人間文化研究情報資源共有化研究会報告集 7

Proceedings of the Study on Information Resources of the Human Science Vol.7

平成 29（2017）年 3 月

March 2017

大学共同利用機関法人　人間文化研究機構

総合情報発信センター

高度連携情報技術委員会

目　　次

人間文化研究情報資源共有化研究会報告集７の刊行にあたって　　榎原　雅治 …… 3

第11回研究会報告集

第11回研究会プログラム …… 5

大型プロジェクトの目指す検索機能の高度化の取り組み …… 7
国文学研究資料館古典籍共同研究事業センター　山本　和明

身装画像データベース「近代日本の身装文化」…… 13
――研究資源データベースの発信と展開――
国立民族学博物館　丸川　雄三

絵入百科事典データベースの構築――『訓蒙図彙』を核として …… 19
国際日本文化研究センター　石上　阿希

洛中洛外図屏風「歴博甲本」人物データベースについて …… 21
国立歴史民俗博物館　小島　道裕

史料編纂所所蔵肖像画模本・歴史絵引データベースの課題 …… 41
東京大学史料編纂所　藤原　重雄

画像内容に基づく検索技術に対する期待と現実 …… 67
国立情報学研究所　北本　朝展

第12回研究会報告集

第12回研究会プログラム …… 91

オープンサイエンスが生まれた背景と最近の政策の動き …… 93
文部科学省科学技術・学術政策研究所　林　和弘

歴史資料のオープンデータ化に関する現在と未来 …… 117
――歴博の総合資料学の取り組みを通じて――
国立歴史民俗博物館研究部　後藤　　真

歴史的典籍NW事業におけるオープンデータ――その戦略と課題―― …… 127
国文学研究資料館古典籍共同研究事業センター　山本　和明

言語研究と「オープン」データ …… 137
国立国語研究所コーパス開発センター　前川　喜久雄

人間文化研究情報資源共有化研究会報告集７の刊行にあたって

大学共同利用機関法人人間文化研究機構では、機構に加わる６研究機関（国立歴史民俗博物館、国文学研究資料館、国立国語研究所、国際日本文化研究センター、総合地球環境研究所、国立民族学博物館）の構築してきたデータベースを横断検索する統合検索システム（nihuINT）や、時空間システム（GT-Map/GT-Time）、参加型データベースであるnDPの開発を進めてまいりました。そして、これらの情報資源の共有化システムの運用と展開のために、平成20（2008）年には研究資源共有化事業委員会を立ち上げました。平成28年度から始まる第３期中期計画期間においては、機構内に新たに設置した総合情報発信センターのもとで高度連携情報技術委員会がこの事業を引き継ぐとともに、機構内外の人間文化研究情報との機動的な統合検索のシステムの構築をめざしております。

また、研究資源共有化事業委員会では、人間文化に関わる研究情報資源の共有化の推進のため、関係学界に呼びかけて、平成21（2009）年より、「人間文化研究情報資源共有化研究会」を開催してまいりました。平成27（2015）年度は、28年２月６日に第11回資源共有化研究会「人間文化研究機構のもつ画像データ共有化の前進に向けて」を京都市のガーデンシティで開催いたしました。本機構内には、絵画や写真などの画像情報を中心としたデータベースが多数あり、nihuINTにも登録されています。ただし、現状では、これらの画像を検索するためには、資料名や描かれた像主などの固有名詞を入力することが必要で、描かれた図柄のさまざまな属性から検索することはできません。もし、画像に「子供」「犬」「病気」のような普通名詞からなるキーワードをメタデータとして埋め込むことによって、共通性のある画像、類似した画像を収集する「絵引きシステム」を作ることができれば、共有化システムは、研究者にとっても一般市民にとっても、格段に使いやすくなると考えられます。当日の研究会では、各機関の具体的な取り組みに即して、本機構のもつ画像情報を、より広く、よりなじみやすく活用していくための方法を求めて、活発な議論が行われました。

平成28（2016）年度は、29年２月３日に、第12回資源共有化研究会「人文科学におけるオープンサイエンスの課題」を名古屋市の愛知工業大学本山キャンパスにおいて開催しました。現在、世界の科学分野においては、論文や、その根拠となるデータをインターネット上で公開・共有化するオープンデータ、オープンサイエンスの動きが急速に始まっています。紙媒体による出版業の機能が大きく変化している中、研究成果の発表、蓄積、検索の新たなあり方が求められるのは必然的なものといえるでしょう。また論文の内容について、第三者による検証がより容易に可能となるよう、データの共有化を求める動きも一層進むものと予想されます。しかしながら人文科学の分野においては、この問題に対する反応はあまり迅速ではないように思われます。当日の研究会では、オープンサイエンスをめぐる動きの展開してきた経緯や国の施策の方向性について報告していただくとともに、人文科学の分野でオープンサイエンス化を進めるうえで、どのようなことが課題となるのか、また研究の進展や研究成果の社会化のためにはどのようなオープン化が必要であり、望ましいのかといった問題について意見がかわされました。

本報告書は、上記２回の研究会で発表された報告を紹介するものです。人間文化研究にかかる情報資源の共有化推進の一助となることを念願いたします。

2017年３月１日

大学共同利用機関法人　人間文化研究機構
理事　榎　原　雅　治

人間文化研究機構のもつ画像データ共有化の前進に向けて

第11回人間文化研究情報資源共有化研究会

○主　催　大学共同利用機関法人人間文化研究機構・研究資源共有化事業委員会
○日　時　平成28年２月６日㈯　13時00分～17時30分
○会　場　TKPガーデンシティ京都「山吹」
京都府京都市下京区烏丸七条下ル721-1　京都タワービル７階（京都駅より徒歩２分）

○開催趣旨

現在、機構内には、画像（絵画、写真）情報を中心としたデータベースが多数ある。nihuINTで「画像・映像・音響」に分類されたもの以外にも、たとえば「所蔵資料目録」の中に「怪異・妖怪絵姿」「近世風俗図」などの画像データベースがある。

人間文化研究機構では、第３期に向け、機構内各機関における研究成果を、より見やすい形で社会に提供することを求められている。機構のもつさまざまな画像データは、専門外の研究者や、文化に関心のある一般市民にもなじみやすい情報資源であり、これをより活用していく方法を探りたい。

機構内のデータベースの連携検索はすでに可能であるが、検索するためには、現状では資料名や描かれた像主などの固有名詞を入力することが必要で、描かれた図柄のさまざまな属性から検索することはできない。

画像に、「子供」「犬」「病気」のような普通名詞からなるキーワードをメタデータとして埋め込むことによって、共通性のある画像、類似した画像を収集する「絵引きシステム」を考えることができれば、共有化システムは、研究者にとっても一般市民にとっても、格段に使いやすくなると考えられる。

すでに一部の機関ではそうした取組が始まっている。各機関の現状と将来計画を俯瞰し、将来的な連携検索を見通す場としたい。

○プログラム

13：00～13：15　問題提起　人間文化研究機構理事　榎原　雅治
13：15～13：45　「大型プロジェクトの目指す検索機能の高度化の取り組み」
国文学研究資料館古典籍共同研究事業センター副センター長　山本　和明
13：45～14：15　「近代日本の身装データベースの画像について」
国立民族学博物館先端人類科学研究部准教授　丸川　雄三
14：15～14：45　「絵入り百科事典データベースの構築――『訓蒙図彙』を核として」
国際日本文化研究センター特任助教　石上　阿希
15：00～15：30　「洛中洛外図屏風『歴博甲本』人物データベースについて」
国立歴史民俗博物館歴史研究系教授　小島　道裕
15：30～16：00　「肖像画模本／歴史絵引データベースの課題」
東京大学史料編纂所古代史料部門助教　藤原　重雄
16：00～16：30　「画像内容に基づく検索技術に対する期待と現実」
国立情報学研究所コンテンツ科学研究系准教授　北本　朝展
16：40～17：30　総合討論

大型プロジェクトの目指す検索機能の高度化の取り組み

国文学研究資料館古典籍共同研究事業センター
副センター長　山　本　和　明

1．大型プロジェクトの概要

国文学研究資料館では現在、文部科学省の大規模学術フロンティア促進事業「日本語の歴史的典籍の国際共同研究ネットワーク構築計画」（略称：歴史的典籍NW事業）が進められている。事業の期間は2014年４月から10ヶ年。本報告書刊行時には３年目が終了ということになる。その目標とするところは大きく言って３点ある。１つ目はあらゆる分野の古典籍30万点の全冊画像を、書誌とともにWeb上で公開するということ。2017年３月時点で延べ約78000点の撮影が完了した。２つ目はその古典籍画像を用いての国際共同研究の推進。そこでは他分野の研究者と共同で推進する異分野融合研究や、国文研が中心となり進めている「総合書物学」といった新学術領域の創出ということも含んでいる。３つ目は、先の２つの成果として国際共同研究ネットワークの構築ということになる。このネットワークを活用して、日本由来の文化や学問領域を国際的な研究の「場」へと引き出し、経常的に国際的に日本研究を推進していこうとするものである。

この論稿の題目にある「検索機能の高度化の取り組み」は、いわば１つ目の目標に関わる取り組みとなる。まだ先のこととは言え、30万点もの古典籍の全冊画像、予算申請の必要上、１点の古典籍を仮に150コマと設定しているのだが、計算すると、30万点すなわち4500万コマの画像のなかから、如何にして必要な画像へたどり着くか、如何に整理するかという問題が浮上してこよう。図書館単位での蔵書であれば、請求記号などから類推し、必要な分野の典籍を絞り込んで、片っ端から閲覧という力業も不可能ではないが、この場合は不可能であろう。そのため画像作成に連動し、検索機能の高度化に関する取り組みが不可欠である。しかも国文学の研究者だけが利用するわけではなく、あらゆる分野の研究者、一般市民にも活用を促すことを考えているため、如何に検索すれば利用しやすいかということは重要な観点だったのである。この点については、科学技術・学術審議会学術分科会研究環境基盤部会に設置された学術研究の大型プロジェクトに関する作業部会からも、平成26（2014）年度の留意事項として「画像へのタグ付けを質・量ともに充実させることはもちろん、異分野の研究者コミュニティのニーズを踏まえた検索機能の充実を、より積極的に図ることが重要」とのご指摘をうけたところであった。検索への期待のあらわれであり、当館もデータベースの機能強化に関わる研究として、主として国立情報学研究所などと精力的に現在も取り組んでいる。

2．検索機能の高度化に向けて

では、利用者のための検索機能をどのように考えたらいいのだろうか。このプロジェクトの場合、一定の期間がさだめられているが、期間を終えても発展的に継続しうる仕組み

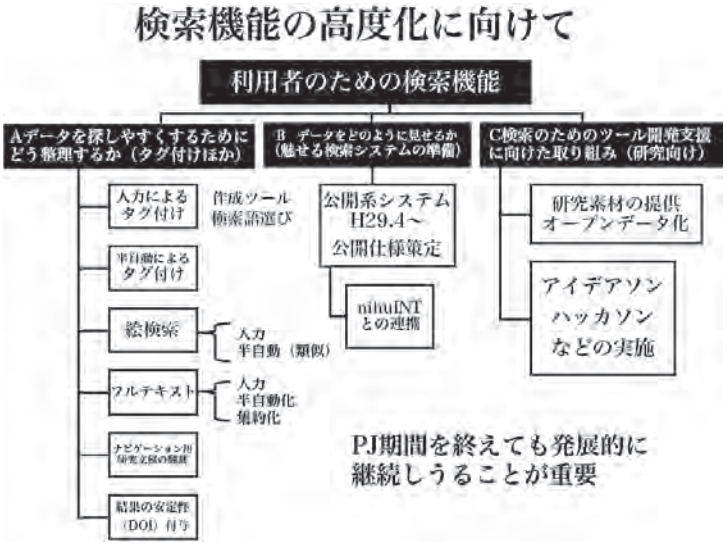
が何よりも大切であることは言う迄もない。私たちは３つの観点からこの検索機能の高度化について考え、取り組むこととした。その３つをまずは列挙しておく。

Ａ　データを探しやすくするためにどのように整理するか（タグ付け等）

Ｂ　データをどのように見せるか、魅せる検索システムの整備

Ｃ　技術の進歩を見据えて、検索のためのツール等の開発をどう支援するか

図表化すると次のようになる。



このうち、本稿ではＡを中心に、戦略としての取り組み４点について簡単に紹介しておきたい。

３．戦略としての取り組み①　タグ付け

検索を有効なものとするためには、タグを付与することがまず想定できよう。国立国会図書館では、たとえば近代文献について、文章中の固有名詞や目次などから検索可能にしており、その情報の蓄積により格段に便利な検索が可能となっていった。そのことを踏まえ、如何にしてタグを付与していくべきか。さまざまな意見もあり、その方針や方法については、当館研究部の先生方を中心に検討いただいたが、そのタグ付け開始当初における主な方針は以下の通りである。これらはPDCAサイクルのもと、常に検討を加え変更していくものであることをあらかじめお断りしておく。

○各分野の古典籍のうち利用頻度と研究上の価値の高いものを選定し、画像（文章と挿絵）に対してタグをつける。タグを付けるのは、データベースに搭載する古典籍全体の５％をひとまずの目安とする。

○タグ付けは、文学及び歴史分野の古典籍から着手し、順次他の分野に拡大していく。他分野についてはその分野の研究者コミュニティの協力を得て、タグ付けワーキング等を実施し、方針をさだめる。

○対象とする古典籍の選定については、すでにデジタル化された本文がひろく普及しているもの、および、タグ付けに際し独自な問題が生ずるものについては、当面のタグ付け作業から除外する。

○思いがけない「発見」を導くため、画像１枚ごとにタグを入力する。
○タグ付けは、原則として、対象とする古典籍に出現する固有名（人名・地名・書名・年号など）をキーワードとしてそのまま拾い出す。

タグについては、検索効果を優先し、統制語彙を用いるなどの意見もあった。事実、現在進めている和算分野のタグ付けWGではそうした用語の検討がなされている。その一方で、国文学分野では、時代ごとや和歌・演劇などの分野ごとでの研究が深化しているため、同じ事を表現するにも専門ごとで異なる用語（時代ごとでの意味の相違など）があることも想定しえた。果たして統制可能かということも難しい問題として浮上してきたことを付記しておきたい。特に今回のプロジェクトは、何も国文学の研究者だけが利用するものではなく、さまざまな研究者や一般の利用も想定されるため、個人個人での異なる見解のもとで統制するには、難しい判断を必要とした。そうした中、このプロジェクトの顧問をお願いしている九州大学前総長の有川節夫先生からのご助言には目から鱗の思いを得た。重要なご指摘と考えるため、以下、当方のメモ書きから抜粋しておく。

モノ自体は昔から何も変わっていないが、これに対する見方が変わりうる。すると途端にキーワードでは拾えなくなってしまいがちだ。たとえば過去の人はこれをそんな風に見ていなかった、という場合には、現代で付けたキーワードではいくら検索しても探せないことになる。つまり物が変わるのではなく、見方が変わるのだという考え方をしてみるとどうだろうか。そうなると、下手にキーワードをつけようとするよりは、全部あらゆる言葉を付けてしまう。それによって可能性が広がるのではないか。

言うなれば、あらゆる言葉でタグを付すことを許す寛容性ということか。たとえば統制された語彙をつけてもいいし、そうでなくても良い。要はタグを付けることが大事なのだと改めて考えさせられたのだった。もちろん難しい問題もある。しかしWeb上での検索であるから、連想検索やさまざまな検索の可能性が今後も開けてくることだろう。人工知能の発達により、今とは全くことなる学習効果が出てくることを思えば、枠組みにとらわれずなによりもスピード感が大事だ、と実感したというのが正直な思いであり、現在までの報告となろうか。

効率よくタグを付与する取り組みとして、人文情報学研究所の永崎研宣氏と当館特任助教の松田訓典によりWeb上でのタグ付与システムが開発されたことで、格段に道が開けたと考えるが、現在はまだ限定的な公開での利用となっている。重要な古典籍を選び、画像上をカーソルで範囲指定し、その位置情報を記録しつつ、選択肢から統制された用語を選ぶか、自由に言葉を記述できるようなツールで、オープンソースを活用して制作された。これを使えば、絵にタグをつけるといったことなどは簡単にできる。また、システム上での一元管理を果たすことで、誰によるタグ付けかの記録情報を備えることも容易だ。現在、国文研教員を中心にこの付与システムを活用し、Web上でタグ付けが進められており、微調整を重ねながらブラッシュアップをしている最中である。当面は研究者コミュニティへタグ付けへの参画を募り、将来のソーシャルタギングの可能性を見据えていきたい。

このツールは、指定した古典籍にタグを付与していくものであるが、これとは別に公開システム上で、画像を見る際に、タグを提案できるボックスの導入も予定している。画像をみた人に、表示されている画像の検索語の提案を自由にお願いするものであり、文字通りのソーシャルタギングとなる。これもまずは研究者コミュニティへの開放からすすめて

いくことになるだろう。

4．戦略としての取り組み②　ワード・スポッティング、テキスト化

先に述べたタグ付けシステムは人力によりタグを付与するものであったが、今後、展開していきたいと考えているのはタグの半自動的付与ということである。この可能性を示してくれたのが公立はこだて未来大学の寺沢憲吾先生のご研究で、現在、国文研とともに共同研究を進めている。その研究はワード・スポッティング技術に関するもの。一知半解の人間がその内容を述べるのもどうかと思うが、大量に蓄えられた文書画像の中から、与えられた画像に似ている画像を検索し、その出現位置情報を提示するというものである。ここでは文書画像をスリット状に切り出し、そのスリット画像を特徴ベクトルで記述し、ベクトルのマッチング問題として、類似画像を検出していく。この技術を応用することで、特定の言葉、たとえばタグ候補の用語を範囲指定し、類似画像を検出させ、その類似画像の候補群でその言葉と確定できるものに対し、一括でタグ用語として付与できれば、タグ付けがかなり省力化できるのではあるまいか。加えての応用で、用語単位でテキスト化していく支援システムともなろう。こうした取り組みを進めるためにも、平成29年４月段階での公開系システムでは実験的導入として寺沢先生の技術を確認していく場を設けることとしている。

