

オープンサイエンスと サステナビリティ

データ駆動型サイエンスの鍵を握る
持続可能なデータプラットフォーム

北本 朝展 (KITAMOTO Asanobu)

国立情報学研究所

情報・システム研究機構

人文学オープンデータ共同利用センター (CODH)

<http://researchmap.jp/kitamoto/>

@KitamotoAsanobu

自己紹介



- 研究分野は**情報学**。画像処理や画像データベースの研究が出发点。
- データ駆動型サイエンスに発展し、**気象情報**、**地球環境情報**、**人文科学情報**などの分野で研究。
- 最近はオープンサイエンスにも関わる。

オープンサイエンスの 背景

オープンサイエンスとは？

- 「オープン」という言葉を梃子にして、サイエンス（研究）の方向を変える。
- 「よりオープンに」という方向性を共有する活動を、一語で束ねると見える世界。
- 個々の活動ごとに「オープンサイエンス」の意味は異なり、単一の定義は困難。
- 大同団結？同床異夢？個々の活動を超える新しい目標を示せるかが問われる。

オープンサイエンスへの収束

透明性

オープンアクセス

共有

オープンピアレビュー

研究の再現性・
透明性・研究
データ保存

オープンデータ

研究データ
データ出版
データリポジトリ

市民科学・クラウド
ファンディング

オープ
ンサイ
エンス

コラボレーション・オー
プンイノベーション

超学際研究

参加

協働

メタ研究＝研究（システム）に関する研究



オープンサイエンス革命
マイケル・ニールセン (著)
紀伊國屋書店, 2013
右の引用はp.340.

- インターネットの登場で社会は変わった。なぜ科学は変わらないのか？
- 科学はオープン化を妨げる独自の問題に直面。
- 科学者の報酬は論文の出版によって得られるものという問題。

現在のオープンサイエンスは、ニールセンの「夢」を実現する方向に進んでいるわけでは必ずしもない。

第5期科学技術基本計画

③ オープンサイエンスの推進

オープンサイエンスとは、オープンアクセスと研究データのオープン化（オープンデータ）を含む概念である。オープンアクセスが進むことにより、学界、産業界、市民等あらゆるユーザーが研究成果を広く利用可能となり、その結果、研究者の所属機関、専門分野、国境を越えた新たな協働による知の創出を加速し、新たな価値を生み出していくことが可能となる。また、オープンデータが進むことで、社会に対する研究プロセスの透明化や研究成果の幅広い活用が図られ、また、こうした協働に市民の参画や国際交流を促す効果も見込まれる。さらに、研究の基礎データを市民が提供する、観察者として研究プロジェクトに参画するなどの新たな研究方策としても関心が高まりつつあり、市民参画型のサイエンス（シチズンサイエンス）が拡大する兆しにある。近年、こうしたオープンサイエンスの概念が世界的に急速な広がりを見せており、オープンイノベーションの重要な基盤としても注目されている。

こうした潮流を踏まえ、国は、資金配分機関、大学等の研究機関、研究者等の関係者と連携し、オープンサイエンスの推進体制を構築する。公的資金による研究成果については、その利活用を可能な限り拡大することを、我が国のオープンサイエンス推進の基本姿勢とする。その他の研究成果としての研究二次データについても、分野により研究データの保存と共有方法が異なることを念頭に置いた上で可能な範囲で公開する。

ただし、研究成果のうち、国家安全保障等に係るデータ、商業目的で収集されたデータなどは公開適用対象外とする。また、データへのアクセスやデータの利用には、個人のプライバシー保護、財産的価値のある成果物の保護の観点から制限事項を設ける。な

32

お、研究分野によって研究データの保存と共有の方法に違いがあることを認識するとともに、国益等を意識したオープン・アンド・クローズ戦略及び知的財産の実施等に留意することが重要である。

また、国は、科学研究活動の効率化と生産性の向上を目指し、オープンサイエンスの推進のルールに基づき、適切な国際連携により、研究成果・データを共有するプラットフォームを構築する。

- 平成28年度～平成32年度の基本計画。
- **研究データの公開や活用、透明化、プラットフォーム**などが記述の中心。
- オープンアクセスや市民科学、オープンイノベーションなどにも触れる。

<http://www8.cao.go.jp/cstp/kihonkeikaku/5honbun.pdf>

「オープン」の3つの側面

1. 他者が使える（再利用）

- オープンデータやオープンアクセスなど。外部の人が研究結果を自分の目的に再利用できる。

2. 他者が検証できる（透明性）

- オープンガバメントや研究再現性など。外部の人がエビデンスを検証し、正当性を判断できる。

3. 他者を受け入れる（参加）

- オープンイノベーションや市民科学など。外部の人を招き入れ、共に価値を生み出す。

データ駆動型サイエンス

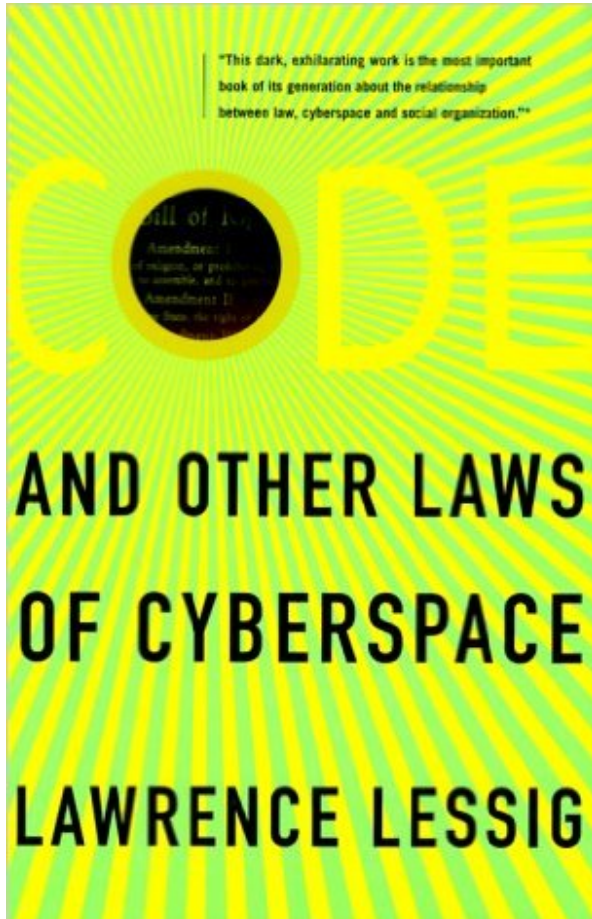
1. **第四の科学**：データが核となるサイエンスへ。
2. **オープンなデータの必要性**：良質のデータがないと、研究やイノベーションが進まない。
3. **情報基盤**：データの生成、流通、分析、共有を支える新しい学術情報基盤が必要。
4. **インセンティブと人材育成**：持続的な活動にはインセンティブと人材育成の見直しが必要。
5. **超学際**：社会の課題解決に貢献するには、多様なステークホルダーが共創することが必要。

PREPARING FOR THE FUTURE OF ARTIFICIAL INTELLIGENCE

- Recommendation 1: **Private and public institutions are encouraged to examine whether and how they can responsibly leverage AI and machine learning in ways that will benefit society.** Social justice and public policy institutions that do not typically engage with advanced technologies and data science in their work should consider partnerships with AI researchers and practitioners that can help apply AI tactics to the broad social problems these institutions already address in other ways.
- Recommendation 2: **Federal agencies should prioritize open training data and open data standards in AI.** The government should emphasize the release of datasets that enable the use of AI to address social challenges. Potential steps may include developing an “Open Data for AI” initiative with the objective of releasing a significant number of government data sets to accelerate AI research and galvanize the use of open data standards and best practices across government, academia, and the private sector.

<https://www.whitehouse.gov/blog/2016/10/12/administrations-report-future-artificial-intelligence>

制度を分析する4つの視点



- Lawrence Lessig (Founder of Creative Commons), *Code: And Other Laws of Cyber Space* (first edition 1999)
- **法** = しなければならない
- **規範** = すべきである
- **市場** = した方が利益がある
- **アーキテクチャ** = せざるを得ない

「法」によるオープン化



NSF Data Management Plan

- **資金提供機関**が「データ管理計画」などを義務化。
- **研究評価**でも、オープン化の進展を指標に含める。
- 上からの押し付けには**副作用**も大きく、**見極め**が必要。

「規範」によるオープン化



- **オープンな文化**：データ共有が不可欠な分野もある。
- **世代の差**：若い世代では共有文化の経験がより強い。
- **文化の差**：異なる文化圏に対する説得力が弱い。

