
- **一般市民**：目的に合ったアプリから答えを取得し、適切な**アクション**を実行。

# オープンサイエンスに向けて

1. オープンにすることで何を目指すのか？ **オープン**は**ゴール**ではない。
2. オープンなエコシステムのどの部分を担当し、**どんな価値を提供**するか？
3. **専門家・学生向けに能力を提供する基盤タイプ**のサービスは、まだ少ない？
4. 一般市民向けは、**能力の提供より答えの提供を目指すアプリ**の構築が必要？

# 関連サイト

- **オープンサイエンス**
  - <http://agora.ex.nii.ac.jp/~kitamoto/research/open-science/>
- **デジタル台風**
  - <http://agora.ex.nii.ac.jp/digital-typhoon/>
- **江戸料理レシピデータセット**
  - <http://codh.rois.ac.jp/edo-cooking/>
- **人文学オープンデータ共同利用センター**
  - <http://codh.rois.ac.jp/>