このテキスト化についてであるが、凸版印刷・公立はこだて未来大学と共同ですすめ、原理検証結果については既にメディアなどにも取り上げられたため、今賛言を重ねない。くずし字を含む新方式OCRの開発や、機械学習の導入などを含め、今後もテキスト化実証試験の取り組みは継続していかねばならないし、30万点の古典籍画像を利活用できるものとするためには何よりもテキスト化の道筋を開くことが重要であろう。その作成されるテキストも、国際テキスト（Text Encoding Initiative）に準拠させることを考えている。

5．戦略としての取り組み③　古典的名著というナビ

異分野融合研究を推進するためにも、他分野の研究者とのセッションは不可欠である。他分野の研究者の方と話をしていると、如何に自分の専門分野が限られた範囲の知識のなかで存在しているかを痛感することがある。研究者は必ずしも他分野の研究者ではない。この自明のことを前にして、他の分野の古典籍を探す時にはどうしたら良いのか。このことを考える時、やはりその分野の古典的名著を紐解くことが良い指針となる。

たとえば古典的名著（著作権の切れたようなテキストを含む）をテキスト化し、閲覧可能にし、その記述の中から古典籍が直接確認できたらどうか。そう考え、ナビゲーションの役割としての古典的名著の導入を考えた。これもまだ器としての機能導入段階であり、今後研究者コミュニティの協力を得て名著や辞書等の記述をナビゲーションに活用していく準備を進めている。

6．戦略としての取り組み④　結果の安定性　DOIの導入

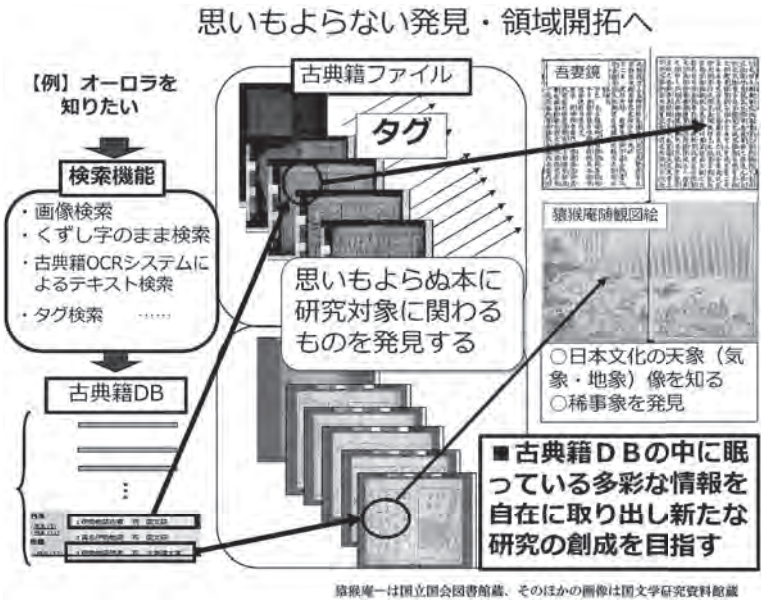
絵検索を含め、あらゆる方法でのタグ付けの方法を模索しているが、今回の画像公開にあたって最も重要であるのは、検索結果の安定性ということではないだろうか。検索結果が、数年後には見当たらないといった事態は、インターネットが盛んになった当初、よく

見られた。せっかくテキスト等、有益な情報を公開していても、再度確認できなければどうしようもないし、後から確認出来ないのなら、それは信用できないものとなる。古典籍などを用いた人文科学の研究では、他人の閲覧ができないような個人蔵の資料紹介などが盛んに行われている。近年はその画像を一部紙面に掲載しているなどし、ある程度改善されたとは言え、他者による検証が可能でなければ、学問としては未成熟なものと言える。人文科学は科学でなければならない。僕はこう思う、こんな見方もあるよという提案型の論文も勿論あるが、その土台となるテキストについては確実なものであってほしい。検索結果の安定性、すなわちいつ見ても結果が同じであることは、人文科学を検証可能な学問にするということを達成する根本的な課題なのである。

私たちは研究会を組織し、どのようにすれば検証可能なものとなるかについて、有識者などの意見を聞きながら検討を重ねてきた。今回の大型プロジェクトではデジタルオブジェクト識別子（Digital Object Identifier = DOI、permanent URL の一種）の導入を取り決めた。加えて、論文等へこの DOI を記載する取り組みも開始している。繰り返しになるが論文等に DOI を引用記載することで、論文を読む人は簡単にその論拠となる作品画像にたどり着けることをめざしており、このことが人文科学を格段に進展させると確信する。

7．おわりに

以上が第11回人間文化研究情報資源共有化研究会「人間文化研究機構のもつ画像データ共有化の前進に向けて」（平成28年２月６日、於 TKP ガーデンシティ京都）で報告した大型プロジェクトにおける検索機能の高度化を巡っての報告の一部である。本稿ではその後の状況もあわせた形で記載したことをお断りしておく。2017年４月（予定）からの公開系システムでは、検索機能について、当初基本的なものからスタートしていけるよう機能を準備した。とは言え、コンテンツ（タグ）が充実しなくては何も始まらない。公開系システムがスタートしてからもコンテンツを充実させることが重要、と肝に銘じ下記の図のような形で、思いもかけない発見が可能な、活用できるシステムとしていきたいと思う。



猿猴庵一は国立国会図書館蔵、そのほかの画像は国文学研究資料館蔵

身装画像データベース「近代日本の身装文化」
── 研究資源データベースの発信と展開 ──

国立民族学博物館　丸　川　雄　三

明治から昭和期（1868～1945年）における日本人の身体と装い（身装）に関する画像とその研究成果を発信するウェブサイト、身装画像データベース「近代日本の身装文化」を開発した。本稿ではシステムの有効性を示すとともに、研究資源データベースのさらなる展開例として、研究情報の付与やモチーフ分析を支援する手法の具体的な提案と検討を行った。

1．はじめに

明治維新以降、約80年間における日本人の「身装 ── 身体と装い」に関する文化変容の実態については、十分に解明されていない点が多い[1)]。MCD プロジェクト[2)] では、明治から昭和期（1868～1945年）における身装に関する画像（身装画像）を対象に、そのデータベース化と発信に取り組んでいる。報告者は、データベースシステムおよびウェブシステムの開発を担当する立場で2012年から MCD プロジェクトに参画し、身装画像データベース「近代日本の身装文化」を開発した[3)]。本稿ではウェブインタフェースの説明を中心に、研究者による書き込み機能やモチーフ分析、身装絵引の可能性について述べる。

2．身装画像データベースの特徴

身装画像データベースには、近代日本の身装文化に関する画像15,000件が登録されている。これらの身装画像は、新聞挿絵や絵葉書、写真などから構成されており、それぞれに詳しい管理情報と身装コードなどの研究情報が付与されている。加えて本データベースには説明文（コメント）が付けられている。コメントは身装文化の研究者による解説情報であり、画像に現れる身装とその時代背景などが丁寧に記述されている。初学者などの前提知識を持たない利用者であっても、コメントを読むことによって、画像に現れるモチーフを読み解くことができ、専門用語などの新しい知識を同時に得ることができる。

またデータベースには、身装に関するテーマごとに、その周辺を読みやすい語り口で生き生きと描写したエッセイ形式の「参考ノート」も収録されている。身装画像のコメントと並んで身装文化を理解する上で大いに役立つものであり、掲載内容をまとめた書籍も出版されている[4)]。

このように身装画像データベースの特徴は、精査された専門用語に加えて、身装文化を理解するための豊富なテキスト情報が付与されている点にある。そのため公開用ウェブサイトにおいては、当該研究分野のための学術情報の共有と同時に、広く一般に向けた情報発信が求められる。

3．ウェブサイト「近代日本の身装文化[5)]」

以上の要件をふまえて、ウェブサイト「近代日本の身装文化」を開発し、2016年５月に一般公開した[6)]。

図１はトップページである。「一覧から探す」と「項目から探す」の二つの入り口を設けている。このうち「一覧から探す」は、身装画像を「身装画像コード」、「年代」、「制作者」の一覧から探すことができるディレクトリ型のインタフェースである。これらは全て統制語を用いて管理されているため、一覧に示された検索語を選ぶだけで、比較的簡単に身装画像を閲覧できる。一方、より詳細な条件で身装画像を探したい利用者は、「項目から探す」を用いることで、テキストによる全文検索および複数の条件を組み合わせた検索が可能である。また参考ノートにもトップページからアクセスすることができる。

図２は詳細表示の一例である。画像表示部分を大きく設け、全体と部分の二種類の画像をタブで切り替えて閲覧できる。またコメントを画像のすぐ下に配置し、キーワードハイライト機能とともに、身装画像の内容やその時代背景などが、初学者や一般の利用者にもわかりやすいよう配慮している。その他のメタデータもすぐその下に項目名とともに配置しており、身装画像データベースの成果である詳細な研究情報を確認することが可能である。

図１　「近代日本の身装文化」トップページ

図２　詳細表示例：「新年梅 第36回」（右田年英筆、東京朝日新聞1902年２月８日号７面）

4．コメントの編集とサブコメント機能

公開用ウェブインタフェースには、右上にログインボタンが付けられている。MCD プロジェクトが認めたユーザにはアカウントとパスワードが発行されており、ログインすることで、コメントの編集およびサブコメントの書き込みができる。

なお全ての情報は画像情報管理システム上で管理者にのみ編集が許されている。しかしコメントについては、公開用ウェブインタフェースから直接編集できるように、限定的な編集機能を組み込んでいる。一方、サブコメントについては全てのログインユーザが書き

込みの権限を持つ。サブコメント機能の本格的な利用はこれからであるが、身装画像に関する素朴な疑問や、コメントやその他のメタデータへの質問や指摘、関連する情報の提供など、様々な利活用が可能である。

5．身装画像コードとモチーフの分析

身装画像には、当時の髪型や衣装、人物同士の関係性などが活写されており、その内容（モチーフ）に対応する「身装画像コード」が各画像にひもづけられている。身装画像コードは概念体系に近いものであり、抽象化されたコードと具体的な名称を表わす幾つかのキーワードから構成されている。しかし身装画像コードは画像全体に対応しているため、その名称に対応する箇所を画像の中に見つけるには、さらに専門的な知識が必要である。

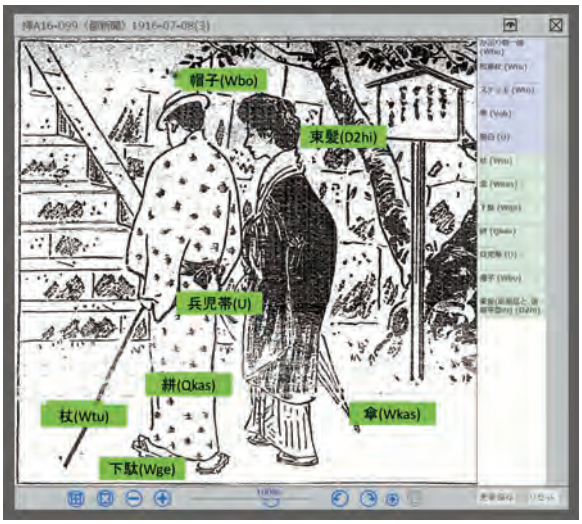


図３　身装画像分類支援システム画面例：「浮雲第43回」（井川洗厓筆、都新聞1916年７月８日号 ３面）

そこで報告者は、身装画像コードやキーワードと画像との対応付けを行う「モチーフ分析」を考案し、そのための作業環境として分類支援システムを開発した[7]。図は分類支援システムの画面例である。データベースに登録されている髪型や持ち物などの名称を、身装画像の上に直接対応づけることができる。この例では女性の髪型である「束髪」や、男性の着物の帯の名称として「兵児帯」などのキーワードが、対応する身装コードとともに配置されている。

6．モチーフから絵引へ

身装画像分類支援システムによって、画像とモチーフ名の対応付け（モチーフ分析）の可能性が見えてきた。さらにそこからモチーフ名を索引とした「身装絵引」を作ることも可能であろう。

民俗学者でもある渋沢敬三は、「字引ならぬ絵引をつくれぬだろうか」と考え、『絵巻物による日本常民生活絵引』を編纂した[8]。絵巻から日常生活の場面を抜き出し、そこに表われるものに名称をつけて言葉から引けるようにした図解辞典（絵引）である。

渋沢栄一記念財団実業史研究センターと国立情報学研究所連想情報学研究開発センターは、「実業史錦絵絵引「衣喰住之内家職幼絵解之図」」を共同で制作し、2009年、ウェブ上に一般公開した。明治初期に子供向けの教材として描かれた錦絵「衣喰住之内家職幼絵解之図」は、一軒の民家が出来るまでを絵入りで説明するものであり、設計から製材、棟上げ、内装工事に至るまでが懇切に取り上げられている。このウェブサイトでは、錦絵に登場する人物や道具500点の名称と図像を「絵引データベース」および「絵引ギャラリー」として発信している。

「身装絵引」を作るためには画像の該当部分を身装画像から抜き出す必要があるが、「実

図４　渋沢栄一記念財団「実業史錦絵絵引[9]」より、絵引ギャラリー（左）と絵引データベース（右）

業史錦絵絵引」は当初より汎用性のある「絵引システム」として構築されており、この身装画像における絵引の作成にも活用が可能である。

７．今後の展望

身装画像データベースの特徴を活かしたウェブサイトを実現し、一般公開を開始することができた。今後は安定した公開運用とともに利用状況をふまえた機能改善を実施する予定である。またログインユーザによるコメントの編集およびサブコメント機能については、国立民族学博物館が進めている「フォーラム型情報ミュージアム・プロジェクト[10]」においても活用が進められている。MCD プロジェクトによる利活用が進むことで、国立民族学博物館の他のプロジェクトにもその効果が波及することが期待できる。さらにモチーフ分析と身装絵引について、分類支援システムを試作し身装画像データベースへの試験的な適用を実施した。文化資源としての学術データベースの展開手法を具体的に提示することにより、活用の幅と可能性を示すことができたのではないかと考えている。

８．謝辞

本研究は、「JSPS 科研費基盤Ｂ課題番号24300099（平成24年度～平成26年度）「近代日本の身装画像デジタルアーカイブの構築 ── 文化変容に視点を据えて」（代表：高橋晴子）」の成果を受け実施されました。関係者の方々に厚くお礼申し上げます。

注

1）高橋晴子『近代日本の身装文化：「身体と装い」の文化変容』三元社、463p、2005。
2）MCD プロジェクトの構成メンバーは次の通りである。
高橋晴子、久保正敏、猿田佳那子、大丸弘、田中昌美、谷本滋、津田光弘、中川隆、野林厚志、八村広三郎、丸川雄三
3）丸川雄三「身装画像データベース「近代日本の身装文化」の構築」情報処理学会・じんもん

こん2013論文集、Vol. 2013、No. 4、pp. 233-238、2013.
4）大丸弘、高橋晴子『日本人のすがたと暮らし：明治・大正・昭和前期の身装』三元社、536p、2016。
5）http://shinsou.minpaku.ac.jp/
6）丸川雄三「身装画像データベース「近代日本の身装文化」の公開と運用 ── 公開用ウェブインタフェースと研究者の参加を促す編集環境の実現 ── 」情報処理学会・じんもんこん2015論文集、Vol. 2015、pp. 233-238、2015。
7）丸川雄三「身装画像におけるモチーフの分析と絵引の研究」情報処理学会・人文科学とコンピュータ2015-CH-105（2）、pp. 1-2、2015。
8）澁澤敬三、神奈川大学日本常民文化研究所編：『絵巻物による日本常民生活絵引』新版、平凡社、1984。
9）http://ebiki.jp/
10）http://www.minpaku.ac.jp/research/activity/project/ifm/index

絵入百科事典データベースの構築
――『訓蒙図彙』を核として

国際日本文化研究センター特任助教　石　上　阿　希

本稿では、筆者が現在構築を進めているイメージデータベースの構想・概要について簡単に紹介する。

1．データベースの目的

近年、国内外において大規模な日本古典籍のデジタルアーカイブが行われている。そのような状況を受け、国際日本文化研究センター（以下、日文研）では各機関で公開されているデジタル画像を利用したデータベース（以下、DB）の構築を進めている。2015年４月にDBの構想・構築をスタートさせ、順調にいけば2017年中に第１次版を公開する予定である[1)]。

本DBのコンセプトは「絵入・図解百科事典DB」であり、その目的と方針をまとめると下記のようになる。

1．多分野の研究の有力な調査・検索・情報収集ツールをめざす。
2．図像の情報を分類し、アクセスするためのプラットフォームをめざす。
3．名と図（と釈）を備えている古典籍を対象とする。
4．日文研所蔵資料に限らず所蔵先で画像が公開されている資料を選ぶ。（安定的なURLを保持するものを優先する）

※歴史的古典籍NW事業（国文学研究資料館）、国立国会図書館デジタルコレクション、早稲田大学図書館古典籍データベースなど。

主に近世期に出版された絵入百科事典やそれに類する書物を対象として、そこに記録されたことばと絵を検索出来るDBを構築する。それらの書物をDB化することによって、文学史、美術史、建築史、医学史など様々な研究の有力な検索ツールの役割を果たすことを目指している。また、日文研が所蔵する資料のみを扱うのではなく、すでに公開されている画像DBから情報を蓄積し、日文研と各機関のDBをつなぐシステムを構想している。