<https://www.icsu-wds.org/>

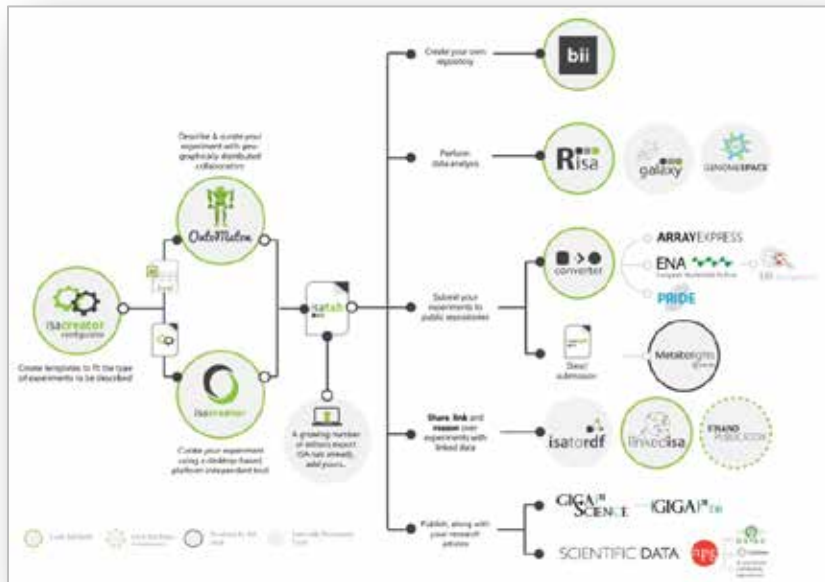
「市場」によるオープン化



- **報酬への期待**：研究成果をオープン化すると、引用も増加する [要出典]。
- **損失への不安**：他者に成果を横取りされるんじゃないの？報酬は労力に見合うの？

Scientific Data (Nature publishing group)

「アーキテクチャ」による オープン化



<http://www.isa-tools.org/software-suite/>

- **選択と誘導**：プラットフォームを選ぶと、可視／不可視なルールによって誘導される。
- **苦痛の軽減**：オープン化は大変だから、有償サービスにお任せ？
- **ベンダーロックイン**：良くも悪くも企業のビジネスチャンス。

オープンアクセスの顛末

オープンアクセス前

- 論文の電子化をきっかけに購読料が高騰。
- 市民による論文アクセスが遮断される。
- グリーン（リポジトリ）、ゴールド（著者支払）で、論文アクセスをオープン化。

オープンアクセス後

- 出版社は方針を変更し、著者支払いで十分な利益を得る。
- 何でも通す出版社のほうが利益が上がる。
- 詐欺的な出版社は、論文への永続的アクセスを放棄。

研究データとデータ駆 動型サイエンス

3種類の研究データ

研究資源データ

研究の入力となるデータ。観測データや評価用データセットなど。**再利用のためのオープン化**が求められる。

論文付属データ

研究の出力となるデータ。論文のエビデンスとなるデータなど。**透明性のためのオープン化**が求められる。

研究過程データ

研究の入力と出力の間で生み出されるデータ。日々の研究活動のエビデンスとなるデータなど。**研究不正防止（透明性）のためのクローズドな長期保存**が求められる。

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

ホーム Earth デジタル台風 English

お知らせ : ご意見 : 壁紙 : 動画 : キッズ : RSS / Atom / Media RSS : OpenSearch
ケータイ版 : PDA版 : Google Earth版 : iPod版 : Twitter版 : Facebook版 : デジタルフォトフレーム版 : アプリケーション
ソーシャル台風 : 台風前線 : 台風画像 : 台風空想 : アイフーン : ツイフーン : 台風なう! : 台風メモリーズ : ふってきったー

リアルタイム台風情報

台風201323号 台風201324号

MTS213100706 MTS213100706

201323 (WNP) 201324 (WNP)

{N27.0, E118.0} {N26.7, E128.7}

1000 hPa / 0 kt 935 hPa / 100 kt

台風発生数
現在 = 24個 (台風経路図) (台風進路予想図)
平年 = 19.8個 (1951-2010)

最新台風情報

- 台風ニュース・ウェブログ
 - 2013年台風23号 (フィートウ [FITOW]) (2013年10月05日)
 - 2013年台風24号 (ダナス [DANAS]) (2013年10月07日)
- 台風ニュース・トピックス
- 地域情報ポータル
- アメラス (アメダス・ランキング)
 - 降水量マップ : 風向・風速マップ : 最新画像 : イベント検索
- レーダー : Google Maps版 : 最新画像
- 災害情報データベース
- その他のニュース

台風データベース

静止気象衛星画像 (地球)

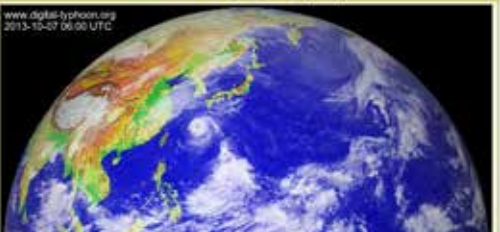
1. メタデータによる検索

- 日時・シーズンで検索
- 名前・番号で検索
- 地図で検索
- 地名 (緯度・経度) で検索
- 最低中心気圧で検索
- 最大風速で検索
- 日付で検索
- 活動カレンダー (存在個数)
- 台風タイムライン

北西太平洋 : 南西太平洋

2013年10月7日15時 (JST)

www.digital-typhoon.org
2013-10-07 06:00 UTC



- 台風に関する「**あらゆる情報**」を集約。
- **最新状況の把握と過去アーカイブの検索**がシームレスに接続。
- 利用実績は**年間約2000万ページビュー**。
- **研究者のみならず、一般市民も多数利用。**

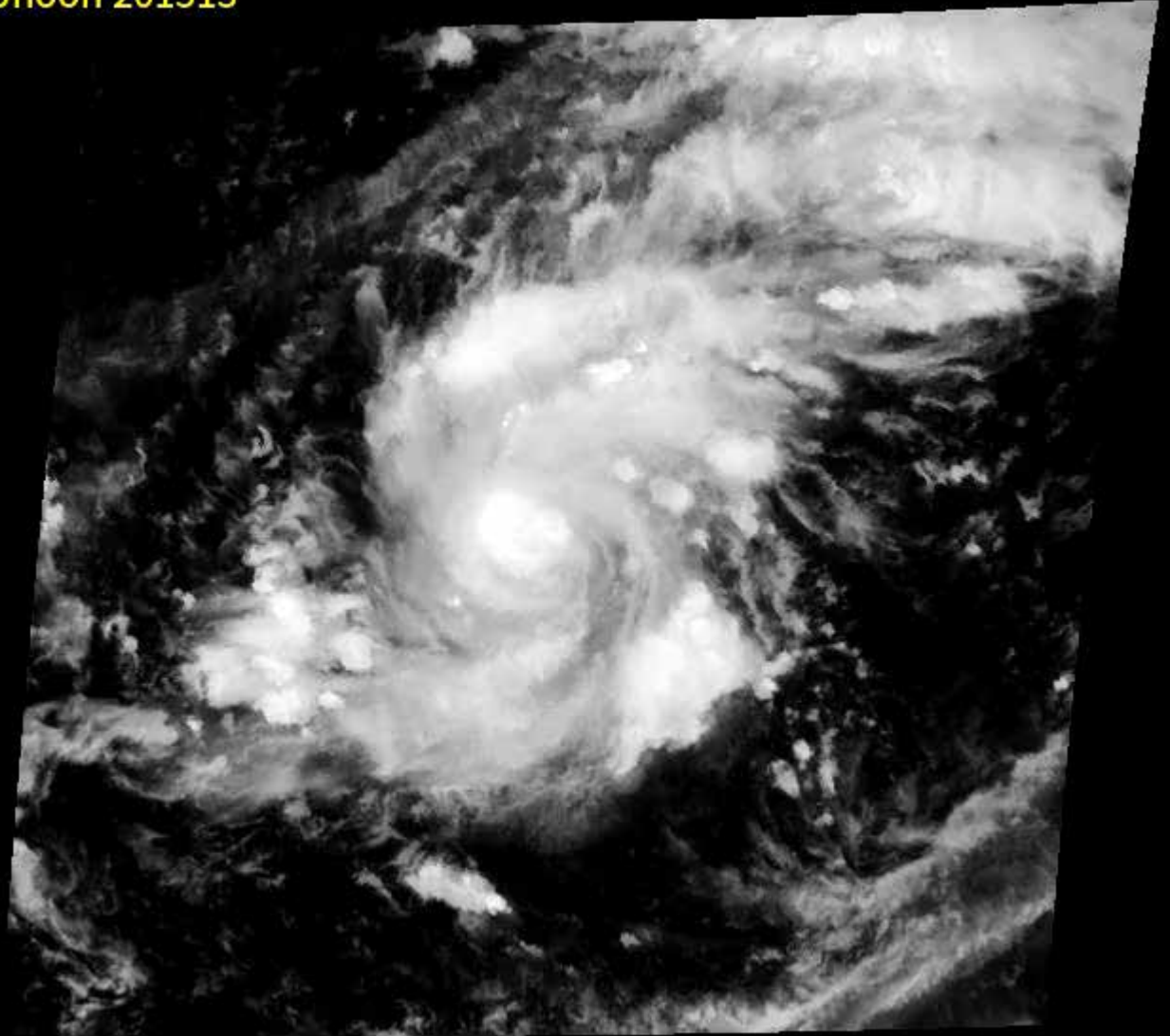
長期データアーカイブ

気象衛星画像（ひまわり）	1978-
台風画像コレクション	1978-
台風ベストトラック	1951-
アメダスデータ	1976-
合成レーダー（レーダーアメダス）	2004-（1998-）
気象庁GPV GSM / MSM	2002-
天気図	1883-
オンラインニュース記事	2003-
ソーシャルメディア（ブログ／ツイッター）	2004- / 2009-
気象災害報告	1971-
気象庁防災情報XML	2012-