例えば本DBの核であり、第一に入力作業をすすめている『訓蒙図彙（きんもうずい）』は国立国会図書館デジタルコレクションDBで公開されている画像を底本としている。

2．『訓蒙図彙』と類書

『訓蒙図彙』は寛文６年（1666）と記された序を持つもので、森羅万象の事物を名称と絵によって明示している。編纂者の中村惕斎（1629-1702）は京都の儒学者であり、二男一女一庶子の父であった。惕斎は序文で本書の制作動機を自分の子どもの教育のためと記している。

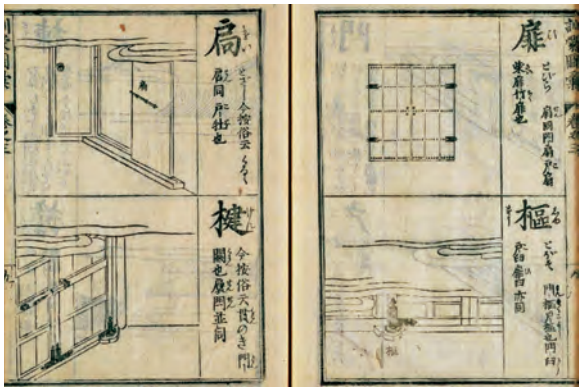


図１　『訓蒙図彙』
（国立国会図書館デジタルコレクション）

項目名	女（ちよ によ）
語釈補記	をんな むすめ 未嫁 日女已嫁曰婦め よめ
分類	人物 公家
書名	『訓蒙図彙』
成立年	寛文6年
巻名	巻之四
丁付	二丁表
所蔵	国立国会図書館蔵

図２　データベース検索結果詳細画面（参考）

本書以降、絵と語が一対になった様々な事典類が刊行され、さらにその形式をもじったパロディが作られるなど、本書から派生した書物は多岐にわたる[2)]。また、『訓蒙図彙』自体も、元禄８年（1695）の『頭書増補訓蒙図彙』、寛政元年（1789）の『頭書増補訓蒙図彙大成』と二度にわたって増補改訂版が作られており、時代の要求に応じながら多くの読者を啓蒙し続けた。

本書は「わが国最初の絵入り百科事典」と評される[3)] 書物で、20巻14冊にわたって17部門（天文・地理・居処・人物・身体・衣服・宝貨・器用・畜獣・禽獣・龍漁・蟲介・米穀・菜蔬・果蓏・樹竹・花草）1484項目を収録している。見開き一丁毎に４つの事物を配し、漢字・ひらがなで名称を記し、形状を図画で表す（図１）。項目数は1484であるが、それぞれに俗称や異称も記載されているため、合計で5000語ほどが収録されている。

DB 作業では、各図に記述された文字を翻刻し、本文通りの表記・よみと通用の表記・よみの両方を入力している。図２は検索結果画面のイメージである。切り出した画像データの横に、文字の翻刻、分類、書名、巻名、丁数、出自の書物の所蔵先を記している。所蔵先をクリックすると各機関の該当ページに移動し、全体像を確認することが可能である。

今後の課題は、類似イメージや関連項目を芋づる式に表示させる方法や名称のわからない事物を検索する方法の検討である。現状ではできる限り詳細なタグ付けを行っているが、さらなる解決方法を模索したい。

注

1）本 DB は、稿者が所属する「広領域型連携基幹研究プロジェクト　異分野融合による「総合書物学」の構築」における日文研ユニット「キリシタン文学の継承：宣教師の日本語文学」の研究の一環でもある。

2）『訓蒙図彙集成』（大空社、1998-2002年）では、訓蒙図彙ものとして下記30点の書物を収録している。本 DB ではこの集成を参考に書物の選定を行っている。立花訓蒙図彙、謡曲画誌、歳旦訓蒙図彙、暗夜訓蒙図彙、外科訓蒙図彙、陰兼陽珍紋図彙、増訓画引和玉図彙、新造図彙、奇妙図彙、戯場楽屋図会、楽屋図会拾遺、戯場訓蒙図彙、花鳥写真図彙、璣訓蒙鏡草、機巧図彙、泰西訓蒙図解、能之訓蒙図彙、改正能訓蒙図彙、好色訓蒙図彙、武具訓蒙図彙、頭書増補訓蒙図彙、女用訓蒙図彙、人倫訓蒙図彙、難字訓蒙図彙、立花訓蒙図彙、仏像図彙、増補仏像図彙、唐土訓蒙図彙、訓蒙図彙、訓蒙図彙大成。

3）杉本つとむ『訓蒙図彙』（早稲田大学出版部、1975年）。

洛中洛外図屏風「歴博甲本」人物データベースについて

国立歴史民俗博物館　小　島　道　裕

１．洛中洛外図屏風「歴博甲本」をどう見せるか

国立歴史民俗博物館が所蔵する、現存最古の洛中洛外図屏風として知られる「歴博甲本」（1525年ころ。スライド２）を、ホームページを媒体としてどのように公開しているかをまず紹介したい。

ホームページに設けられた「WEB ギャラリー」の頁では、「歴博甲本」は、次の５種類の方法で見ることができる（スライド３・４）。

①屏風全体を単純に拡大する高精細画像（ブラウザ）
②描かれた名所などの事物に説明を加えた読み解き画像
③テーマごとに名所などの個別画像を切り出して解説する「電子企画展」版
④パワーポイントのスライドショーと音声解説によるｅラーニング版
⑤描かれた人物（1426人）一人一人の、画像と属性などのデータを検索できる「人物データベース」

今回の報告では、この中から、⑤の「人物データベース」について説明する。

２．洛中洛外図屏風「歴博甲本」人物データベースによる検索

まず、実際にどのような検索が行え、どのようなデータが見られるかをご覧いただきたい。

スライド５はオープニングの検索画面であり、「性別」以下の項目に対して、最大７つのキーワードを表示している（スライド６はその部分拡大）。任意の語句を左下の空欄に入れて検索することもできるのだが、一般の利用者にとっては、あらかじめキーワードを例示した方が利用しやすいと思われ、特にこのデータベースは、2012年の企画展示「洛中洛外図屏風と風俗画」に際して、コンテンツのひとつとして作られたものであるため、専門家以外の利用を強く意識している。

実際に検索した場合の例を示す。「性別」から「男」を選択すると、スライド７のような結果を表示したサムネイル画面になる。その中から検索者が興味を持った人物、たとえば中央の団扇を持った人物（矢印）を選択すると、スライド８のような、個別人物データの画面になる。

上端の右に書かれた、「甲＿右＿１_68」という表示は、この人物の ID であり、「甲本」「右隻」「第１扇」の（右上から数えて）68番目の人物、ということになる。その下にある屏風の模式図で、どの部分に描かれているかも分かる。（各扇を上中下に分割したこの模式図も、そのまま検索項目になっている。）

その下の表が各項目のデータであり、この場合、「性別」は男、「身分・職業等」は不詳、「服装」は小袖、「被り物」は無し、「髪型」は「たぶさ髪」（単純に後頭部で結わえたもの）、

「髭」は有り（口髭）、「持ち物」は団扇、「場所」は、祇園会の長刀鉾（の付近）、となり、「行為」は、「長刀鉾のそばで拍子を取る。」「団扇を持つ。」と記述されている。「備考」は特に記述がない。（ある場合は、ここでは省略した「その他」の記述や、研究的な付加情報などを記している。）

左側の人物画像はブラウザーになっているので、表示された倍率から縮小して、周囲に何が描かれているのかを直接見ることも可能である。この場合、記述にあった、祇園祭の長刀鉾のそばで拍子を取っていることを、画面でも確認することができる。

このデータ表示画面からも検索を行うことが可能であり、「性別」から「場所」までの欄は、そのまま表示されたキーワードが検索される。またここでも表の下方に検索語句の入力欄があるので、それによって検索することもできる。例えば「鉾」で検索すると、結果は、またサムネイル画面で表示される（スライド９）。

そこからまた関心を持った人物、たとえば左下の横笛を吹く人物（矢印）を選択すると、スライド10のように、個別人物のデータと画像が、また表示される。

このように、例示されたキーワードから、関心に従って次々と検索を進めていくことができるが、最初から特定のキーワードで検索をかけることももちろん可能で、たとえば、「絵師」という語句で検索してみると（スライド11）、一件だけがヒットし、個別人物のデータとしては、スライド12のようなものになる。「行為」や「備考」欄には、現在も「狩野元信邸跡」として知られる場所であることや、顔に後補があるらしいことなどの研究的情報が盛り込まれている。ちなみに、画像をブラウザ機能で縮小して、屋敷の全体を見るとスライド13のようになるが、これは現地ではスライド14のような碑が立っている、狩野元信邸跡と一致する。将来的には、このような位置情報を加えることも考えられよう。

検索機能としては、複数の語句による検索も可能なので、たとえば、表紙の検索画面に使用している被衣（かずき）を被った女性について知りたい場合は、スライド15のように「女」と「被衣」という二つの語句を入力すると、スライド16のような結果となるので、サムネイル画像から知りたかった人物を選択することができる。スライド17のように、この屏風が伝来した三条家に関わる人物ではないか、という説を紹介している。屏風全体の中では、スライド18・19のように、幕府から外出する場面として描かれた女性である。

３．関連するデータベース

現在、この洛中洛外図屏風「歴博甲本」人物データベースに関連する二つの画像データベースを作成中なので、それについて紹介したい。共に、2016年度中（2017年２月ころ）の公開を予定している。

①　「歴博乙本」人物データベース

ひとつは、国立歴史民俗博物館が所蔵する洛中洛外図屏風のひとつ、「歴博乙本」（1580年代ころ）について、「歴博甲本」と同様に、描かれた人物（1172人）ひとりひとりの画像とデータを検索表示できるデータベースである。

まだ実際の画面をお見せできないが、「歴博乙本」の場面を一つお見せするとスライド21のようであり、ここに見える「傘」をキーワードとして、「歴博甲本」と比較してみたい。「歴博甲本」人物データベースで「傘」を検索するとスライド23のようになり、37件がヒッ

トしているが、描かれているのは、男性、僧侶、尼僧のみであり、「歴博乙本」のような、着飾った女性が色や模様の付いた傘をさしている情景は見られないことが分かる。おそらく、16世紀初めの「歴博甲本」ではまだ身分表象であった傘が、60年ほど後の「歴博乙本」では、身分を離れた一種のファッションとなっているためと思われ、社会や文化の「近世化」の一例であろう、と解釈することができる。

このように、「歴博甲本」と「歴博乙本」という、時代の離れた二つの屏風を同時に見ることで、歴史的な変化などを分析することが可能になる。

②「近世職人画像データベース」

もうひとつの関連データベースは、屏風ではなく、近世の版本に載せられている職人関係の図像を取り出して、書誌データや詞書きなどと共にデータベース化したものである。ここでの「職人」は、狭義のそれではなく、生業という程度の意味で用いている。対象とした版本は、主に国文学研究資料館が所蔵するもので、作業も主に同館が行った。

検索画面はスライド26のようであり、たとえば「フリーワード」の項目に「売り」を入力して検索すると、スライド27のような、一覧表形式のサムネイル画面になる。「飴売り」を選択すると、スライド28のような個別データ画面になる。記述されている項目は、主に掲載されている版本の書誌に関するもので、絵の内容については、項目を分けずに「キーワード」で表示している。対象が必ずしもひとりひとりの人物ではなく、多くが複数の人物を描いた場面単位の挿絵であることがひとつの理由であり、服装・持ち物などについても、たとえば「小袖」などのようにすべてを記述するのではなく、特徴のある情報のみを記載している。

画像をクリックすれば、スライド29のように画像を拡大することができ、これもブラウザを実装しているので、さらに拡大することも可能である。

データとして記載した内容として、画中に見える「詞書」も取り上げており、スライド30の例のように、全文を翻刻している（画像はスライド31）。この「魚屋」の場合は、一人の人物ではなく、生け簀のある魚屋、というひとつの場面を対象にしているため、個別の人物に関するデータは多くない。なお、この図では魚についても具体的に描写しているので（スライド32）、蝦、蛸、など判明するものについてはキーワードに記載している。わずかだが、「魚データベース」にもなっていると言えるかもしれない。

統合検索について

以上のような、関連する複数のデータベースをどのように結びつけるかが課題となるが、機構の共通検索システムによって、テキストベースでは横断的な検索が可能になるはずである。画像の表示については難しい面があると思われるが、「歴博甲本」人物データベースと「歴博乙本」人物データベースの二つについては、報告時点では未定であったが、国立歴史民俗博物館のシステム内で、完全な横断検索と表示が可能となり、統合した検索画面も作ることとなった。

参考文献

大薮　海　「洛中洛外図屏風歴博甲本人物データベース各項目の立項方法と入力語」
宮田公佳　「画像・文字情報融合手段としての人物データベース」

（共に、小島道裕編『洛中洛外図屏風歴博甲本の総合的研究』国立歴史民俗博物館研究報告 第180集、2014年２月）
森下佳奈 「洛中洛外図屏風歴博乙本人物データベースの作成と課題」（『国立歴史民俗博物館研究報告』第206集、2017年３月）
大高洋司他 『日本中世・近世の地誌と風俗』（平成21年度総合研究大学院大学文化科学研究科イニシアティブ事業教員学生連携研究「日本中世・近世の地誌と風俗」報告書、2010年、総合研究大学院大学文化科学研究科）
大高洋司・小島道裕・大久保純一編『鍬形蕙斎画 近世職人尽絵詞 ── 江戸の職人と風俗を読み解く ──』（勉世出版、2017年）

第11回人間文化研究情報資源共有化研究会
「人間文化研究機構のもつ画像データ共有化の前進に向けて」
2016年2月6日（土）TKPガーデンシティ京都

洛中洛外図屏風「歴博甲本」人物データベース について

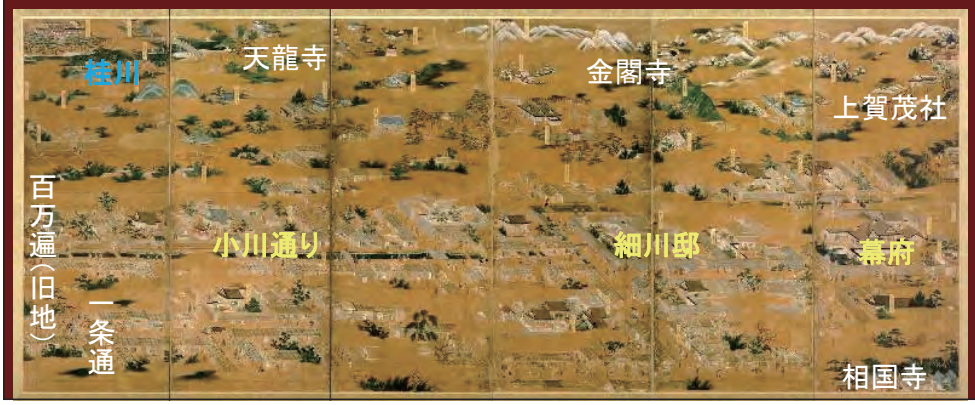
国立歴史民俗博物館
小島 道裕

1

洛中洛外図屏風（歴博甲本　1525年か）六曲一双

右隻
（東隻）

←北



左隻
（西隻）

北→

2

洛中洛外図屏風「歴博甲本」をどう見せるか

WEBギャラリー ｜資料｜研究・教育・資料

HOME ＞ 研究・教育・資料 ＞ 資料 ＞ WEBギャラリー ＞ 資料形態での分類

資料形態での分類

このページの目次

WEBギャラリーについて　屏風　絵図　絵巻　絵画　古文書　拓本　おことわり

ー WEBギャラリーについて

博物館収蔵の資料をHP上で気軽にご覧頂けます。画質を優先にしておりますので、表示に時間がかかることもございますが、予めご了承ください。

1. 〔高精細画像Flash版〕は、資料を高精細画像でご覧いただけます。〔高精細画像Flash版〕と〔説明入り画像Flash版〕の閲覧にはFlash Player が必要となります。Flash Playerに関する情報はこちらをご覧下さい。
2. 〔高精細画像Flash版〕および〔説明入り画像Flash版〕がご覧になれない場合は、〔高精細画像順次拡大版〕をご覧下さい。
3. 〔電子企画展版〕は、館蔵品から興味深い資料やテーマを選び、教員の独自の切り口で紹介するページです。

PAGE TOP

3

洛中洛外図屏風（歴博甲本）　【重要文化財】　[16世紀前期]

現存最古の洛中洛外図屏風です

- 洛中洛外図屏風（歴博甲本）左隻 〔高精細画像Flash版〕
- 洛中洛外図屏風（歴博甲本）左隻 〔説明入り画像Flash版〕
- 洛中洛外図屏風（歴博甲本）右隻 〔高精細画像Flash版〕
- 洛中洛外図屏風（歴博甲本）右隻 〔説明入り画像Flash版〕
- 洛中洛外図屏風（歴博甲本）　〔高精細画像順次拡大版〕
- 洛中洛外図屏風（歴博甲本）　〔電子企画展版〕
- 洛中洛外図屏風 「歴博甲本」を読む　前編 〔eラーニング版〕
- 洛中洛外図屏風 「歴博甲本」を読む　後編 〔eラーニング版〕
- 洛中洛外図屏風 歴博甲本 人物データベース

〔eラーニング版〕について

洛中洛外図屏風歴博甲本（重要文化財）を、当館の小島道裕教授（歴史研究系）が画像を見ながら読み解いていきます。現在の京都の地図をチェックしながら見ると、屏風との対比がしやすくなります。閲覧にはFlash Player が必要となります。Flash Playerに関する情報はこちらをご覧下さい。

※この教材は総合研究大学院大学（SOKENDAI）の特別教育研究経費（広い視野を有する博士育成のためのテーラーメイド教育システムの構築事業）の支援を受けて制作しています。

4

5

せいべつ 性別	おとこ 男	おんな 女	こども 子供	あかんぼう 赤ん坊			
みぶんなど 身分等	くげ 公家	ぶし 武士	そうりょ 僧侶	にそう 尼僧	のうみん 農民	いぬじにん 犬神人	かよちょう 駕輿丁
ふくそう 服装	かりぎぬ 狩衣	ひたたれがた 直垂型	かたぎぬばかま 肩衣袴	ほうい 法衣	こそで 小袖	どうぶく 胴服	つけひも 付紐
かぶりもの 被り物	えぼし 烏帽子	あみがさ 編笠	ぬりがさ 塗笠	かずき 被衣	ぬの 布	ずきん 頭巾	かぶと 兜
かみがた 髪型	たぶさがみ たぶさ髪	ていはつ 剃髪	すいはつ 垂髪	たばねがみ 束ね髪	ほうはつ 放髪	ほうはつ 蓬髪	まげ 髷
ひげ 髭	くちひげ 口髭	あごひげ 顎髭	ほおひげ 頬髭				
もちもの 持ち物	かたな 刀	おうぎ 扇	やり 槍	かさ 傘	ふくろ 袋	つえ 杖	おうご 杓
ばしょ 場所	だいり 内裏	ばくふ 幕府	てい 邸	けいだい 境内	とおり 通	たんぼ 田圃	かわ 川
こうい 行為	あるく 歩く	すわる 座る	みる 見る	まつ 待つ	のむ 飲む	かつぐ 担ぐ	おがむ 拝む
そのた その他	こうほ 後補	ほひつ 補筆	はくめん 白面	さげお 下げ緒	させき 左隻	うせき 右隻	

2016年1月25日曜日

6

「男」に該当する人物の一覧：1117人
◀◀最初の一覧
◀前の一覧
一覧ページ　2 / 28
次の一覧▶
最後の一覧▶▶
トップへ
一覧から，見たい人物を選んで下さい
国立歴史民俗博物館
National Museum of Japanese History

7

「男」に該当する人物情報　1117人の60人目：甲_右_1_68
左隻（させき）
右隻（うせき）
第6層 第5層 第4層 第3層 第2層 第1層
中
水色の部分をクリックすると再検索できます
性　別
男
身分・職業等
-
服　装
小袖
被り物
無
髪　型
たぶさ髪
髭
有（口髭）
持ち物
団扇
場　所
長刀鉾
行　為
• 長刀鉾のそばで拍子を取る。
• 団扇を持つ。
備　考
-
検索実行
リセット
トップへ
再表示
◀前の人物
▲一覧へ
次の人物▶
国立歴史民俗博物館
National Museum of Japanese History

8

「鉾」に該当する人物の一覧：88人
◀◀最初の一覧
◀前の一覧
1
2
3
次の一覧▶
最後の一覧▶▶

9

「鉾」に該当する人物情報　88人の74人目：甲_右_2_87
左隻（させき）
第6層 第5層 第4層 第3層 第2層 第1層
右隻（うせき）
第6層 第5層 第4層 第3層 第2層 第1層
中
水色の部分をクリックすると再検索できます
性　別
男（子供）
身分・職業等
-
服　装
小袖
被り物
無
髪　型
束ね髪
髭
無
持ち物
横笛
場　所
函谷鉾
行　為
• 函谷鉾に乗る。
• 横笛を吹く。
備　考
• 白面

10

洛中洛外図屏風「歴博甲本」人物データベース
－「歴博甲本」に登場する1426人の人物像について，キーワードで情報を検索できます－
1426人の全てを表示
せいべつ 性別 おとこ 男 おんな 女 こども 子供 あかんぼう 赤ん坊
みぶんなど 身分等 くげ 公家 ぶし 武士 そうりょ 僧侶 にそう 尼僧 のうみん 農民 いぬじにん 犬神人 かよちょう 駕輿丁
ふくそう 服装 かりぎぬ 狩衣 ひたたれがた 直垂型 かたぎぬばかま 肩衣袴 ほうい 法衣 こそで 小袖 どうぶく 胴服 つけひも 付紐
かぶりもの 被り物 えぼし 烏帽子 あみがさ 編笠 ぬりがさ 塗笠 かずき 被衣 ぬの 布 ずきん 頭巾 かぶと 兜
かみがた 髪型 たぶさがみ たぶさ髪 ていはつ 剃髪 すいはつ 垂髪 たばねがみ 束ね髪 ほうはつ 放髪 ほうはつ 蓬髪 まげ 髷
ひげ 髭 くちひげ 口髭 あごひげ 顎髭 ほおひげ 頬髭
もちもの 持ち物 かたな 刀 おうぎ 扇 やり 槍 かさ 傘 ふくろ 袋 つえ 杖 おうご 杓
ばしょ 場所 だいり 内裏 ばくふ 幕府 てい 邸 けいだい 境内 とおり 通 たんぼ 田圃 かわ 川
こうい 行為 あるく 歩く すわる 座る みる 見る まつ 待つ のむ 飲む かつぐ 担ぐ おがむ 拝む
そのた その他 こうほ 後補 ほひつ 補筆 はくめん 白面 さげお 下げ緒 させき 左隻 うせき 右隻
甲本 左隻 第1扇 於：幕府（柳の御所）東通り
キーワードを選択すると関連する画像が見られます
絵師 絵師
検索実行
リセット
とじる
国立歴史民俗博物館
National Museum of Japanese History

11

「絵師」に該当する人物情報 1人の1人目：甲_左_5_80
左隻（させき）
第6層 第5層 第4層 第3層 第2層 第1層
下
右隻（うせき）
第6層 第5層 第4層 第3層 第2層 第1層
水色の部分をクリックすると再検索できます
性別 男
身分・職業等 絵師
服装 小袖カ
被り物 （見えず）
髪型 （見えず）
髭 無
持ち物 絵筆
場所 誓願寺辻子（狩野辻子）の狩野邸内
行為
• 誓願寺辻子（狩野辻子）にある狩野邸内で座る。
• 扇に絵を描く。
備考
• 顔などに後補あるカ。
• 狩野元信カ。
検索実行
リセット
トップへ
再表示
-
▲一覧へ
-
国立歴史民俗博物館
National Museum of Japanese History

12

「絵師」に該当する人物情報　1人の1人目：甲_左_5_80
左隻（させき）
第6層 第5層 第4層 第3層 第2層 第1層
下
右隻（うせき）
第6層 第5層 第4層 第3層 第2層 第1層
水色の部分をクリックすると再検索できます
性別
男
身分・職業等
絵師
服装
小袖カ
被り物
（見えず）
髪型
（見えず）
髭
無
持ち物
絵筆
場所
誓願寺辻子（狩野辻子）の狩野邸内
行為
• 誓願寺辻子（狩野辻子）にある狩野邸内で座る。
• 扇に絵を描く。
備考
• 顔などに後補あるカ。
• 狩野元信カ。
検索実行
リセット
トップへ
再表示
-
▲一覧へ
-
国立歴史民俗博物館
National Museum of Japanese History

13

狩野元信 屋敷跡 の碑
此附近 狩野元信屋敷
（地理情報も入れられるかもしれない）

14

－「歴博甲本」に登場する1426人の人物像について，キーワードで情報を検索できます－
1426人の全てを表示
性別 男 女 子供 赤ん坊
身分等 公家 武士 僧侶 尼僧 農民 犬神人 駕輿丁
服装 狩衣 直垂型 肩衣袴 法衣 小袖 胴服 付紐
被り物 烏帽子 編笠 塗笠 被衣 布 頭巾 兜
髪型 たぶさ髪 剃髪 垂髪 束ね髪 放髪 蓬髪 髷
髭 口髭 顎髭 頬髭
持ち物 刀 扇 槍 傘 袋 杖 杓
場所 内裏 幕府 邸 境内 通 田圃 川
行為 歩く 座る 見る 待つ 飲む 担ぐ 拝む
その他 後補 補筆 白面 下げ緒 左隻 右隻
甲本 左隻 第1扇 於：幕府（柳の御所）東通り
キーワードを選択すると関連する画像が見られます
女 被衣
検索実行
リセット
とじる

「 女 被衣 」で検索

15

「女」に該当する，かつ，「被衣」に該当する人物の一覧：102人
◀◀最初の一覧
◀前の一覧
1
2
3
次の一覧▶
最後の一覧▶▶

16

17

「公方様」（幕府＝将軍邸「柳の御所」）門前から外出する女性の一行

18

三条家息女　武家の懇望で上臈となる
1525年（大永5）12月13日（義晴入居と同日）新御所に輿入れ

服装は、小袖・被衣（かづき、まだ小袖的）、垂髪

19

開発中の関連データベース

①「歴博乙本」人物DB

1580年代ころ？
作者は不明だが、狩野松栄工房か
登場人物：1172人

入力作業は終了　2016年度公開予定

※機構連携研究「都市風俗と『職人』」の成果の一部

20

新しい風俗の例：女性の傘
→「歴博甲本」と比較すると？

21

企画展示「洛中洛外図屏風と風俗画」　－歴博甲本　登場人物コレクション－

い人物を選んで下さい.

1426人の全てを表示

性　別	男	女	子供	赤ん坊			
身分・職業等	僧侶	尼僧	武士	農民	犬神人	駕輿丁	貴族
服　装	小袖	肩衣袴	法衣	直垂	袴	付紐	胴服
被り物	笠	帽子	被衣	頭巾	兜	冠	布
髪　型	たぶさ髪	剃髪	垂髪	束ね髪	放髪	逆髪	髢
髭	口髭	顎髭	頬髭				
持ち物	刀	扇	槍	傘	桶	団扇	杖
場　所	幕府	内裏	境内	門前	小路	田圃	川
行　為	担ぐ	歩く	座る	見る	待つ	飲む	拝む

「歴博甲本」人物データベース　で　キーワード「傘」を検索

22

甲本人物データベース 「傘」の表示結果
（男性、僧侶、尼僧のみ）

23

「歴博乙本」の新しい風俗への関心が分かる
身分表象としての持ち物が、ファッションに （近世化）

「甲本DB」と同様の検索項目を予定（※歴博DBで横断検索可能）
機構共有DBを通して、「甲本DB」「近世職人画像DB」などとも共通検索を可能にする予定

24

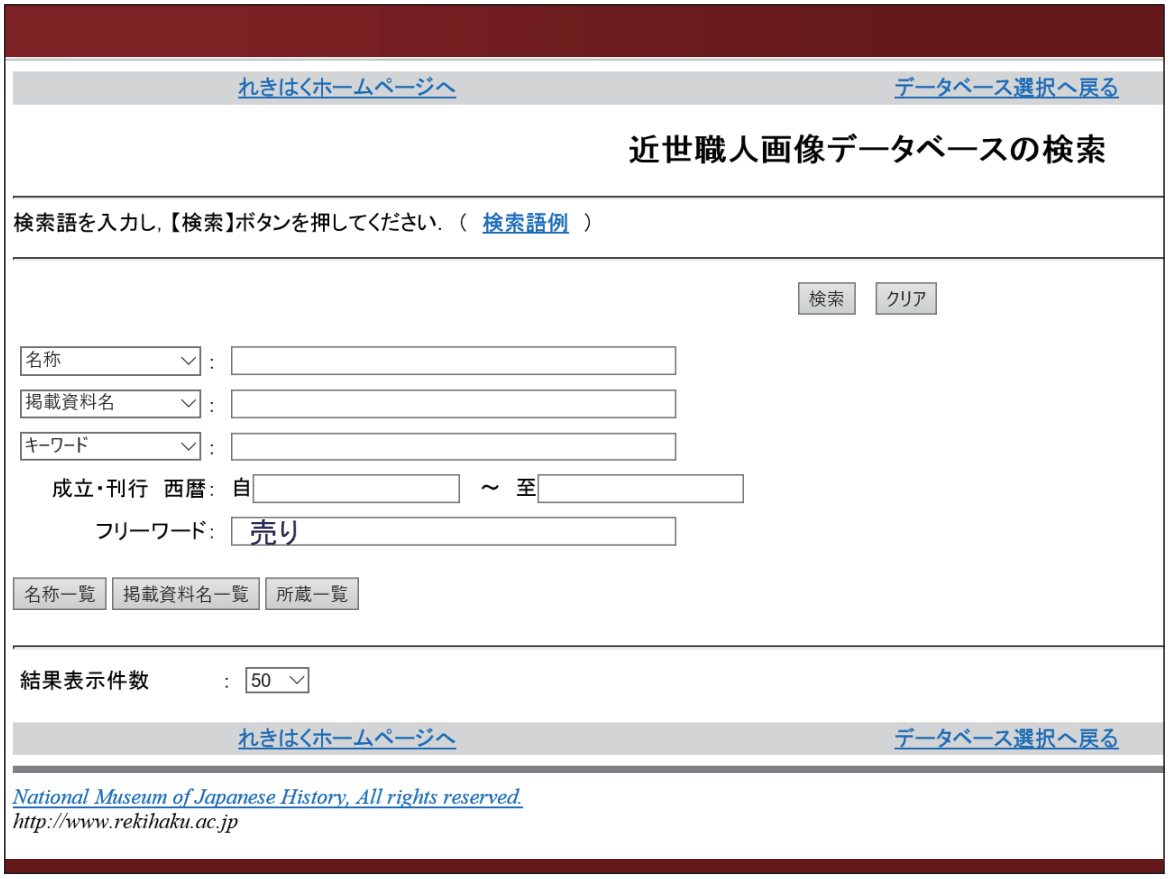
開発中の関連データベース

②近世職人画像データベース

国文研等所蔵の近世版本の所収画像から、「職人」関係の図像を取り出し、詞書きなどと共にDB化

約3,000件　システム開発中
2016年度早期に公開予定
※機構連携研究「都市風俗と『職人』」の成果の一部

25



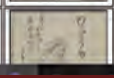
26

検索条件:フリーワード=売り
検索結果:4件データが見つかりました。

ソート　絞り込み検索　検索履歴・AND/OR検索

番号をクリックすると詳細が見られます.

1～4)/4件数

番号	画像	名称(名称ふりがな)	掲載資料名	成立・刊行	成立・刊行　西暦	所蔵
1		扇売り(おうぎうり)	京童	明暦4年刊	1658	国文学研究資料館
2		飴売り(あめうり)	四季交加	寛政10年刊	1798	個人蔵
3		魚屋(さかなや)	摂津名所図会(修訂再印本)	寛政10年	1798	国文学研究資料館
4		綿売り(わたうり)	七十一番職人歌合(刊本)	明暦3年刊	1657	国文学研究資料館

27

画像をクリックすると拡大画像を表示します。

【名称】　あめうり 飴売り
【掲載資料名】　しきのゆきかい 四季交加
【所蔵】　個人蔵
【成立・刊行】　寛政10年刊
【成立・刊行　西暦】1798
【巻冊等】　上巻 01
【丁付】　4ウ
【詞書】　「二月」
【キーワード】　江戸、京橋、春、二月、男、唐人飴、朝鮮飴、担い売り、天秤棒、笠、チャルメラ、脚絆、草鞋、羽織
【掲載資料書誌】　2巻2冊。風俗絵本。山東京伝著・北尾重政画。〈江戸〉鶴屋喜右衛門刊。
【備考】　–

絵の内容については、項目を分けずに、「キーワード」で検索。他のDBとも連動、の予定。

次へ　一覧へ戻る

28

飴売り

29

【名称】 さかなや
魚屋

【掲載資料名】 せっつめいしょずえ(しゅうていさいいんばん)
摂津名所図会(修訂再印本)

【所蔵】 国文学研究資料館

【成立・刊行】 寛政10年

【成立・刊行　西暦】 1798

【巻冊等】 8上巻10冊 10

【丁付】 34ウ35オ

【詞書】 「兵庫の生洲(いけす)は今在家(さいけ)町の浜濱にあり。常に諸魚を生て湿気(しけ)の時不漁(ふぎよ)の用とす。 より捧(さゝぐ)る事多し。」
36オ「兵庫生洲(ひやうごのいけす)　当津南浜今在家町にあり。長サ十三間巾(はゞ)四間計(ばかり)、四方を圍(かこ
中に潮(うしほ)水を湛(たゝ)へ鯛(たい)鱧(はも)鱚(すゞき)及び諸魚(しよぎよ)を多く放生(はなちいけ)て常に貯(た
ふ。禁裏(きんり)調進(てうしん)の手当(てあて)とす。往来(ゆきゝき)の旅人(りよしん)こゝに来ッて目(め)を悦(よろこ

【キーワード】 摂津、矢田部郡、八部郡、生洲、生け簀、、漁師、魚屋、魚売り、たも網、鯛、鱧(はも)、鱚(すずき)、蛸(たこ)、えび

【掲載資料書誌】 請求記号ヤ-6-128-1～12
本作初印本の書誌事項は別掲のとおりだが、寛政6-7年頃絵師竹原春朝斎に故障が起こったと見られる。その結果、
工が分担しているが、寛政10年刊の巻1-6下では、春朝斎の欠を丹羽桃渓が一人で補っている。これに伴って、巻7-9
る。

【備考】 修訂再印本では、初印本の秀雪亭画を丹羽桃渓画に差し替える。

30

↓ 詞書き

31

Database Search Results
Database Search Results
証明書エラー
dbsvr.rekihaku.ac.jp/up-cgi/getimgrd.pl?tn=08&ti=78&h=./history/w11453718523_52678&ch=08&p=param/kins/db_param&o=08

32

史料編纂所所蔵肖像画模本・歴史絵引データベースの課題

東京大学史料編纂所　藤　原　重　雄

【概要】

人間文化研究機構による統合検索システムnihuINTに属さない機関の研究者として、「画像データ共有化の前進」が検討されるにあたり、東京大学史料編纂所から公開している画像データベースの一部について、その概要と課題とを紹介して参考事例に供するとともに、日本中世絵画史料の研究者として、史資料画像の記述や〈名づけ〉について、参考となる営みを紹介しつつ、研究実践上の方法的な課題を述べた。