ひまわり 8号に よる高 頻度観 測動画

2015-07-30 01:25:00 UTC
Typhoon 201513

www.digital-typhoon.org



Himawari-8 [B13]

NII/NICT

市民参加型スマホアプリ

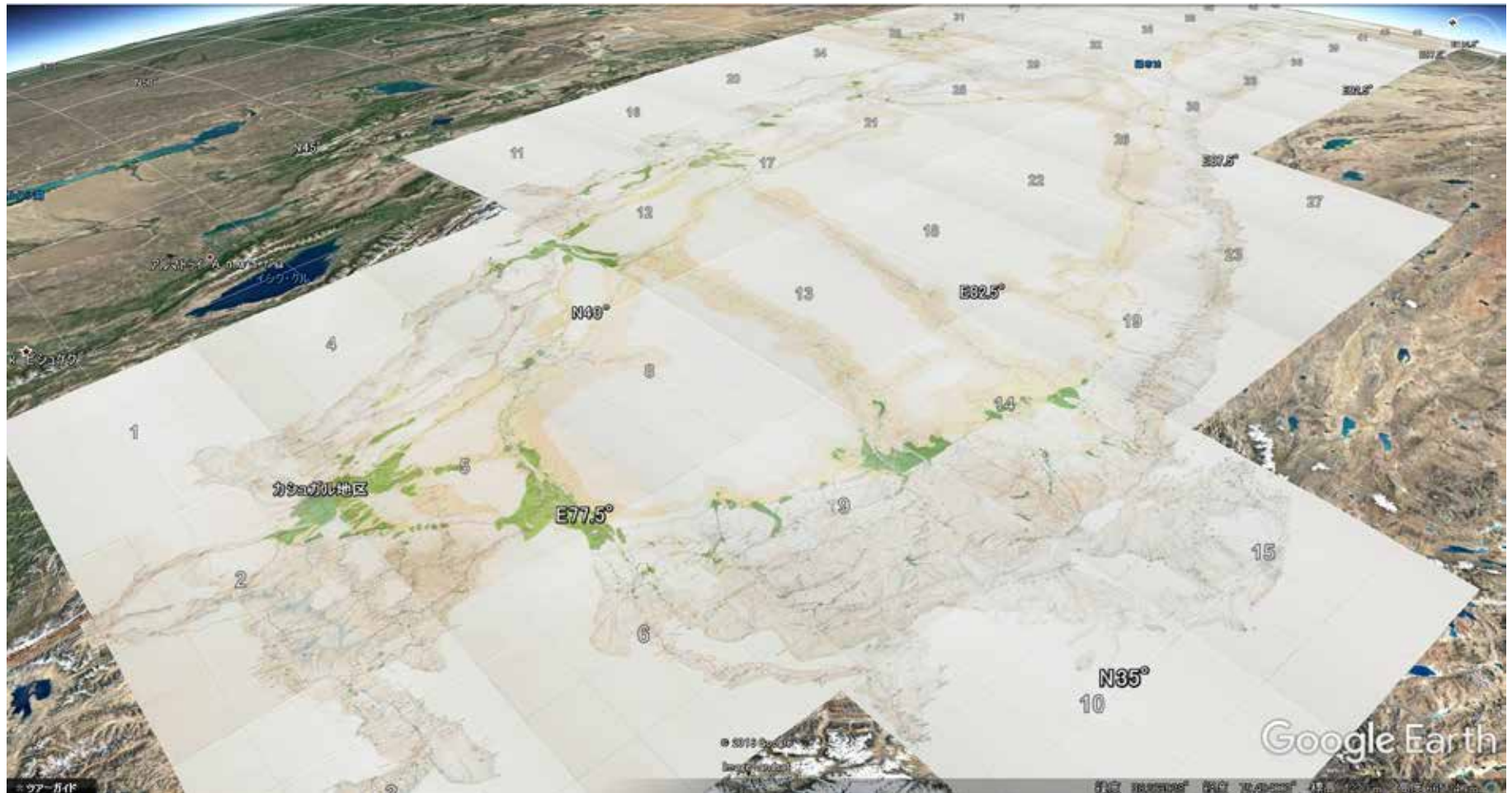


1. スマホで積雪状況を継続的に写真観測するアプリ「雪ログ」。
2. 札幌の気象キャスターや放送局と連携し、昨冬テスト観測を実施。
3. 今年度もテストする予定で頑張ってます。



デジタル・シルクロード

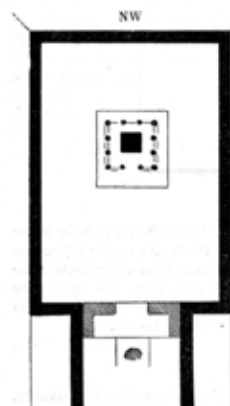
<http://dsr.nii.ac.jp/>



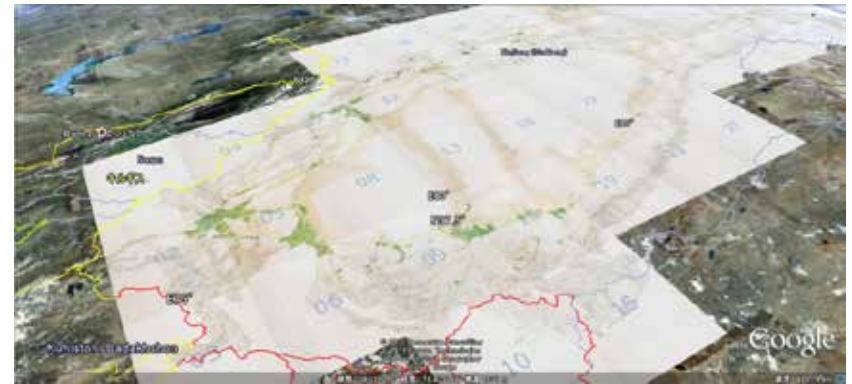
多様な人文学データ

テキスト

地図



die obere sich wie eine in eine niedrigere 3,10 m tiefe Plattform eingesaute Bank darstellt (auf der Skizze schraffiert) und die Mitte offen läßt. Vor dieser großen Unterstufe liegt der Rest eines mächtigen Sockels, in welchem ein tiefes Loch sich zeigt: hier hat also wohl eine große Statue oder eine Fahne gestanden. 12 m nach innen zu vom S.-Rand der Plattform des Hauptbaues, 5,50 m von den Seitenmauern und 7 m vor der Rückmauer, erhebt sich eine niedrige, 8 m ins Geviert betragende Stufe, auf deren Mitte ein jetzt zerstörter, 2 m großer, viereckiger Sockel steht; um diesen Sockel geht ein Gang herum, vorne und an den Seiten je 1,50 m breit, hinten aber nur 90 cm breit. Dieser Umgang ist nach außen von einer Mauer umgeben, welche durch zwölf kleine Stäben in kleine Abteile geteilt ist, von denen der mittlere der Frontseite den Eingang bildet. Auf der Rückseite ist dies aus zwei Eck- und zwei Mittelsäulen bestehende System sehr zerstört. Vor den sechs Interkolumnien der Seiten und den zwei Interkolumnien neben dem Eingang sind je noch Sockel für Statuen erhalten: auch mancherlei dekoratives Bei-