1997年に設立された附属施設である画像史料解析センターでは、前近代日本に関わる視覚的な諸史料、もしくは文献史料の視覚的側面に即した調査研究をプロジェクト形式にて蓄積し、成果のデータベースによる公開も進めてきた。

このうち、「史料編纂所所蔵肖像画模本データベース」は、主に戦前期に作成した肖像画の模本約900点につき、画像とともに書誌・像主・賛の翻刻・描写の記述・関連論文等の研究情報を付加したものである。ある水準での完了をみたが、専従の研究者が必要となる画像の記述など、充実させるべき要素も多く残る。これを補完するのが「肖像情報データベース」で、模本として所蔵しない絵画・彫刻・写真をも把握すべく、自治体史・図録等で掲載・言及された人物肖像につき、典拠ごとに一点づつテキストデータとしたものである（多数の掲載がある肖像画はそれだけ件数が増える）。随時データを更新している。

「歴史絵引データベース」は、肖像画の理解に必要な服飾史を中心とする学習・知識提供型のデータベースとして当初は構築されたが、費用やシステム維持の面から機能を縮小し、現在は収載データの再点検を進めている。トレース図に名所（などころ）・解説を付した図解と、図典類刊行物の図版の語彙による索引との二本立てになっている。

ところで、絵画を歴史的な知識・認識の形成に援用すること、とりわけ事実をありのまま描いたように利用するかに見えた歴史家の態度は、美術史家からの厳しい批判にさらされ、絵画を歴史の理解に資するための素材として用いる方法は、その実践を通じて深められてきた。また、西洋・東洋・日本の図像を扱う学問分野では、それぞれに〈名づけ〉を営んできた。しかるに今回NIJL-NW projectより披露された「画像へのタグ付け」は、図像学の蓄積はもとより、降っては1980年代末からの論争を一顧だにしている気配がない。

現在、人間文化機構の諸機関が公開する画像データにつき、安易な統合検索を進めることは、人文学が版築してきた知の土台を突き崩しかねない。まずは特徴的なコレクションそれ自体の性格に即して、適切な画像データを形成し、共通フォーマットに載せうるメタデータと、専門分野に即した記述とを充実させるのが先決である。視覚的資料に〈名づけ〉することへの方法的な思索に加え、インターネットを通じて広く画像を公開することへの省察も必要となろう。調査研究による対象の選定や、書誌情報・インデックスの付与をしないまま、膨大な画像をWeb公開してしまう方式は、国立国会図書館が先陣を切り、国文

学研究資料館がNIJL-NW projectに飲み込まれる形で遂行しているように見受けられる。Webを介して社会と向き合うことを得手とはしない人文的諸学といえども、研究活動ないし研究資源・成果と社会との動的な作用については、文化人類学・民俗学や考古学、あるいは抜き差し難い非対称性下にある異文化研究などにおいて、既に深い経験と考察の蓄積を備えており、人間文化研究機構の叡智を結集すべき課題とも思える。

史料編纂所所蔵肖像画模本／歴史絵引データベースの課題

藤原重雄（東京大学史料編纂所）

●はじめに

○個人研究領域

日本中世史。文化史・社会史。日本中世における絵画史料論（12～16 世紀頃：年中行事絵巻～上杉本）。

国際日本文化研究センター　共同研究「デジタル環境が創成する古典画像資料研究の新時代」研究会

「中世史研究における史料画像のデジタル環境―個人的観察―」（2012 年 5 月 13 日、史料編纂所）

「史料校訂に関わるデジタル環境」（楊暁捷・小松和彦・荒木浩編『デジタル人文学のすすめ』勉誠出版、2013 年）

○「編纂」業務の基本的な理解

直接的には、『大日本史料』『大日本古文書』等の史料集の刊行。

歴史学研究の素材＝史料を調査研究し、諸レベルでの利用可能とする。史料採訪・史料情報の提供（調査撮影・保存修復・目録書誌解題・翻刻校訂編纂・写真帳・データベース・画像公開）。

現下の状況。人員減・予算減・業務増・後継者難。煩瑣・細切れの説明・成果。組織・事業（および関係者）の突然死へと至らない、緩やかな縮小への着地の模索。

デジタルアーカイブ・データベース・Web 公開についての、かくある「べき」論の無効性。

各史料が置かれた状況に寄り添った多様性。優先度の自己決定、視野とバランス感覚。（←別の「べき」論…。）

目的や思考ベクトルの相違：理念や理想的状態の実現／史料・作品の理解を深めること、それらモノあるいはまつわる知識・情報を将来に伝えること。作品か、イメージ情報か。

●画像史料解析センター

○東京大学史料編纂所（東京大学の附置研）に附属する施設

所内公募のプロジェクトによる個別の調査研究、関連 DB の維持。

「解析」はやや誤解を呼び込む。1997 年の発足時は電算系業務も担当

（→2006 年、前近代日本史情報国際センター）。

前近代日本史研究の視点から、対象となる史料を視野に収め、基幹となる史料ジャンルを扱う。

肖像画、荘園絵図、古写真、文字史料の視覚的・物質的側面（花押・くずし字）。

金石文拓本、近世絵図、錦絵・刷物、屏風絵、ガラス乾板・採訪記録、その他。

→井上聡「東京大学史料編纂所「電子くずし字字典データベース」の概要と展望」（『情報の科学と技術』65-4、2015 年：PDF あり）、同「史料編纂所所蔵荘園絵図模本データベースと地理情報蓄積システムの連携について」（『画像史料解析センター通信』72、2016 年）ほか。

具体的な調査対象は区々であるが、史料類型ごとに何らかの研究を継続してゆくことで、特定課題の研究に携わるセンターとしての全体性を意識した PJ の認定・予算配分。科研費の申請。

所蔵する画像史料（各種の複製：模本・写真など）の調査・保存・公開を軸にした研究活動。

○画像関係 DB・閲覧システムと公開範囲（報告者が些か関係したもの）　（DB 一覧）

荘園絵図模本 DB、肖像画模本 DB、歴史絵引 DB 他（Web 公開／未校正データ／未登録）
一遍聖絵閲覧システム（来館利用のみ、メンバーにもデータなし）　2005 年 9 月公開
洛中洛外図・合戦図屏風閲覧システム（コアメンバー/メンバーに DVD/展示用タッチスクリーン/来館）
東郷荘絵図総合閲覧システム（DVD 一般配布）
近世武家儀礼指図（DVD 配布：インデックス付き／所蔵史料目録 DB：画像・簡単な細目）
鷹司本「年中行事絵巻」（本所公開保留→書陵部より画像 Web 公開：所蔵者指定業者による撮影）
所蔵史料目録 DB、Hi-cat plus（生成した画像を公開するには簡便：所蔵史料の Web 公開／他機関所蔵史料の来館利用）

○閲覧・公開システムにおける情報・機能付与
閲覧のみ／関連情報の付与／検索機能の高度化／画像利用の仕掛けの組み込み（→プラットフォーム維持のコストが上がる）
スタンド・アローン（単体利用）、来館利用の評価・位置づけ
「マイクロフィルムの閲覧」は研究手段として定着したが、「デジタル画像の閲覧」は定着しない。
東博・東文研などでも、図書室でのデジタル画像閲覧、高精細画像閲覧の利用頻度は低い。
事例が増えて一定のゾーンを作るか。一方で、アクセス可能性は段階的・階層的にならざるを得ない。
〈公開＝Web 公開〉への流れ・環境醸成。相反するリスク：史料の秘匿、閲覧・撮影の謝絶・高額化。
理念に基づく「正しい」主張は、史料の利用・公開と後世への継承を疎外する方向へも働き得る。

●肖像画模本／歴史絵引 DB の概要と成り立ち
○概要
・史料編纂所所蔵肖像画模本 DB　（ヘルプ）　（スライド）
主に明治末年から昭和戦前期にかけて作成した模本類のうち、肖像画約 900 点を対象。
画像データは所蔵史料目録 DB と共通。
模本の書誌的な情報、像主に関する情報、描かれた内容・画賛等、参考文献。
没年、男女子供・法体、身分、賛の有無、キーワードからの検索・絞り込み。
像主略伝は「知の開放」向けの便宜的なデータで、本来は公開停止（削除）すべきもの。
画像の記述（ディスクリプション）が一番難しい。肖像画は定型化しており、項目による単語登録で簡易化して全体に付与。
模本の問題点：Web 公開画像としては模本が原本に先行してしまう。
模本のメリット：一度人の眼と手を通して描く。解釈・整理がなされる（ミスリードもありうる）。
精細画像化の必要性：肖像画は服飾史の資料。文様などの細部。賛・印章が判読できる画像サイズ。

・肖像情報 DB　（ヘルプ）　（スライド）
全国に所在する肖像画・肖像彫刻・肖像写真に関する情報蓄積。
自治体史・辞典・図録から、出典ごとに作品 1 点づつデータ化（出典の複数ページからまとめる）。
カード取り。像主、作品名、出典、本文項目（掲載章段）、所蔵者、材質法量、文字情報有無、図版有無（カラー・モノクロ）、裏面にコピー貼付。

DB 化の際に、名寄せ・作品寄せ（作品目録化）を試みたが、未了のままで、未公開データ多数。

・歴史絵引 DB　（ヘルプ）　（スライド）

画像検索：服飾史・有職故実関係の電子的図典。トレース図の作成、名所（などころ、イメージネーミング）、長めの解説文。知識提供型（史料編纂所 DB では例外的）。

キャプション検索：工具書を中心とした関連書籍の口絵・挿図の横断索引。図中注記・図に対するキャプションもキーワード検索可能。図版自体は持たない。（『古事類苑』は入力休止：日文研）

※当初は画像検索と語彙検索とが循環する仕掛けとし、利用者による自学自習に供することを目標とした（旧版ヘルプ参照）。

ぐるぐると廻るが、全体像が見えないため、自分の居場所が分からなくなるデメリット。

主題ごとの書籍の方が入門には適切。画像検索→キャプション検索のみを残す。

○経緯

1992～95 年度科学研究費一般研究 (A)「中世・近世肖像画の調査・データベース化と歴史図像学的研究」（研究代表者・黒田日出男）

研究成果報告書（1996 年 3 月）：肖像画模本全点の目録、半数（僧侶を割愛）程度の図版。

肖像画模本約 900 点の調査、撮影（モノクロ）、DB 化（四切焼付→nexusDBnet 画像プロセッサー）。

肖像情報カード 27,000 枚（自治体史・図録・辞典からの情報抽出）→肖像情報 DB へ

黒田編『肖像画を読む』（角川書店、1998 年）←1996 年 3 月「肖像画と歴史学」シンポジウム

1997 年：画像史料解析センターの設立、東京大学創立 120 周年「知の開放」パビリオン・デジタルミュージアム

模本 300 点（武将など）の記述を充実。像主略伝、画賛テキスト、描かれた内容（主に服飾）の記述。

カラー 4 × 5 フィルムによる全点再撮影、DB の Web 公開。

画像センターにおける科研費獲得・PJ 継続

歴史絵引 DB の立ち上げ、高精細画像閲覧・研究システムの作成。拡充期。

（中心となった黒田日出男、高橋典幸、佐多芳彦らの異動）

リプレイス時などに生じたシステム・データ上の障害の復旧、維持可能な機能への縮減。

模本全点の Web 公開画像の精細化（フィルム再スキャン・再撮影）

画賛フルテキストの作成（2015 年度完了予定）

＊くずし字 DB への文字切り出し。未収録の禅林墨蹟系の文字などをカバー。

「e 国宝」など所外リソースから採取の可能性。パーマネントリンク。

歴史絵引 DB「キャプション検索」・肖像情報 DB 用データの蓄積

＊辞典的な史料（図絵・図彙）の電子化。所蔵史料目録 DB、肖像画模本 DB の画像から切り出し。

○現段階の構え

便利だが手の込んだシステム設計よりも、テキストベースで維持・継承な情報化を進める。

言葉にする、電子データにする、定義づける（メタ情報を言語化し、それを備えた画像データにする）。

総合的なものを作ろうとしても、構想倒れに終わる。（情報の密度が薄いか、素材の質がまちまち。）

●〈名付け〉をめぐる問題

千野香織「言葉とイメージ―物語絵画研究の現在―」(『列島の文化史』7、1990 年→『著作集』)

※日本中世史からの絵画への言及の擡頭に対する、日本美術史・絵画史からの反撥の一例。

「言葉とイメージは本来別個のものなのであるから、表されたイメージの全体を言葉に置き換え、両者が過不足なく一対一の対応関係を示すと期待する方が間違っている。」

「この点を特に強調するのは、イメージを言葉に置き換え、その言葉を集積したデータをコンピューターで整理して、言葉からイメージを呼び出す、というシステムが、遠からず利用されるようになるだろうと思われるからである。…しかし、キーボードの操作によってディスプレイ機器の画面上に無機的に並んだ言葉が、実は誰か個人の解釈による著作物であるという意識を、一体何人の研究者が保ち続けられるだろうか。データがコピーされ、その過程で複数の人間がデータを付加するということが繰り返されていけば、もとのデータが個人の著作物であったという色合いはますます薄くなっていく。同時に、誰かがある時点でイメージを言葉に置き換えた、その言葉が、まるでそのイメージの唯一正確な文字データであるような錯覚が、次第に蔓延していくのではなかろうか。コンピューターの操作を繰り返しても文字データそのものが変質してしまうことはないが、それだけに、イメージから言葉に置き換えられた、そもそもの出発点の曖昧さを忘れてはならない。」

※デジタル環境に固有の問題でない事柄をない交ぜにしてしまう議論の甘さ。ただし危惧はそのまま。

黒田日出男「図像の歴史学」(『歴史評論』606、2000 年→『増補　姿としぐさの中世史』平凡社ライブラリー、初出ロナルド・トビ訳 1995 年英文)

「絵画という史料の性質からして、そこに描かれている〈もの〉や〈こと〉を正確に把握することは容易ではない。絵画史料の利用に慎重ないし否定的な研究者たちは、すぐにそれは「絵空事」であると言い兼ねないし、そうした推定・確認についてのいい加減な研究は、絵画史料論や歴史図像学を否定するための恰好の攻撃材料を提供することになる。」

※ことあるごとに「史料批判がなされていない」「方法論が確立していない」という論難。

藤原「絵巻のなかの《伊予簾》」(『月刊百科』407、1996 年)

藤原「軒端の鞠―『絵巻物による日本常民生活絵引』のひとこま―」(『非文字資料研究』14、2006 年:PDF あり)

○画像と言葉との一対一対応を前提とすることの危険性。

絵画の表現は、過去の森羅万象を扱う。画像記述には水準の多層性。

元の作品の性格、制作・受容の文脈。見る側の関心で〈名付け〉は変わる面もある。

〈名付け〉を支える・背景となる体系。シソーラスは一元的でない。

「絵画史料」研究は、美術史学からの視点のずらしによっても成立。

範囲の限定、趣旨・切り口の明確化、対象の制約。(歴博甲本人物 DB は成功例)

○イコノグラフィー、「アート・ドキュメンテーション」との共通性／差異

作品分類：地域／時代／様式／流派／作家

イコノグラフィー：作品を主題・モチーフで分類、記述。

イコノロジー：広範な（比較）文化学の領域へ。imagery 研究。視覚芸術の前提・背景・深層。
多義的な解釈を呼び込む造形。関連概念の網の目。アレゴリー、シンボル、古典主題。
研究の方法、手段・ツールとして始まる。自己目的化しない。　ICONCLASS
ルーロフ・ファン・ストラーテン、鯨井秀伸訳・解説『イコノグラフィー入門』（ブリュッケ、2002 年）
（鯨井「ICONCLASS：イコノグラフィー的分類システム」『情報の科学と技術』58-2、2008 年：PDF あり）
愛知県美術館　コレクション検索「主題で探す」　（スライド）

○東洋画題
『歷代名画記』以下の中国の画論、『本朝画史』以下の日本の画論。分類・階層化も。
斎藤隆三『画題辞典』（1919 年、1925 年、1977 年復刊）　＊五十音順、分類なし。拡張→
金井紫雲編『東洋画題綜覧』11 冊（芸艸堂、1941～43 年、1997 年合冊復刊）
　＊雨・梅・馬など一般名詞に下位項目あるも五十音順。
台北・故宮博物院の所蔵資料 DB（典藏資料庫系統＞書畫典藏資料檢索系統＞主題）　（スライド）
　＊描かれた内容の簡単なキーワードを登録（現状ではキーワード検索はできない）。
　→いま各所蔵機関が地に足着けて進めるに相応しい業務のヒントになるのでは？

○「絵引」
澁澤敬三・神奈川大学日本常民文化研究所編『新版　絵巻物による日本常民生活絵引』（平凡社、1984 年）5 巻
神奈川大学 COE（非文字資料研究センター）編『マルチ言語版　絵巻物による日本常民生活絵引』第 1～3 巻
　『日本近世生活絵引』北海道編、北陸編、東海道編、奄美・沖縄編
　『東アジア生活絵引』中国江南編、朝鮮風俗画編
　『18 世紀ヨーロッパ生活絵引』「都市の暮らしと市門、広場、街路、水辺、橋」
神奈川大学デジタルアーカイブ「『絵引』原画」（日本常民文化研究所）
　＊登録キーワードは『新版』での標題のみ。
渋沢栄一記念財団情報資源センター「実業史錦絵絵引」　＊対象史料が 1 点。

○芸能史関係
ジャンルの特性に沿った細かな項目設定、（Web 以前の）作品記述方法の定型化。
東京都立図書館「TOKYO アーカイブ：浮世絵検索」
早稲田大学演劇博物館「浮世絵閲覧システム」
国立劇場「文化デジタルライブラリー：収蔵資料」

●「画像データ共有化」？
○現段階（5 年 10 年スパン）で横断的検索、システム的な統合については疑問。
モノ・イメージ自体のコレクション、デジタル化された蓄積の偏り・僅少さ。そこに付与されたテキスト
　トはさらに貧弱。インデックスなしで大量の画像 Web 公開の倫理性。
公的機関所蔵、公開を希望する民間等所在の資料の画像データの受け皿。
補助金による文化財修理の報告書作成と画像公開など、個別的に底上げする課題。

5

たとえば、日文研「外像データベース」の情報を、かつての日本・アジアの実情としてそのまま利用できない。慎重な史料批判。欧米所在の土産品・輸出向け商品の画像公開によるオリエンタリズムの垂れ流し。史料的価値・限界に対する認識、史料批判・ヴィジュアルリテラシーの意識が吹っ飛ぶ。

古典文学の江戸期版本の挿図は、受容レベルの問題を扱う際に利用可能（初学者相手に許容される範囲もあろうが）。

google 画像検索の結果表示タイプは、現時点での対象蓄積に照らし、学術的機関が行うべき方向か？

個別 DB の趣旨、素材の文脈を理解できるように提示する、ゲートウェイ的なもので充分。

各 DB の個性・特徴・限界などが、明確に把握される形での連携。

答え探しの DB でなく、素材探しの DB。「気づきのきっかけ」としての検索結果を期待する程度でよい。

○画像 Web 公開の罠

予算獲得・執行、数値目標・評価、Web 上の評判。

原本史資料を、研究者による特権的な利用から解放するものであったとしても、史資料と社会とを媒介する能力を備えた利用者へ向き合ってはいない施策。

原本へアクセス可能な「資格・立場」から、「能力・関心」への転換・移行がうまくいっていない。

DB なり Web を介し、直接に市民個々人へ利用を呼びかけるべきものなのか。

研究・著作・教育・展示などを通して、市民社会へと戻し、教えを得る。学習・展示コンテンツの制作。

誰もがその気になれば、典拠・詳細データへと立ち戻って検証できる、アクセス可能性を支える。

研究で集めた素材は、原史料であれ書籍であれデータであれ、整理・保存をしてゆく体制の充実。

DB を作るのでなくて、素材を利用可能にする形態のひとつとして、Web ベースの仕事。

調査研究保存なき公開活用利用はない。

画像公開が研究や新しい需要を生みだす側面もあるが、偶然に頼り、場当たり的。

視覚的資料を教育的・啓蒙的活動の最前線に動員することの弊害。

藤原「画像資料と歴史研究・叙述・教育」（『岩波講座　日本歴史』別巻二・史料論、岩波書店、2015 年）

【付記】　2016 年 2 月 10 日記／9 月 3 日 Web へアップ

・当日配布資料の字句を若干修正し、リンクを付加した。

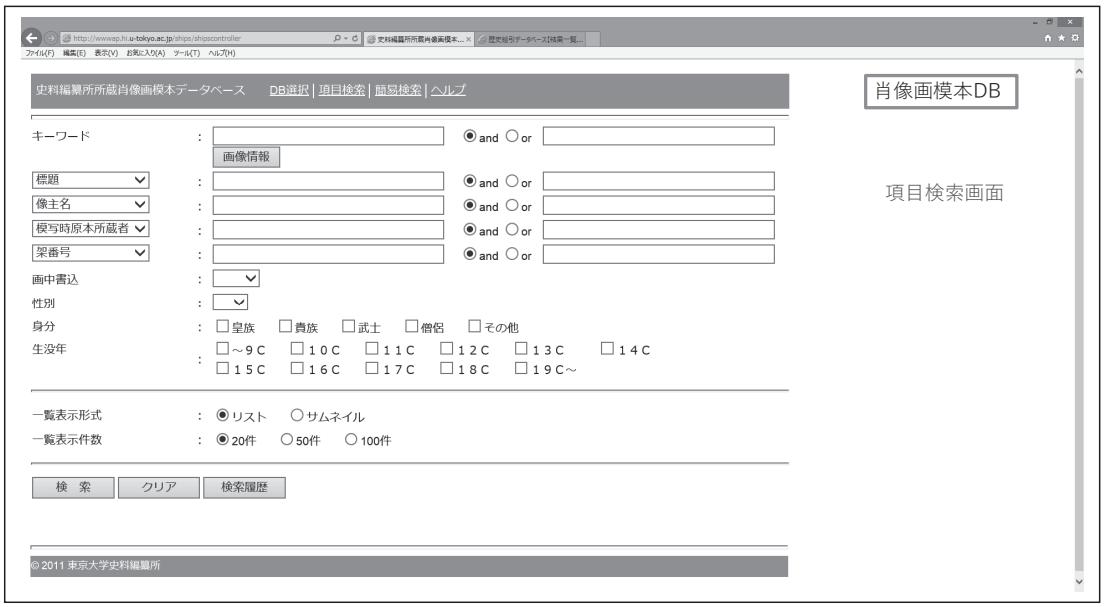
・実際の口頭報告では、5 頁以下の最後の項目については、ごく簡単に、google 画像検索の検索結果表示のような一覧を、人間文化研究機構のデータベースから提供することへの強い危惧を述べるに止めた。

・むろんこの部分では、国文学研究資料館（および国立国会図書館）における大規模デジタル化事業を念頭に置いているが、現時点における日本文化研究に関するデジタル環境の一般的な論点ともなりうると考える。

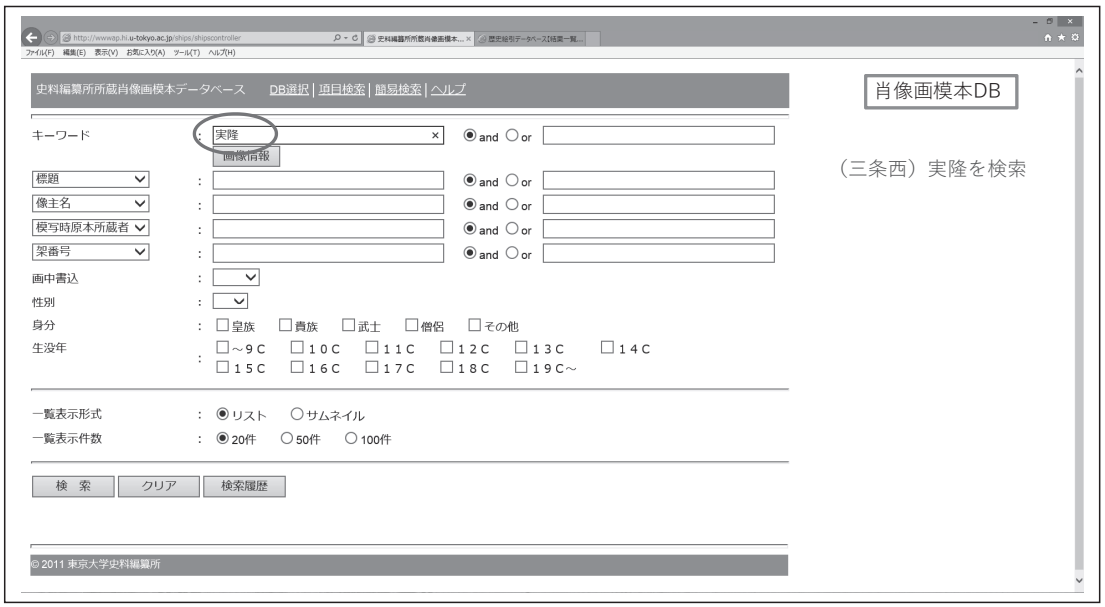
国文学研究資料館からの資料画像 Web 公開については、「周知の如く，システムの能力の問題であろうか，動作速度が実用に堪えられなくっており，評判が芳しくないのは残念である。だだし，これもすぐに改善されることとなろう。」（赤間亮「立命館大学アート・リサーチセンター の古典籍デジタル化―ARC 国際モデルについて―」『情報の科学と技術』65-4、2015 年 4 月）とあるように、2016 年 1 月 29 日のリプレースによって解消されたと思しいが、それ以前の状況においても、実用環境を顧慮せずに公開画像の新規追加が続けられ、現在も例えば、『源氏物語』全帖（書陵部[伏 204]本 2518 コマなど）や『続群書類従』（一部欠で 4644＋30285 コマ）が同一フォルダである。実際に DB を利用することのない評価者・世間の評判に向かっているのではないか、あるいはもはや相手は人間ですらなく、計画・目標・評価という仕組みが暴走し始めているのではないか、そうした疑念を人文学研究者として感じるところである。

史料編纂所データベース選択画面
歴史絵引DB
肖像情報DB
肖像画模本DB
データベース選択　ホーム | 横断検索 | ログアウト
史料の所在
所蔵史料目録データベース
▼所内 Hi-CAT Plus
日本古文書ユニオンカタログ
正倉院文書マルチ支援データベース
Catalogue Database of the Batavia's Uitgaand Briefboek, 1621-1792
▼所内 マイクロフィルム簡易検索
できごとを主題に
大日本史料総合データベース
編年史料カードデータベース
近世編年データベース
維新史料綱要データベース
近世史編纂支援データベース
▼所内 編年史料集データベース
（摺物データベース）
（錦絵データベース）
図像をさぐる
歴史絵引データベース
肖像情報データベース
史料編纂所所蔵肖像画模本データベース
史料編纂所所蔵荘園絵図模本データベース
金石文拓本史料データベース
▼所内 古地図・絵図所在情報アンケートデータベース
摺物データベース
錦絵データベース
古写真データベース
全文の検索
古記録フルテキストデータベース
古文書フルテキストデータベース
奈良時代古文書フルテキストデータベース
平安遺文フルテキストデータベース
鎌倉遺文フルテキストデータベース
▼所内 未刊古文書欧文フルテキストデータベース
▼所内 明治前日本科学史フルテキストデータベース
人を主題に
（大日本史料総合データベース）
中世記録人名索引データベース
花押カードデータベース
▼所内 花押彙纂データベース
忘形見データベース
大日本史料７編人名カードデータベース
（肖像情報データベース）
（史料編纂所所蔵肖像画模本データベース）
辞典・字典
応答型翻訳支援システム（On-line Glossary of Japanese Historical Terms）
欧文日本古代史料解読辞典データベース（Dictionary of Sources of Classical Japan）（Dictionaire des sources du Japon）
電子くずし字字典データベース
『電子くずし字字典データベース』『木簡画像データベース・木簡字典』連携検索
▼所内 個人史データベース
▼所内 人名典拠サービスモジュール
▼所内 古事類苑－総目録・索引
▼所内 和暦西暦変換テーブル閲覧
© 2011 東京大学史料編纂所　史料編纂所データベースについて | 利用規程 | データベース更新スケジュール | よくあるご質問

1


肖像画模本DB
項目検索画面
史料編纂所所蔵肖像画模本データベース　DB選択 | 項目検索 | 簡易検索 | ヘルプ
キーワード
画像情報
標題
像主名
模写時原本所蔵者
架番号
画中書込
性別
身分　皇族　貴族　武士　僧侶　その他
生没年　～９Ｃ　１０Ｃ　１１Ｃ　１２Ｃ　１３Ｃ　１４Ｃ　１５Ｃ　１６Ｃ　１７Ｃ　１８Ｃ　１９Ｃ～
and or
一覧表示形式　リスト　サムネイル
一覧表示件数　20件　50件　100件
検　索　クリア　検索履歴
© 2011 東京大学史料編纂所

2


肖像画模本DB
（三条西）実隆を検索
史料編纂所所蔵肖像画模本データベース　DB選択 | 項目検索 | 簡易検索 | ヘルプ
キーワード　実隆
画像情報
標題
像主名
模写時原本所蔵者
架番号
画中書込
性別
身分　皇族　貴族　武士　僧侶　その他
生没年　～９Ｃ　１０Ｃ　１１Ｃ　１２Ｃ　１３Ｃ　１４Ｃ　１５Ｃ　１６Ｃ　１７Ｃ　１８Ｃ　１９Ｃ～
and or
一覧表示形式　リスト　サムネイル
一覧表示件数　20件　50件　100件
検　索　クリア　検索履歴
© 2011 東京大学史料編纂所

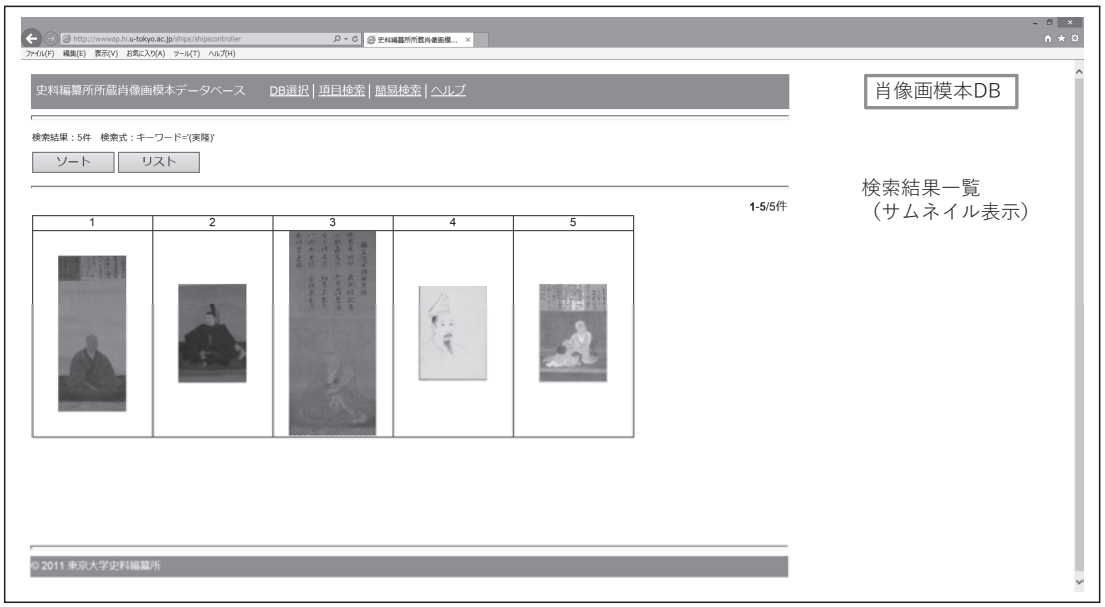
3

No	詳細	架番号	標題	模写時原本所蔵者	像主名
1	詳細	呂-43	三条西実隆画像（法体）　賛	京都府二尊院	三条西実隆，三条西公延，三条西公世，堯空，耕隠，逍遥院
2	詳細	呂-44	三条西実条画像	京都府二尊院	三条西実条，香雲院
3	詳細	波-47	三条西公条画像（法体）　自賛	京都府二尊院	三条西公条，仍覚，称名院
4	詳細	波-309	三条西実隆画像［線描］	酒井宇吉（→東京大学史料編纂所）	三条西実隆
5	詳細	波-310	三条西実隆像［現状模写］	三条西実謙(横浜市都筑区)	三条西実隆