写真

地名辞書

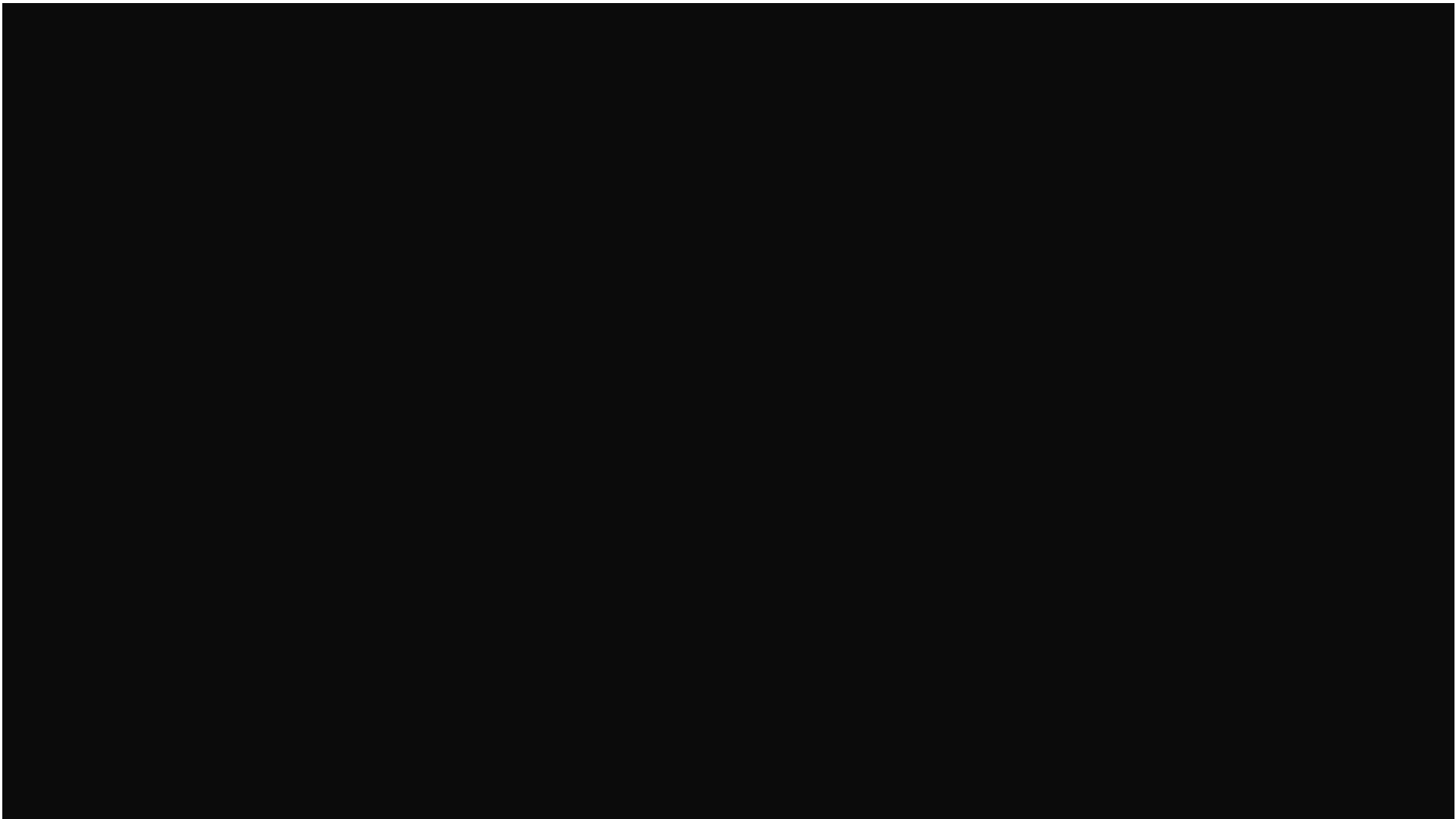


Y
1904(Le Coq, 1913, Tafel. 70, I)

Abāb-langar, habit., 14. B. 3.
Abād (of Ak-su), market-town, 12. A. 3.
Abād (of Kara-yulghun), vill., 12. B. 1.
Abād (of Karghalik), vill., 5. C. 4.
Abād (of Kāshgar), vill., 5. A. 2.
Abād (of Turfān), vill., 28. C. 3.
Abād (of Yārkand), vill., 5. C. 2.
Abād-jilga, valley, 12. B. 2.
Abdal, vill., 30. B. 2.
Abdalkash-mazār, shrine, 14. C. 3.
Abdul-ghafūr-langar, loc., 10. C. 1.
Abdul-rahmān-jilga, valley, 9. A. 4.
Abshak-bēl, Pass, 2. B. 1.
Aeh-tāgh, hill and vill., 7. C. 2.
Aeha-dong (of Chizghān), hill, 19. C. 3.
Aeha-dong (of Yārkand R.), loc., 7. D. 4.
Aeha-kuduk, loc., 7. D. 4.
Aeha-shipang, loc., 22. D. 4.
Aehak-aghzi, loc., 5. A. 4.
Aehal (of Ak-su), vill., 12. A. 3.
Aehal (on Charchak R.), loc., 21. C. 2.

Aehchik-bulak (of Turfān), spring, 28. B. 4.
Aehchik-bulak (of Yai-döbe), spring, 4. C. 4.
Aehchik-daryā, river, 21. A. 2.
Aehchik-dawān, pass, 9. B. 3.
Aehchik-jilga (of Duwa), valley, 9. B. 3.
Aehchik-jilga (of Kara-tāsh), valley, 2. D. 3.
Aehchik-jilga (of Khotan), valley, 9. C. 3.
Aehchik-jilga (of Sampula), valley, 14. A. 3.
Aehchik-jilga (of Tawak-kēl), loc., 14. A. 1.
Aehchik-köl, lake, 15. D. 1.
Aehchik-kuduk (of Kapa), well, 23. A. 1.
Aehchik-kuduk (of Kuruk-tāgh), well, 28. C. 4.
Aehchik-kuduk (of Marāl-bāshi), well, 5. D. 2.
Aehchik-otan, loc., 7. C. 2.
Aehchik-su, loc., 31. A. 4.
Aehchik-tügemen, loc., 5. D. 2.
Aehi-tāgh, hill, 32. B. 1.
Aehik-aghzi, loc., 9. D. 3.
Aehma (of Hanguya), vill., 14. A. 2.

Memory Hunt in Aceh (Collaboration with CIAS, Kyoto University, Prof. Nishi and Prof. Yamamoto, Video by Mr. Mahruza Murdani)



人文学オープンデータ共同利 用センター（CODH）

- 情報・システム研究機構（国立情報学研究所の上位組織）が2016年4月1日に準備室を開設。2017年4月1日にセンター発足予定。
- **情報学＋統計学によるデータ駆動型人文学**で、人文学コミュニティの研究に新しい道を開く。
 1. 情報学の技術を用いて人文学の研究を行う。
 2. 人文学のデータを用いて情報学の研究を行う。
- **国文学研究資料館「歴史的典籍NW事業」**が、オープンデータの一つの中心となる。

国文研歴史的典籍NW事業

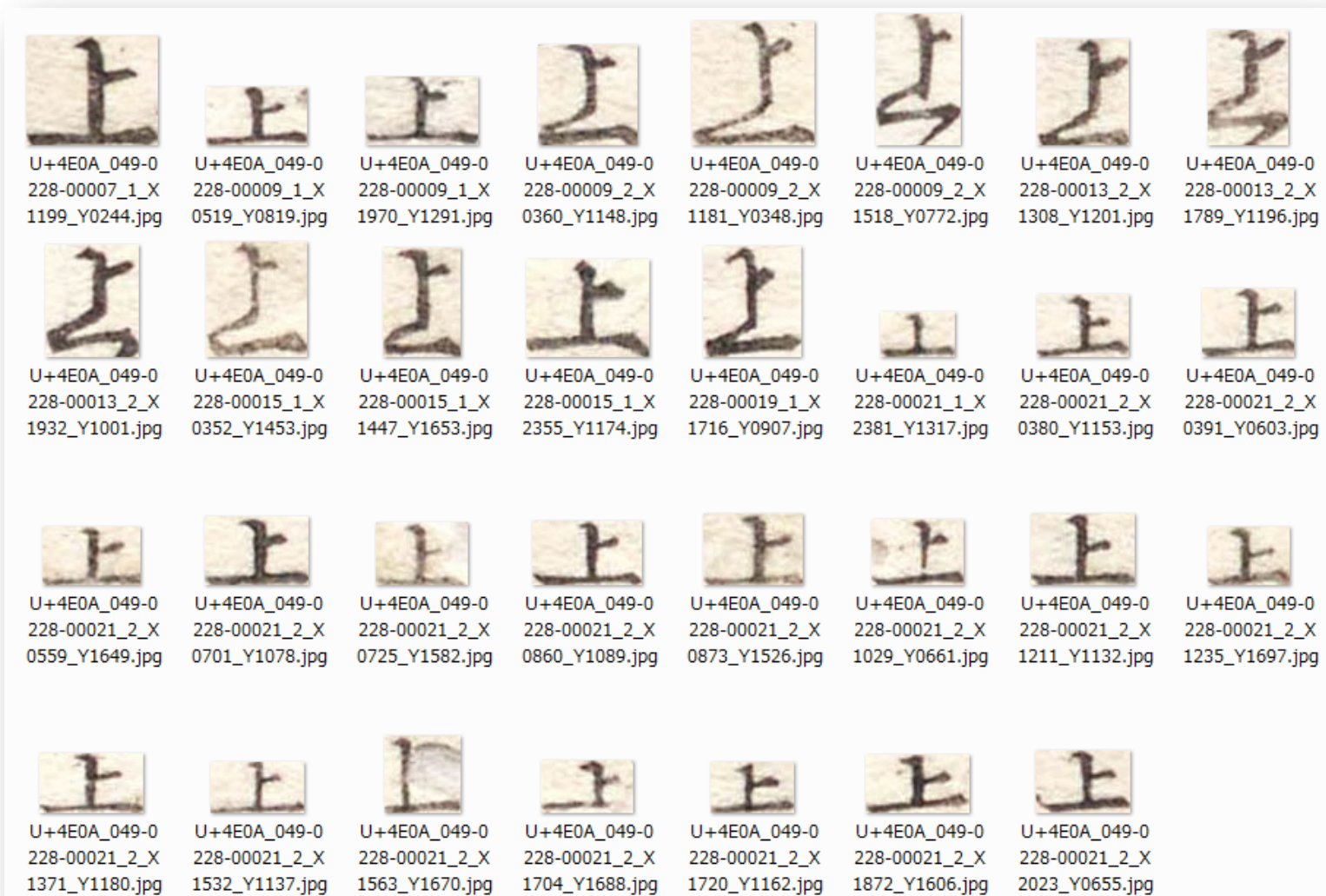
- 10年間で約30万冊の古典籍をデジタル化し、公開する（当初予算通りなら）。
- 国文研所有の古典籍は、350冊をCC- BY-SAで公開済み。近日中に700冊（158,455画像）に拡大。
- くずし字で書かれた版本を、一般人は読めない。翻刻（テキスト化）や現代語訳もまだ少ない。
- 人間が学習する＝モバイルアプリ（KuLA）
- 機械が学習する＝広義の文字認識（ディープラーニング等）