肖像画模本DB

検索結果一覧
（テキスト表示）

4



肖像画模本DB

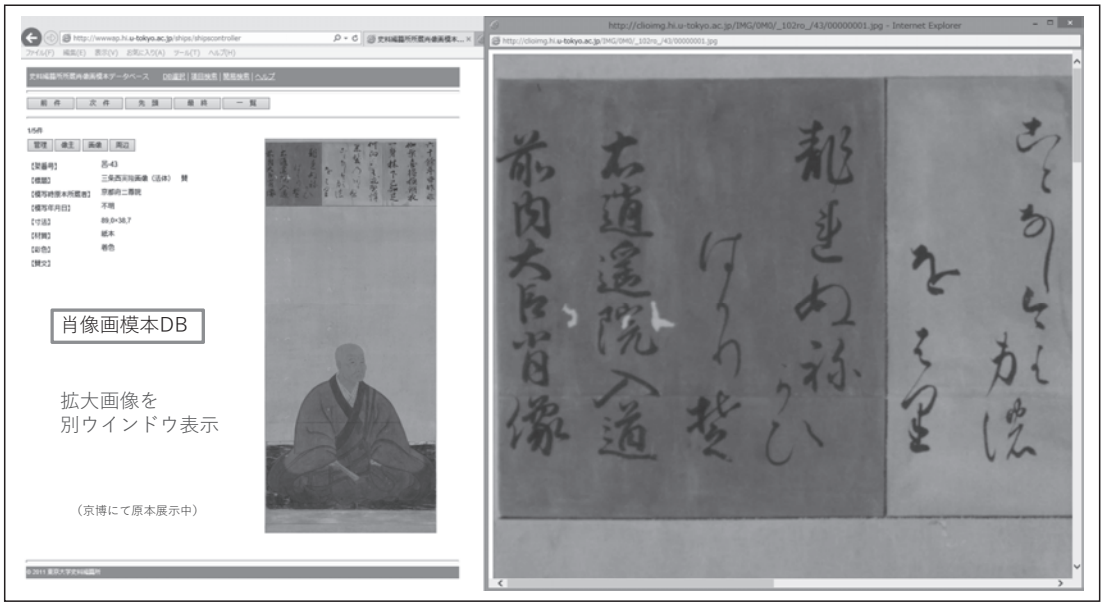
検索結果一覧
（サムネイル表示）

5

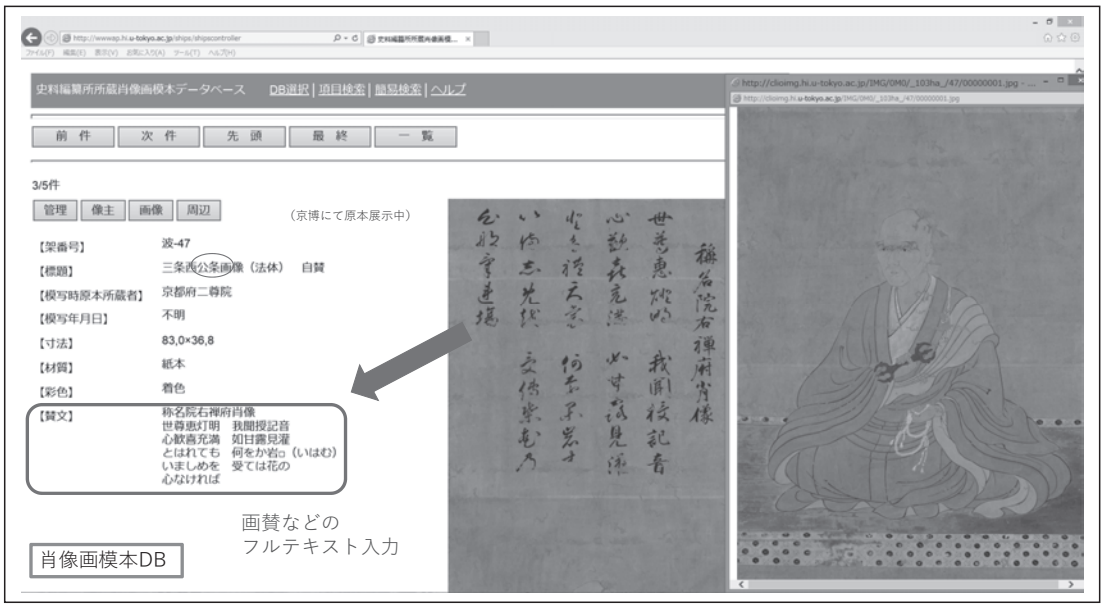
肖像画模本DB

検索結果
詳細個別表示
（管理情報）

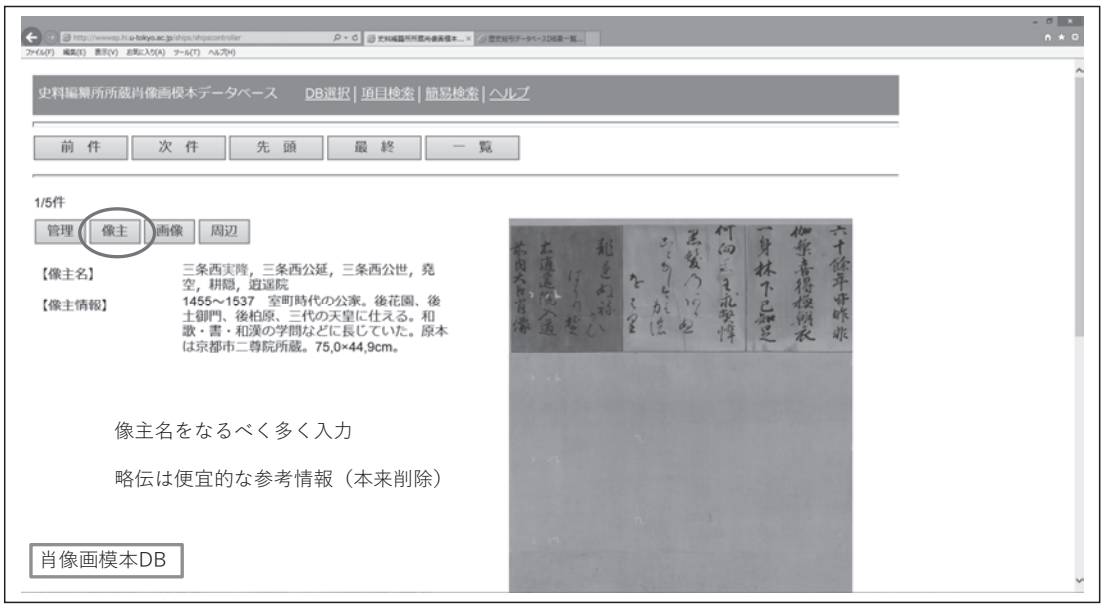
6


肖像画模本DB
拡大画像を
別ウインドウ表示
（京博にて原本展示中）

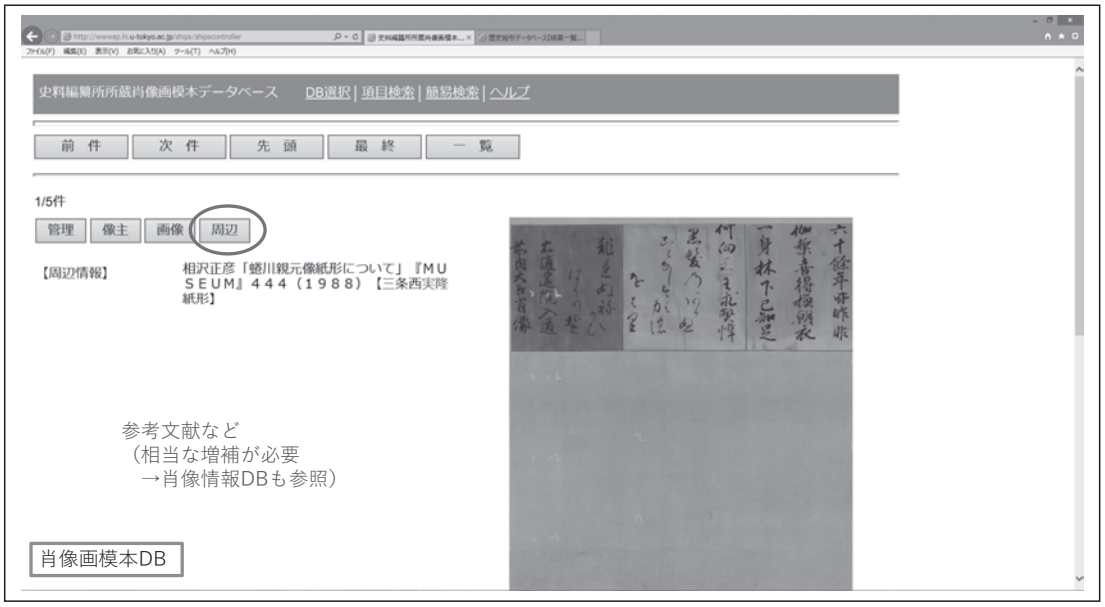
7


史料編纂所所蔵肖像画模本データベース
DB選択 | 項目検索 | 簡易検索 | ヘルプ
前 件
次 件
先 頭
最 終
一 覧
3/5件
管理
像主
画像
周辺
（京博にて原本展示中）
【架番号】 波-47
【標題】 三条西公条画像（法体） 自賛
【模写時原本所蔵者】 京都府二尊院
【模写年月日】 不明
【寸法】 83,0×36,8
【材質】 紙本
【彩色】 着色
【賛文】 称名院右禅府肖像
世尊恵灯明 我聞授記音
心歓喜充満 如甘露見灌
とはれても 何をか岩ぢ（いはむ）
いましめを 受ては花の
心なければ
画賛などの
フルテキスト入力
肖像画模本DB

8


史料編纂所所蔵肖像画模本データベース
DB選択 | 項目検索 | 簡易検索 | ヘルプ
前 件
次 件
先 頭
最 終
一 覧
1/5件
管理
像主
画像
周辺
【像主名】 三条西実隆，三条西公延，三条西公世，堯空，耕隠，逍遥院
【像主情報】 1455～1537 室町時代の公家。後花園、後土御門、後柏原、三代の天皇に仕える。和歌・書・和漢の学問などに長じていた。原本は京都市二尊院所蔵。75,0×44,9cm。
像主名をなるべく多く入力
略伝は便宜的な参考情報（本来削除）
肖像画模本DB

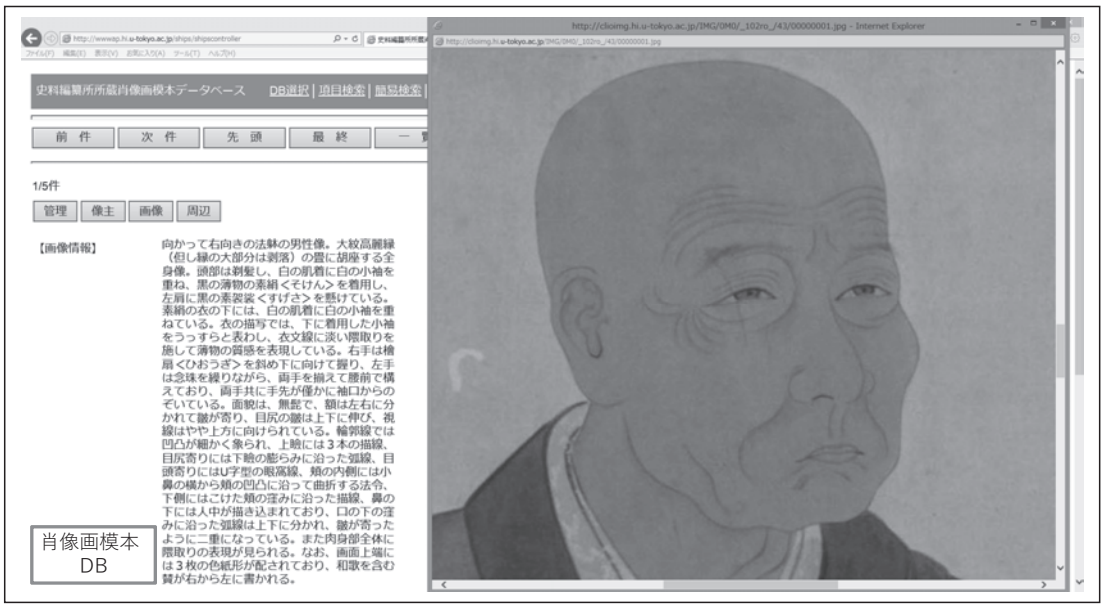
9


史料編纂所所蔵肖像画模本データベース
DB選択 | 項目検索 | 簡易検索 | ヘルプ
前 件
次 件
先 頭
最 終
一 覧
1/5件
管理
像主
画像
周辺
【周辺情報】
相沢正彦「蜷川親元像紙形について」『MUSEUM』444（1988）【三条西実隆紙形】
参考文献など
（相当な増補が必要
→肖像情報DBも参照）
肖像画模本DB

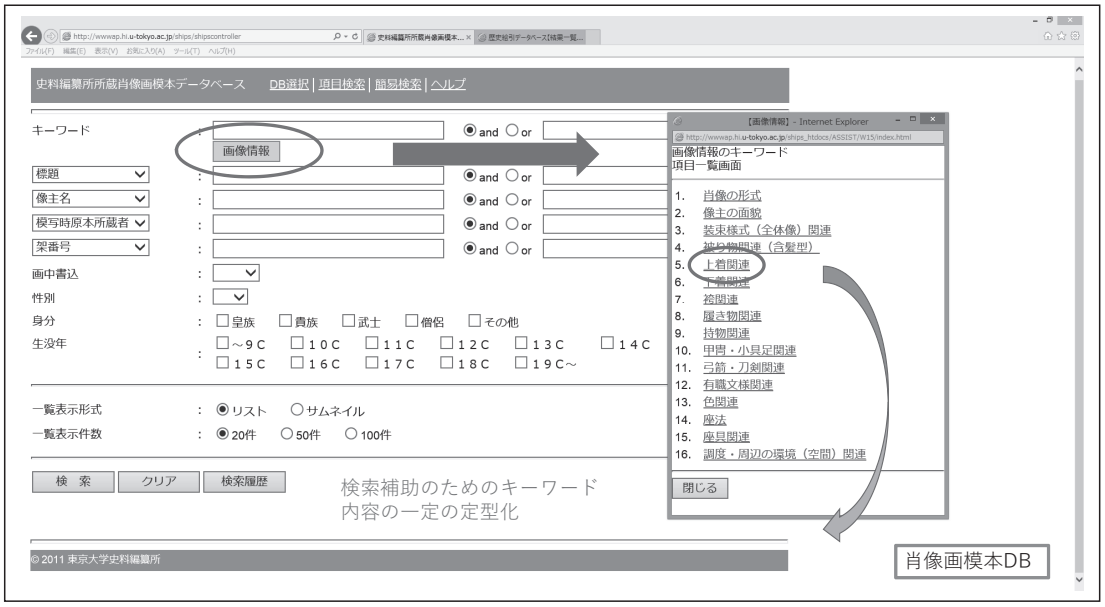
10

史料編纂所所蔵肖像画模本データベース
DB選択 | 項目検索 | 簡易検索 | ヘルプ
前 件
次 件
先 頭
最 終
一 覧
1/5件
管理
像主
画像
周辺
【画像情報】
向かって右向きの法躰の男性像。大紋高麗縁（但し縁の大部分は剥落）の畳に胡座する全身像。頭部は剃髪し、白の肌着に白の小袖を重ね、黒の薄物の素絹<そけん>を着用し、左肩に黒の素袈裟<すげさ>を懸けている。素絹の衣の下には、白の肌着に白の小袖を重ねている。衣の描写では、下に着用した小袖をうっすらと表わし、衣文線に淡い隈取りを施して薄物の質感を表現している。右手は檜扇<ひおうぎ>を斜め下に向けて握り、左手は念珠を繰りながら、両手を揃えて腰前で構えており、両手共に手先が僅かに袖口からのぞいている。面貌は、無髭で、額は左右に分かれて皺が寄り、目尻の皺は上下に伸び、視線はやや上方に向けられている。輪郭線では凹凸が細かく象られ、上瞼には３本の描線、目尻寄りには下瞼の膨らみに沿った弧線、目頭寄りにはU字型の眼窩線、頬の内側には小鼻の横から頬の凹凸に沿って曲折する法令、下側にはこけた頬の窪みに沿った描線、鼻の下には人中が描き込まれており、口の下の窪みに沿った弧線は上下に分かれ、皺が寄ったように二重になっている。また肉身部全体に隈取りの表現が見られる。なお、画面上端には３枚の色紙形が配されており、和歌を含む賛が右から左に書かれる。
画面に描かれた内容の詳細な記述（ディスクリプション）
服飾と面貌表現、環境描写
肖像画模本DB

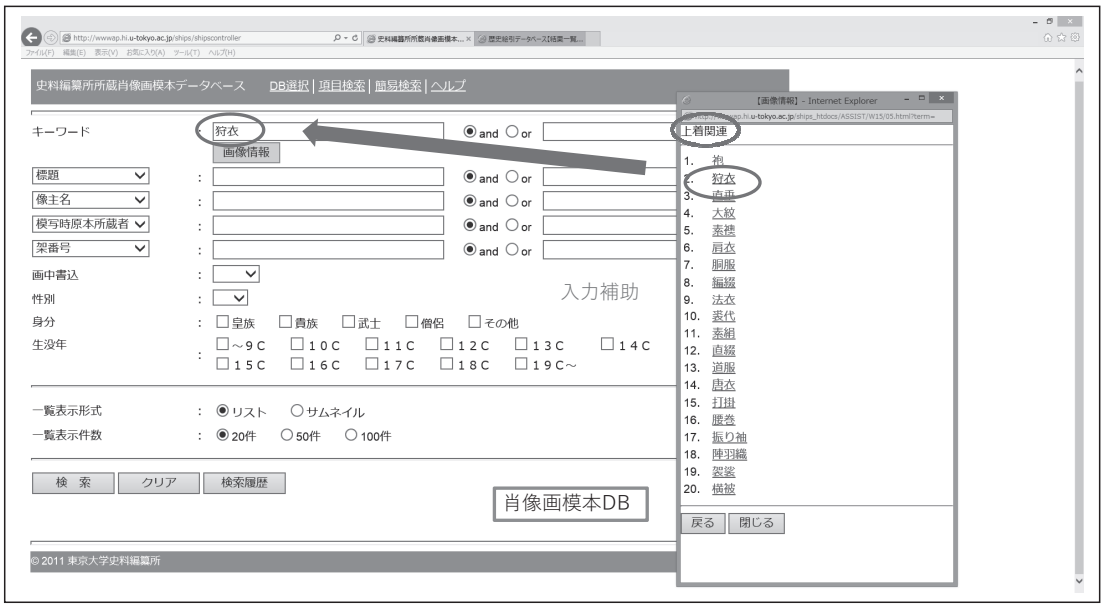
11


史料編纂所所蔵肖像画模本データベース
DB選択 | 項目検索 | 簡易検索
前 件
次 件
先 頭
最 終
1/5件
管理
像主
画像
周辺
【画像情報】
向かって右向きの法躰の男性像。大紋高麗縁（但し縁の大部分は剥落）の畳に胡座する全身像。頭部は剃髪し、白の肌着に白の小袖を重ね、黒の薄物の素絹<そけん>を着用し、左肩に黒の素袈裟<すげさ>を懸けている。素絹の衣の下には、白の肌着に白の小袖を重ねている。衣の描写では、下に着用した小袖をうっすらと表わし、衣文線に淡い隈取りを施して薄物の質感を表現している。右手は檜扇<ひおうぎ>を斜め下に向けて握り、左手は念珠を繰りながら、両手を揃えて腰前で構えており、両手共に手先が僅かに袖口からのぞいている。面貌は、無髭で、額は左右に分かれて皺が寄り、目尻の皺は上下に伸び、視線はやや上方に向けられている。輪郭線では凹凸が細かく象られ、上瞼には３本の描線、目尻寄りには下瞼の膨らみに沿った弧線、目頭寄りにはU字型の眼窩線、頬の内側には小鼻の横から頬の凹凸に沿って曲折する法令、下側にはこけた頬の窪みに沿った描線、鼻の下には人中が描き込まれており、口の下の窪みに沿った弧線は上下に分かれ、皺が寄ったように二重になっている。また肉身部全体に隈取りの表現が見られる。なお、画面上端には３枚の色紙形が配されており、和歌を含む賛が右から左に書かれる。
肖像画模本DB