日本古典籍データセット

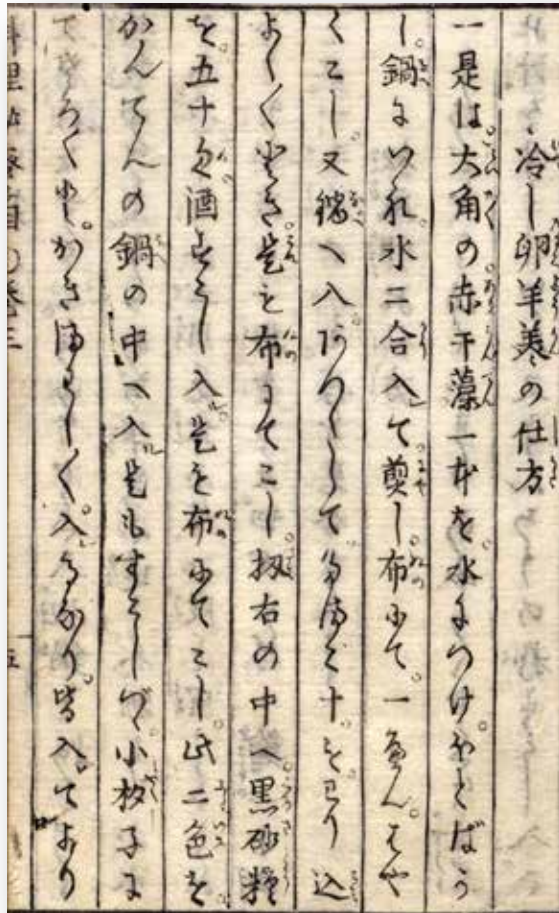
くずし字で書かれた文字を自動的に読めれば、人文学研究は大きく進展するだろう。

『画本虫撰』 日本古典籍データセット（国文研蔵）

字形データセット



江戸料理レシピデータセット



『万宝料理秘密箱』日本古典
籍データセット（国文研蔵）

卵：10個
赤寒天（大角一
本）：8g
葛粉
黒砂糖：約200g
酒：少し
水：360cc



研究者と市民
が江戸レシピ
を共に探求



人文学とオープンサイエンス

- **人文学者向けオープンデータ**：誰でも公平にデータを利用でき、アクセスの権力化を防ぐ。
- **情報学者向けオープンデータ**：機械学習データセットを用意し、技術を競う土俵を作る。
- **市民向けオープンデータ**：日常に密接したデータセットを用意し、多様な展開を狙う。
- **超学際的コラボレーション**：研究分野やセクターを越えた協働で、新しい見方を生み出す。
- 人文学は最後のフロンティア？

研究データ人材・データライブラリアン

研究データ人材・基盤

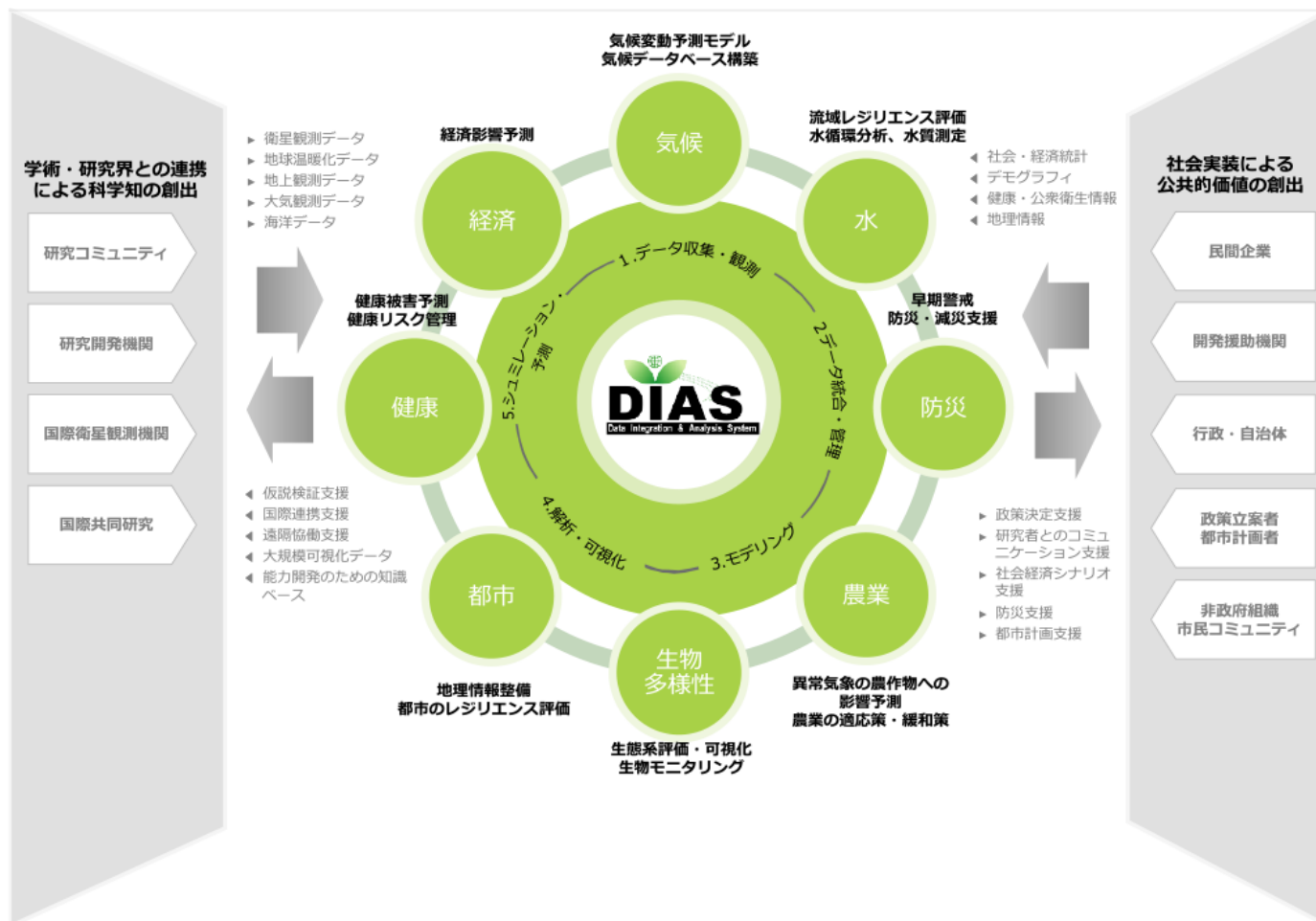
- **公的研究資金の公開義務化**：研究データ公開に関わる作業を、誰かがやらねばならない。
- **研究者の仕事？**：片手間に作業するだけだと、進まないしクオリティも上がらない。
- **研究データ専門職**：データインフラに関する多岐の仕事は、専門職の人に任せたい。
- **どんな人材？**：データ整理に関する仕事は、ライブラリアンと似ている？
- **データ基盤**：データ利活用には、オープンサイエンスを支える新しい学術基盤も必要。

研究データ人材

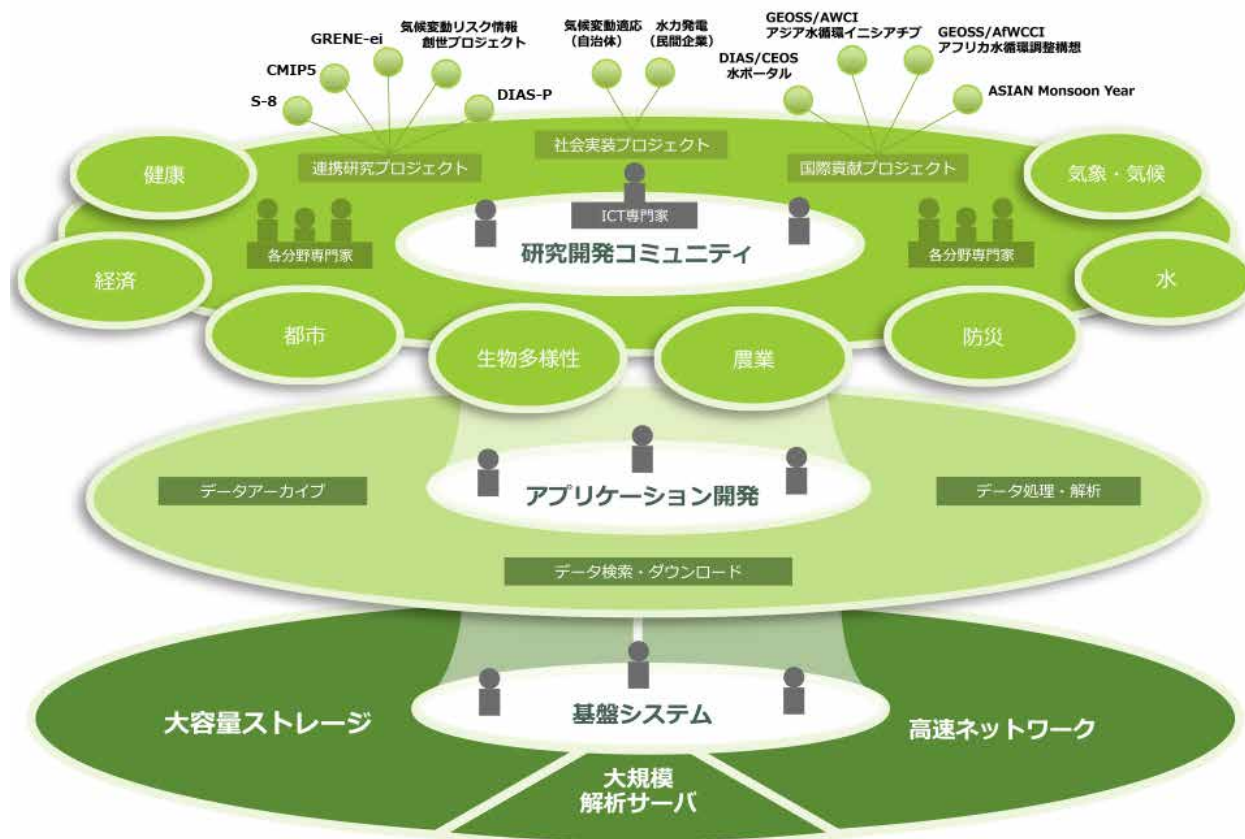
- **データサイエンティスト**：データ分析を科学的に行える。研究開発よりビジネス向き？
- **データライブラリアン**：図書の整理からデータの整理へ。
- **データキュレーター 1**：美術品の展示からデータの展示へ。
- **データキュレーター 2**：美術品の解釈からデータの解釈へ。
- 名前は同じだが、既存の職業の延長なのか？

DIAS : データ統合解析システム

<http://www.diasjp.net/>



DIAS = データ + コミュニティ



コミュニティ基盤

データ基盤

<http://www.diasjp.net/about/system/>

DIASのデータマネジメント

1. **ドメインの専門家にヒアリング**しながら、メタデータのフォーマットを決める。
2. **メタデータの国際標準**を参照しながら、相互運用性を確保する。
3. **メタデータシステム（管理・検索）**を構築する（または仕様書を書いて外注する）。
4. **メタデータの内容をチェック**し、システム全体として一貫性を保つように改善する。
5. **メタデータ入力を推進**するため、普及活動を開催しながらコミュニティと付き合う。

メタデータ形式の検討

- ⊕ データ提供者が作成しているデータセット紹介のドキュメントを分析
- ⊕ ドキュメントに含まれているメタデータの収集
- ⊕ ISO19115（地理空間メタデータ標準）の項目との対応づけ
- ⊕ 検索に役立つキーワードについての検討

試行錯誤

メタデータ ワーキンググループ

データ提供者

情報システム
研究者

当初検討
サンプル
データ
A1 - A6



データ提供者作成

データセットドキュメンテーション

入力項目確認

メタデータ標準

キーワード

インタフェース検討

メタデータ
入力ツール



アプリケーション



データ分類マトリクス

検索と表示

ドキュメンテーション中の章立て、内容进行分析し、メタデータ項目との対応付けを行う。データセットごとに内容が異なるため当初のサンプルデータに対して詳細な検討を行う。

2016/11/04

メタデータ入力ツールは必須メタデータ項目を含めドキュメントの章立てに対応

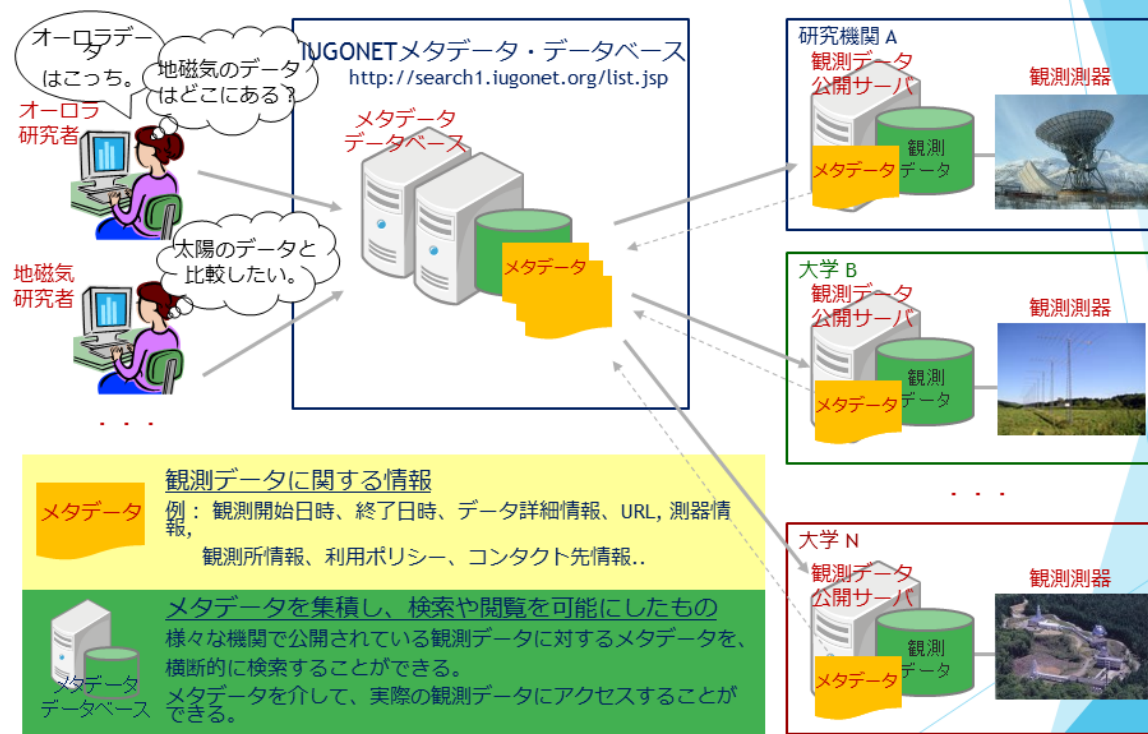
メタデータ入力キャンプ



- **メタデータの登録が滞っているデータ提供者に対し、入力方法・入力内容などに関する疑問に答え、入力を促進する場を設定。**
- **2014年7月、11月、2015年2月に実施。**

IUGONETの事例

IUGONET紹介



サイエンス成果の創出に向けたデータアーカイブと解析環境の融合. 梅村直生ほか
第3回オープンサイエンスデータ推進ワークショップ@京大を一部改変

11

南山 泰之, 超高層大気観測データのメタデータ作成実験経過報告, 第2回 SPARC Japan セミナー2016

IUGONETの事例

概要

▶ 図書館職員によるIUGONETメタデータ作成を通じて、

1. 研究者へのインセンティブ

- ▶ 研究者、特にメタデータ作成業務従事者の負担軽減
- ▶ 図書館が持つ流通面での知見の提供

2. 図書館による新規サービス展開

- ▶ 研究者との協働による、より直接的な研究支援

の可能性を探る。

南山 泰之, 超高層大気観測データのメタデータ作成実験経過報告, 第2回 SPARC Japan セミナー2016

IUGONETの事例

パイロット開始 (6月～)

図書職員とのコラボレーション

No.	難易度	内容	所感 (図書)	所感 (IUG)
1	低	既存のメタデータ(XML)を機械的に修正する	◎ 容易	◎ OK
2	中-	既存のメタデータ(XML)を新しいスキーマバージョンに則したものに变更する	○ 出来た	○ OK
3	中	PIにインタビューしながら新しくサイエンスメタデータを作成する	○ 出来た	○ OK
4	中	外部サーバを使って単純なデータ公開・更新を行う	(予定)	
5	中+	目前サーバを使って単純なデータ公開・更新を行う	(予定)	
6	高	IUGONETのようなサービスのデータ公開部分を運用する	(予定)	
7	高+	IUGONETのようなサービスのデータ登録部分を運用する	(予定)	

サイエンス成果の創出に向けたデータアーカイブと解析環境の融合. 梅村直生ほか
第3回オープンサイエンスデータ推進ワークショップ@京大を一部改変

南山 泰之, 超高層大気観測データのメタデータ作成実験経過報告, 第2回 SPARC Japan セミナー2016

IUGONETの事例

ここまでの検証

- ▶ 既存の情報源をメタデータ化する作業は、図書館の目録作成業務の延長
 - ▶ 標準化されたスキーマや入力ルールが定められていれば、
分野を問わず対応可能
 - ▶ メタデータ作成業務を集約でき、機関としてのコストが下がる
 - ▶ もっとも、データ取得者による確認作業は省略できない
- ▶ 人的な持続可能性の問題
 - ▶ 作業は慣れや労力を要するが、マニュアル作成によって
標準化・引き継ぎが可能
 - ▶ メタデータ作成に要する時間は、図書や地図とあまり変わらない
 - ▶ 図書館のネットワークで共有することで、全体の知見になる

コード	一般資料種別	コード	特定資料種別
なし	下記のいずれでもないもの	なし	下記のいずれでもないもの
		a	大活字本 (largoprint)
		b	触知資料 (tactil)
		c	天体図 (celestial globe)
		d	惑星儀 / 月球儀 (planetary or lunar globe)
		e	地球儀 (terrestrial globe)
		f	ダイアグラム (diagram)
		g	地図 (map)
		h	断面図 (profile)
		i	模型 (model)
		j	リモートセンシング画像 (remote-sensing image)
		k	断面図 (section)
		l	観測図 (view)
		m	上記のいずれでもないもの
b	文字資料 (点字) (brail)		
		a	スコア (full score)
		b	ミニチュアスコア (miniature or study size)
		c	鍵盤楽器伴奏譜 (accompaniment reduced for keyboard)
		d	ヴォイススコア (voice score)
		e	コンダクトスコア / ピアノコンダクタースコア (condensed score or piano-conductor score)
		f	クローススコア (close score)
		g	邦楽譜
		h	複合形態の楽譜 (multiple score format)
		i	不明 (unknown)
		j	上記のいずれでもないもの (other than score format)
d	文字資料 (書写資料) (manuscript text)		GMD「なし」に対応するSMDのうち適切なコードを用いる
e	地図 (書写資料) (manuscript map)		GMD「a」に対応するSMDのうち適切なコードを用いる
f	楽譜 (書写資料) (manuscript music)		GMD「c」に対応するSMDのうち適切なコードを用いる
		h	フィルムストリップ (カートリッジ) (filmstrip cartridge)
		i	フィルムストリップ (filmstrip)
		j	フィルムストリップ (その他) (other filmstrip type)
g	静止画像 (投影) (projected image)		

目録システムコーディングマニュアル

http://catdoc.nii.ac.jp/MAN2/CM/furoku1_1.html

南山 泰之, 超高層大気観測データのメタデータ作成実験経過報告, 第2回 SPARC Japan セミナー2016

求められるスキル@2016年

1. メタデータ規格や語彙に関する国際標準との相互運用性を踏まえた知識。
2. メタデータの入力や運用に関して研究コミュニティと交流する能力。
3. メタデータに関するシステムを実装または外注する能力。
4. もっと欲張るなら：**データキュレーター**=データ利活用全般（広報や価値プレゼンなど）、**データサイエンティスト**=データ分析やアプリ化などでもできると嬉しいが。。

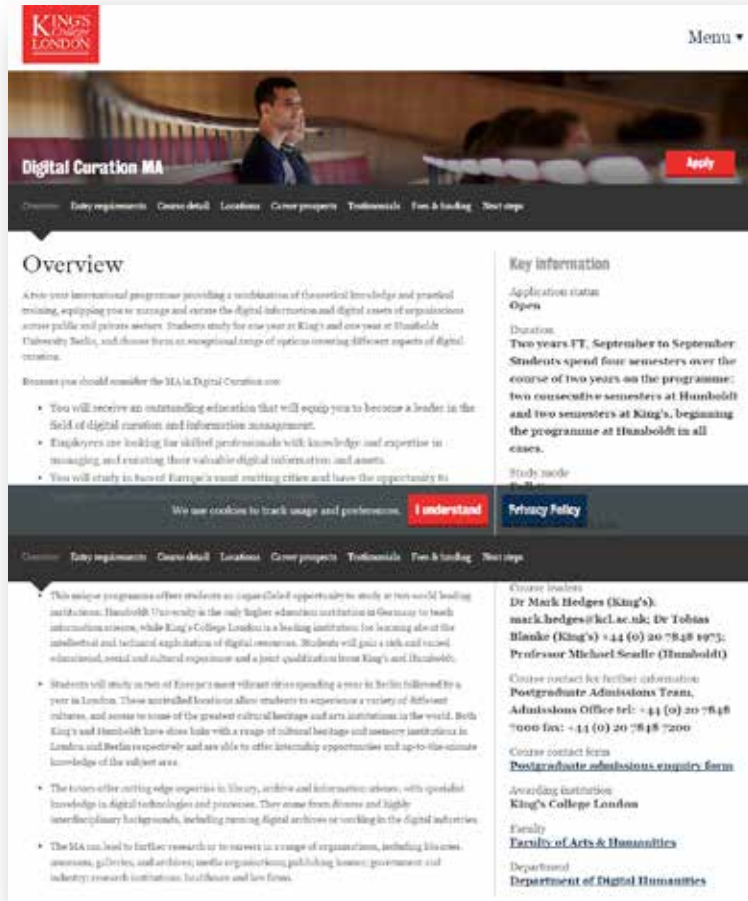
役割・スキル・価値観

- 「データライブラリアン」は「ライブラリアン」の延長にあるのか？あるいは、全く別のスキルを持った人がスライドするのか？
- データの世界における社会的役割は、「ライブラリアン」や「キュレーター」と言える。
- 社会的役割が同じなので、価値観や目指すべき方向性も似ている。
- 具体的な対象が異なるため、今までと同じスキルだけでは対処できない。

求められるスキル@2026年？

1. **人工知能がすべてを代替する？** 代替可能なスキルを学ぶのは、人生戦略上まずい。。
2. **データ整理はアクセス手段が目的？** ならば、現状の方法にこだわる必要はない。
3. **人工知能を使いこなせる** スキルは欲しい（たぶんすべてのライブラリアンがそう）。
4. **人工知能の下で働く？** 人工知能をより良くするために、より良いデータを入力する。
5. 長期的に維持できる価値は何か？

Digital Curation @ King's College London



Digital Curation MA

Overview

A two-year international programme providing a combination of theoretical knowledge and practical training, equipping you to manage and curate the digital information and digital assets of organisations across public and private sectors. Students study for one year at King's and one year at Humboldt University Berlin, and choose from an exceptional range of options covering different aspects of digital curation.

Reasons you should consider the MA in Digital Curation are:

- You will receive an outstanding education that will equip you to become a leader in the field of digital curation and information management.
- Employers are looking for skilled professionals with knowledge and expertise in managing and curating their valuable digital information and assets.
- You will study in two of Europe's most exciting cities and have the opportunity to...

Key information

Application status: Open

Duration: Two years FT, September to September. Students spend four semesters over the course of two years on the programme: two consecutive semesters at Humboldt and two semesters at King's, beginning the programme at Humboldt in all cases.

Study mode: Full-time

Privacy Policy

Course leaders: Dr Mark Hedges (King's), mark.hedges@kcl.ac.uk; Dr Tobias Blanke (King's) +44 (0) 20 7848 1973; Professor Michael Sauter (Humboldt)

Course contact for further information: Postgraduate Admissions Team, Admissions Office tel: +44 (0) 20 7848 7000 fax: +44 (0) 20 7848 7200

Course contact form: [Postgraduate admissions enquiry form](#)

Awarding institution: King's College London

Faculty: Faculty of Arts & Humanities

Department: Department of Digital Humanities

- Libraries, museums, galleries, and archives.
- Digitally literate professionals to work in education and heritage institutions.
- Media organisations, publishing houses, government and industry; research institutions and healthcare and law firms.

<http://www.kcl.ac.uk/study/postgraduate/taught-courses/digital-curation-ma.aspx>

担い手は若手かシニアか？

- **若手によるデータマネジメント**：データライブラリアン、データキュレーションなどのキャリアパスをちゃんと設計しよう！
- **シニアによるデータマネジメント**：引退前（？）の最後のご奉公として、コミュニティに貢献してもらおう。
- **若手案の利点**：キャリアを積んで大きく伸びる可能性があり、新しい可能性も開拓できる。
- **シニア案の利点**：経験豊富なので安定した作業ができるし、キャリアも考えなくていいし。。

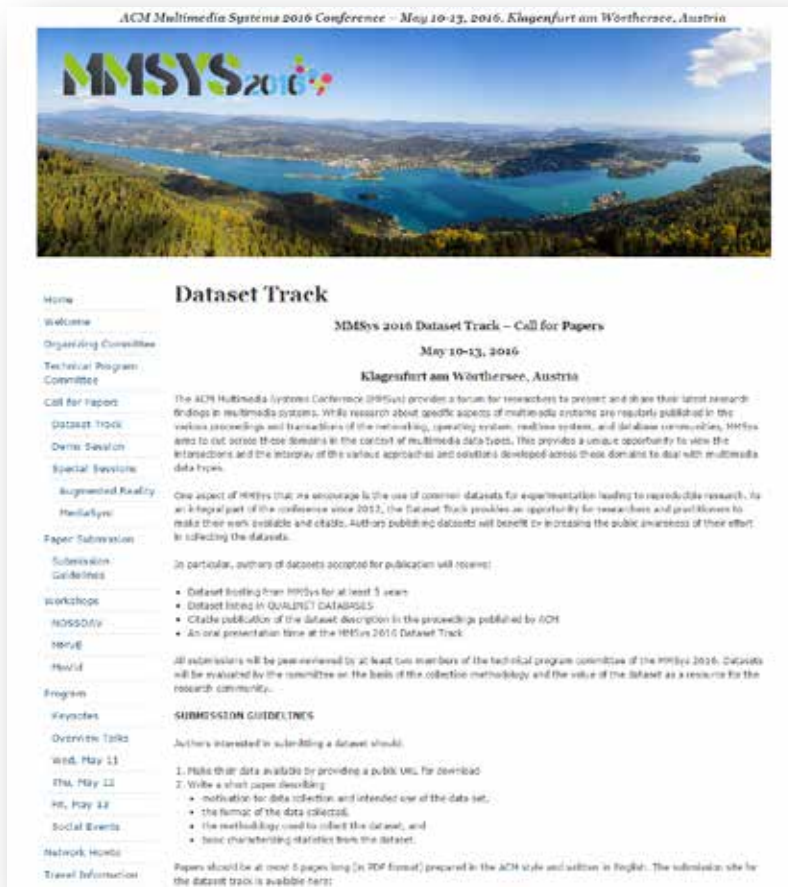
オープンサイエンスと 研究評価・持続可能性

研究資源データセット

Olga Russakovsky*, Jia Deng*, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg and Li Fei-Fei. (* = equal contribution) ImageNet Large Scale Visual Recognition Challenge. IJCV, 2015



データ論文



- **新規性より再利用性：**データセット公開と通常の研究成果では、評価基準が異なる。
- **Data Descriptor：**データセットの特徴を論文化する新フォーマット。
- **査読の導入：**高品質のドキュメントを残し、再利用性を高める。

<https://mmsys2016.itec.aau.at/dataset-track/>

データ引用

1. **ビブリオメトリクス**：引用（被引用数）は、昔から評価の最重要項目の一つである。
2. **データ引用**：論文向けの引用計測インフラに、データも相乗りできる（謝辞はできない）。
3. **データDOI**：データ（論文）に識別子（ID）を付与し、引用の計測を簡単にする。
4. **インセンティブ**：引用が増えて評価が高まれば、データ基盤の研究者もやる気が出る。
5. **持続性**：データ駆動型科学の持続性が高まる。

グローバルID連携と研究評価

1. **論文識別子**：DOI (Digital Object Identifier)は、論文識別子のデファクトスタンダード化。
2. **研究データ識別子**：DOIの付与が最近数年で徐々に普及、データ論文も増えつつある。
3. **研究者識別子**：ORCIDの普及が進むが、スタンダードとみなせるかは微妙な段階。
4. **ソフトウェア識別子**：GitHubへのDOI付与やソフトウェア論文など、まだ試行段階。
5. **研究助成識別子**：すべてのIDを結合すれば、プロジェクトの費用対効果も計算可能？

“Shoulders of Giants”



☒ ウェブ全体から検索 ☐ 日本語のページを検索

巨人の肩の上に立つ



☒ Articles ☒ include patents ☐ Case law

Stand on the shoulders of giants

- 研究とは先人の成果の上に、自分の成果を積み重ねていくことである。
- 研究とは、より大きな「肩」を作ることだとすれば、その「肩」とは何なのか？

「肩」はこんなもの？



株式会社稲葉製作所ウェブサイトより

- 100人乗っても大丈夫な理論、10000人が乗っても大丈夫なデータ基盤など。

Giant Shoulder Index (GSI)

- 「巨人の肩」インデックスは、研究の価値を表す究極の指標である。
- 他人の成果にただ乗りするだけの論文ではなく、多くの人が上に乗っている論文を評価すべき。
- よいデータをオープン化し、たくさんの人が使えば、たくさんの人が乗っているデータという観点で評価を高くする。
- 論文に偏りすぎな指標を適正化する。

「デジタル台風」におけるキュレーションとオープンサイエンス

持続可能なデータプラットフォームに向けた課題

Curation and open science in "Digital Typhoon"
Issues toward a sustainable data platform

北本 朝展^{1,2}
KITAMOTO Asanobu^{1,2}

¹ 国立情報学研究所 ² 総合研究大学院大学
¹ National Institute of Informatics ² SOFVENA

「デジタル台風」とは、台風に関するあらゆる情報を整理し、誰でもアクセス可能なデータプラットフォームの構築を目指すプロジェクトである。気象データや社会データなどさまざまな大規模データを結合するだけでなく、データの意味を拡張性やリンクといった相対的な文脈で解釈する機能など、「デジタル台風」にしかないユニークなデータや機能も人々を集めている。本稿はこのデータプラットフォームと2つの観点から考察する。第1に、公開から約13年間の約1億6,300万ページビューのアクセス集計に基づき、利用実態や情報流通におけるソーシャルメディアの役割なども考察する。第2に、デジタル台風が推進するさまざまな持続可能性の課題も、プラットフォームによるキュレーションや、市民科学とオープンサイエンスのかわり合いといった観点から論じるとともに、持続可能なデータプラットフォームに向けて目指すべき方向を展望する。

デジタル台風、キュレーション、オープンサイエンス、持続可能性、データ統合、データプラットフォーム、バスマインダー、アクセス解析、ソーシャルメディア、市民科学

原稿受理: 2016-06-06
刊行予定: 2016 vol. 59 no. 5, pp. 293-304 doi:<http://doi.org/10.1241/johokanri.59.293>

1. はじめに

「デジタル台風」¹⁾ (図1) とは、台風に関するあらゆる情報を整理し、誰でもアクセス可能なデータプラットフォームの構築を目指すプロジェクトである。1999年4月に最初のアイデアを整理して以来、データプラットフォームの構築を7年以上にわたって継続的に続けてきた。研究プロジェクトや産業プロジェクトでは、一般的に革新性と持続性を両立させることは難しいが、デジタル台風はこの相反する要求を満たす解決策を自らのプラットフォームを対外的に実行始めてきた点に独自性がある。しかし持続性は個人的な努力に依存する面が大きく、長期的に維持可能な仕組みはまだ確立できていない。そこでデータ

プラットフォームの持続可能性という点について、キュレーションやオープンサイエンスなどのさまざまな観点から論じてみたいというのが本稿の目的である。

本稿の構成を述べる。2章でデジタル台風の歴史、3章ではデジタル台風の特徴を簡単に紹介した後、4章ではデジタル台風の利用を分析する。そして、持続可能性に影響を与える要因として、5章ではキュレーション、6章ではオープンサイエンスの現状から現状と課題を論じ、最後に7章で本稿をまとめる。

2. デジタル台風の歴史

デジタル台風の出発点は、気象衛星画像の検索と

情報管理 2016 vol. 59 no. 5 293

持続可能性

- デジタル台風の歴史を紹介しつつ、持続可能性を現在の重要課題とした。
- (研究的な) 革新性と(事業的な) 持続性を両立させるのが難しい。
- オープン化に関する長期的な取り組みを支援する仕組みがない。

北本 朝展, "「デジタル台風」におけるキュレーションとオープンサイエンス: 持続可能なデータプラットフォームに向けた課題", 情報管理, Vol. 59, No. 5, pp. 293-304, doi:10.1241/johokanri.59.293, 2016年8月

サステナビリティの問題

1. **規範に基づくオープンサイエンス**：私自身は重要だと信じているが、他人はどうか？
2. **属人性の問題**：自分が倒れたら、後を継ぐ／継げる人がいない。後継者をどう育成する？
3. **評価の問題**：社会的に多くの人に利用されているが、それを研究評価に取り込む道がない。
4. **資金の問題**：専用の研究費がないため、他のテーマで研究費を取り続ける必要がある。
5. **信用の問題**：いつまでも「個人サイト」では信用されず、研究コミュニティに支えられるサイトにならねば続かない。
6. **データの外部依存性の問題**：データ配信が止まったら／有料化されたら継続できない。
7. **サービスの外部依存性の問題**：外部サービスが止まったら、サービスが継続できない。
8. **基盤システム互換性の問題**：セキュリティアップデートなどで互換性が打ち切られると、システムが動かなくなる。

オルタナティブとしてのオープンサイエンス

- **持続可能性**：現状のサイエンスは短期的な指標に最適化し、長期的な視点が減っている。
- **慣性**：古い制度の影響が強いため、新しい制度へ向かう力を継続して加える必要がある。
- **外部環境の変化**：図書館は否応なく変化を迫られているし、研究者も変化を迫られる。
- **変化の先取り**：将来を見越して、現在は主流でない道に踏み出すことも必要？
- **あるべき姿と代替案**：現在主流の考え方に変わる、代替案としてのオープンサイエンス。

おわりに

まとめ

- オープンサイエンスは様々な動きの収束点であり、その目標は人によって異なる。
- サイエンスのオープン化は、現在のクローズドなサイエンスを逆転する考え方。単なる技術論では収まらず、制度からの見直しが必要。
- オープン化への動きは、評価され、支援され、持続的な活動になることが不可欠である。
- 持続可能性の必要条件は、活動自体が魅力的であること。その点で、まだまだ努力と成功が足りない気もしている。

関連ウェブサイト

- オープンサイエンス

- <http://agora.ex.nii.ac.jp/~kitamoto/research/open-science/>

- 人文学オープンデータ共同利用センター

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

- デジタル台風

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

- DIASプロジェクト

- <http://www.diasjp.net/>