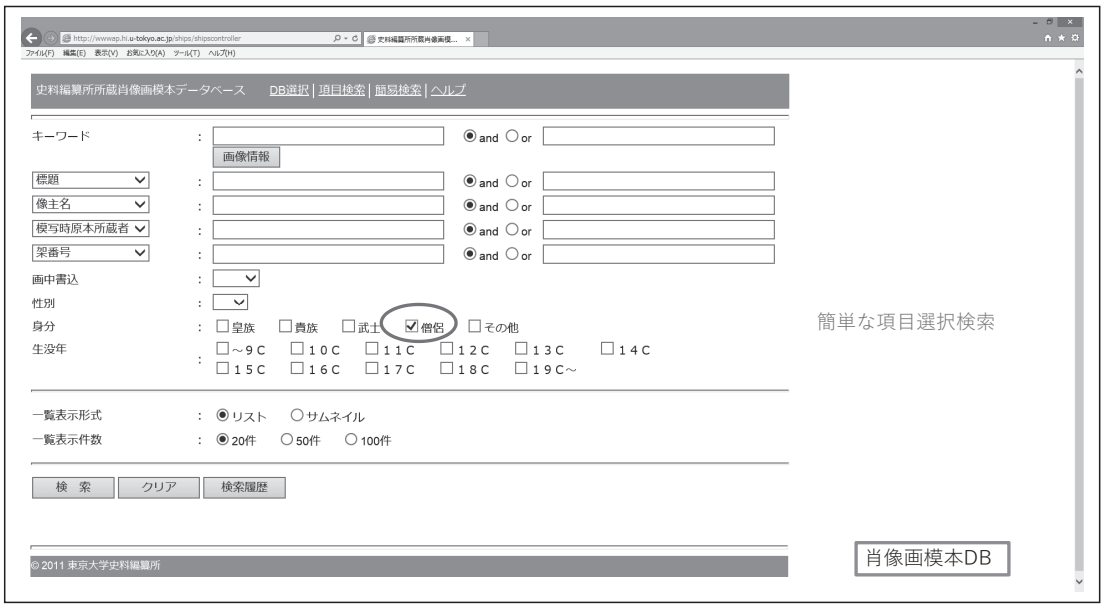
12


史料編纂所所蔵肖像画模本データベース
DB選択 | 項目検索 | 簡易検索 | ヘルプ
キーワード
画像情報
標題
像主名
模写時原本所蔵者
架番号
画中書込
性別
身分
生没年
一覧表示形式
一覧表示件数
検 索
クリア
検索履歴
検索補助のためのキーワード
内容の一定の定型化
画像情報のキーワード
項目一覧画面
1. 肖像の形式
2. 像主の面貌
3. 装束様式（全体像）関連
4. 被り物関連（含髪型）
5. 上着関連
6. 下着関連
7. 袴関連
8. 履き物関連
9. 持物関連
10. 甲冑・小具足関連
11. 弓箭・刀剣関連
12. 有職文様関連
13. 色関連
14. 座法
15. 座具関連
16. 調度・周辺の環境（空間）関連
閉じる
肖像画模本DB

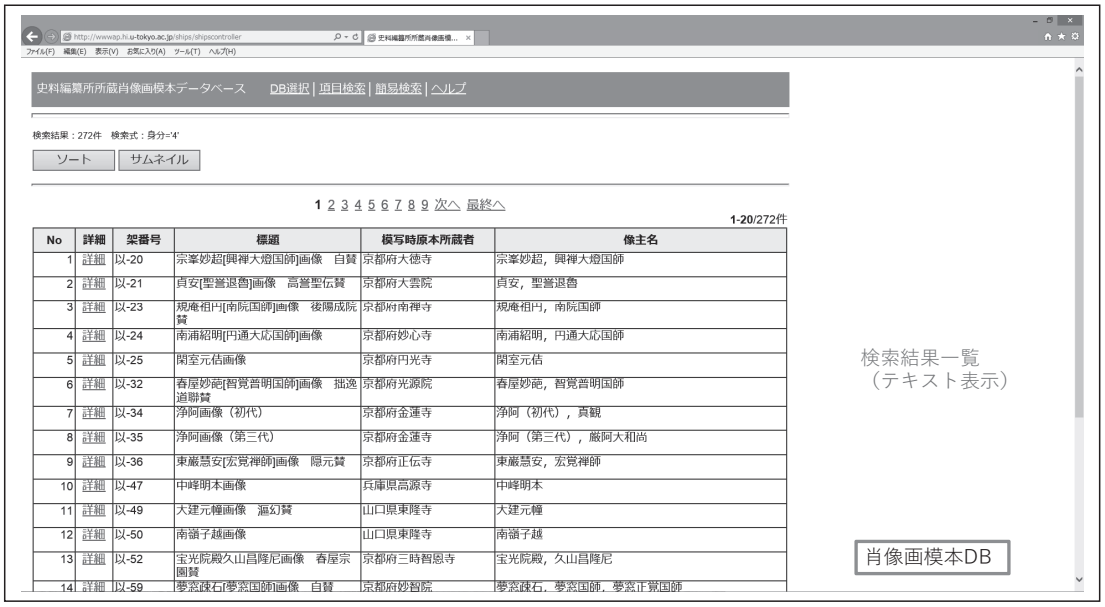
13


史料編纂所所蔵肖像画模本データベース
DB選択 | 項目検索 | 簡易検索 | ヘルプ
キーワード
狩衣
画像情報
入力補助
上着関連
1. 袍
2. 狩衣
3. 直垂
4. 大紋
5. 素襖
6. 肩衣
7. 胴服
8. 編綴
9. 法衣
10. 裘代
11. 素絹
12. 直綴
13. 道服
14. 唐衣
15. 打掛
16. 腰巻
17. 振り袖
18. 陣羽織
19. 袈裟
20. 横被
戻る
閉じる
肖像画模本DB

14


史料編纂所所蔵肖像画模本データベース
DB選択 | 項目検索 | 簡易検索 | ヘルプ
簡単な項目選択検索
肖像画模本DB

15

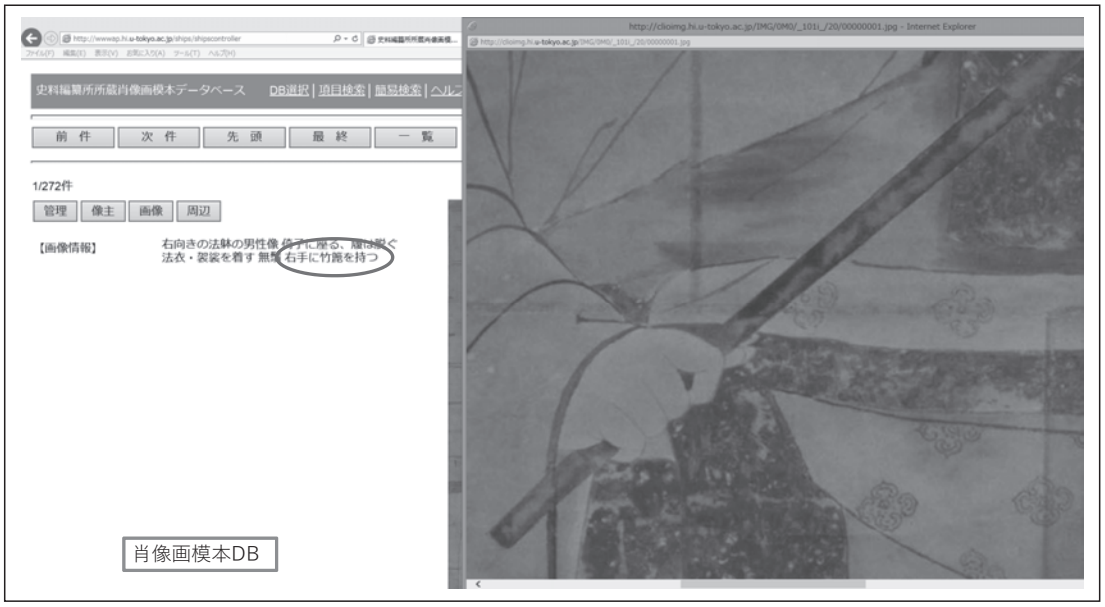


No	詳細	架番号	標題	模写時原本所蔵者	像主名
1	詳細	以-20	宗峯妙超[興禅大燈国師]画像　自賛	京都府大徳寺	宗峯妙超，興禅大燈国師
2	詳細	以-21	貞安[聖誉退耆]画像　高誉聖伝賛	京都府大雲院	貞安，聖誉退耆
3	詳細	以-23	規庵祖円[南院国師]画像　後陽成院賛	京都府南禅寺	規庵祖円，南院国師
4	詳細	以-24	南浦紹明[円通大応国師]画像	京都府妙心寺	南浦紹明，円通大応国師
5	詳細	以-25	関室元佶画像	京都府円光寺	関室元佶
6	詳細	以-32	春屋妙葩[智覚普明国師]画像　拙逸道聯賛	京都府光源院	春屋妙葩，智覚普明国師
7	詳細	以-34	浄阿画像（初代）	京都府金蓮寺	浄阿（初代），真観
8	詳細	以-35	浄阿画像（第三代）	京都府金蓮寺	浄阿（第三代），厳阿大和尚
9	詳細	以-36	東厳慧安[宏覚禅師]画像　隠元賛	京都府正伝寺	東厳慧安，宏覚禅師
10	詳細	以-47	中峰明本画像	兵庫県高源寺	中峰明本
11	詳細	以-49	大建元幢画像　漏幻賛	山口県東隆寺	大建元幢
12	詳細	以-50	南嶺子越画像	山口県東隆寺	南嶺子越
13	詳細	以-52	宝光院殿久山昌隆尼画像　春屋宗園賛	京都府三時智恩寺	宝光院殿，久山昌隆尼
14	詳細	以-59	夢窓疎石[夢窓国師]画像　自賛	京都府妙智院	夢窓疎石，夢窓国師，夢窓正覚国師

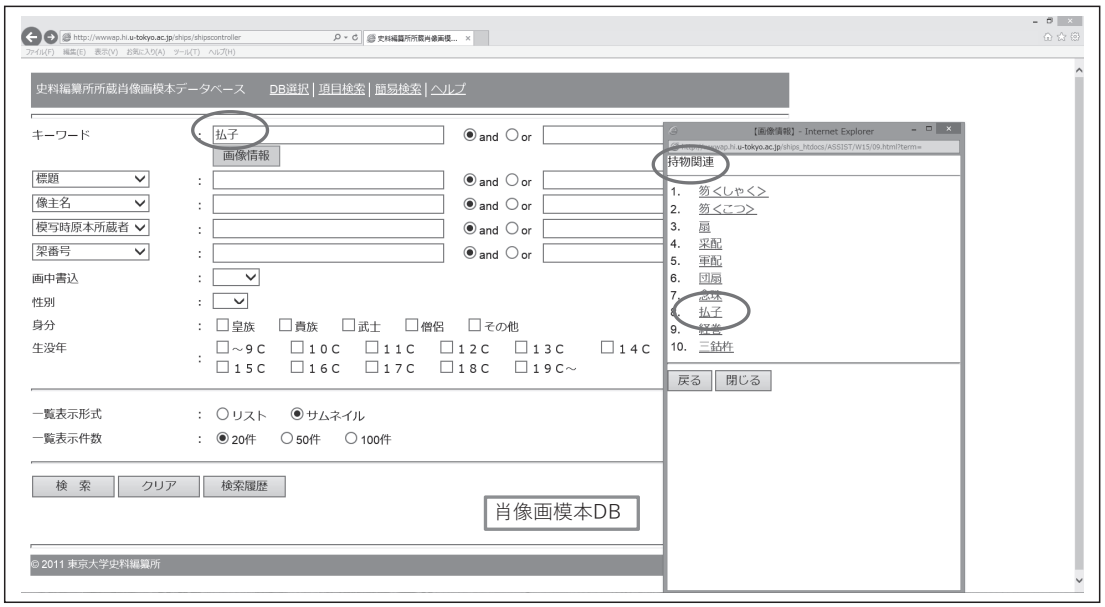
16

17

18


史料編纂所所蔵肖像画模本データベース
DB選択 | 項目検索 | 簡易検索 | ヘルプ
前 件
次 件
先 頭
最 終
一 覧
1/272件
管理
像主
画像
周辺
【画像情報】
右向きの法躰の男性像 椅子に座る、履は履く
法衣・袈裟を着す 無冠 右手に竹箆を持つ
肖像画模本DB

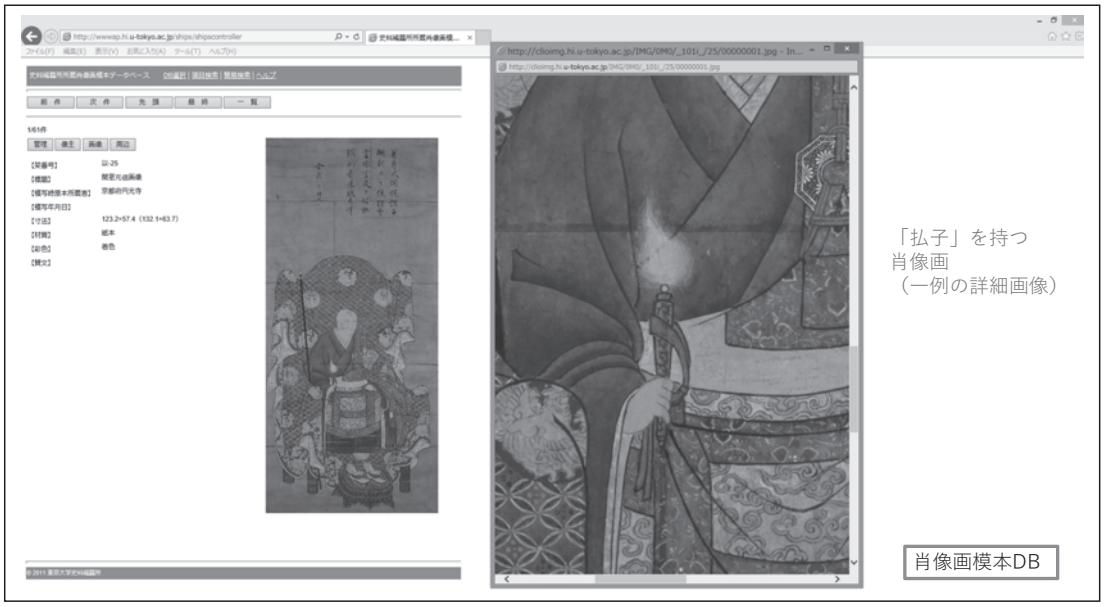
19


史料編纂所所蔵肖像画模本データベース
DB選択 | 項目検索 | 簡易検索 | ヘルプ
キーワード
払子
画像情報
and or
標題
像主名
模写時原本所蔵者
架番号
画中書込
性別
身分
皇族
貴族
武士
僧侶
その他
生没年
～9C
10C
11C
12C
13C
14C
15C
16C
17C
18C
19C～
一覧表示形式
リスト
サムネイル
一覧表示件数
20件
50件
100件
検 索
クリア
検索履歴
肖像画模本DB
© 2011 東京大学史料編纂所
[画像情報] - Internet Explorer
持物関連
1. 笏<しゃく>
2. 笏<こつ>
3. 扇
4. 采配
5. 軍配
6. 団扇
7. 念珠
8. 払子
9. 経巻
10. 三鈷杵
戻る
閉じる

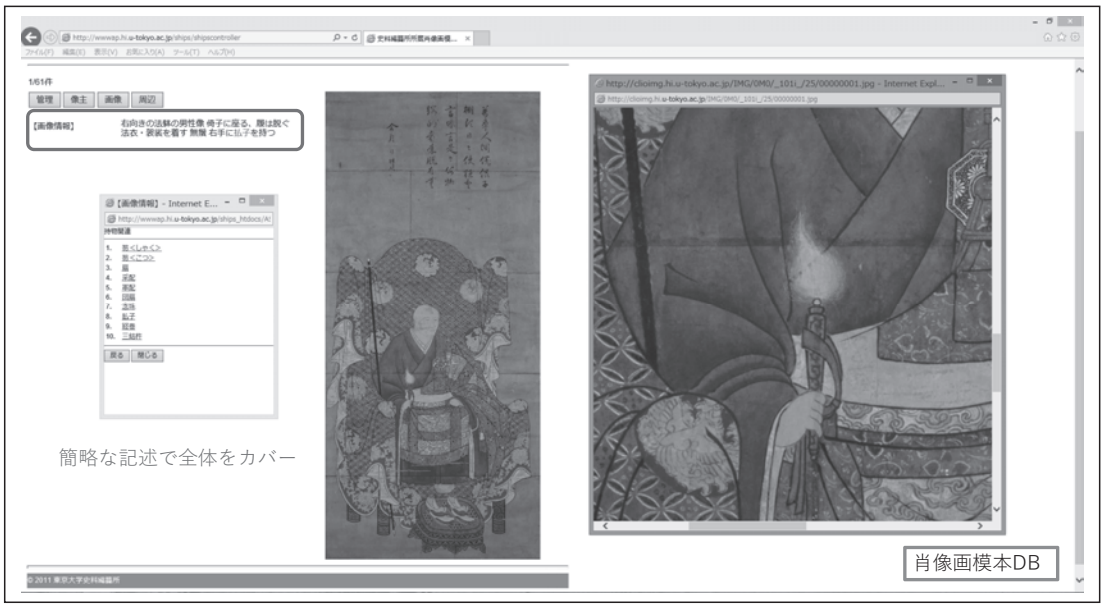
20

「払子」を持つ肖像画
（サムネイル表示）
描かれたモノを集める検索
肖像画模本DB

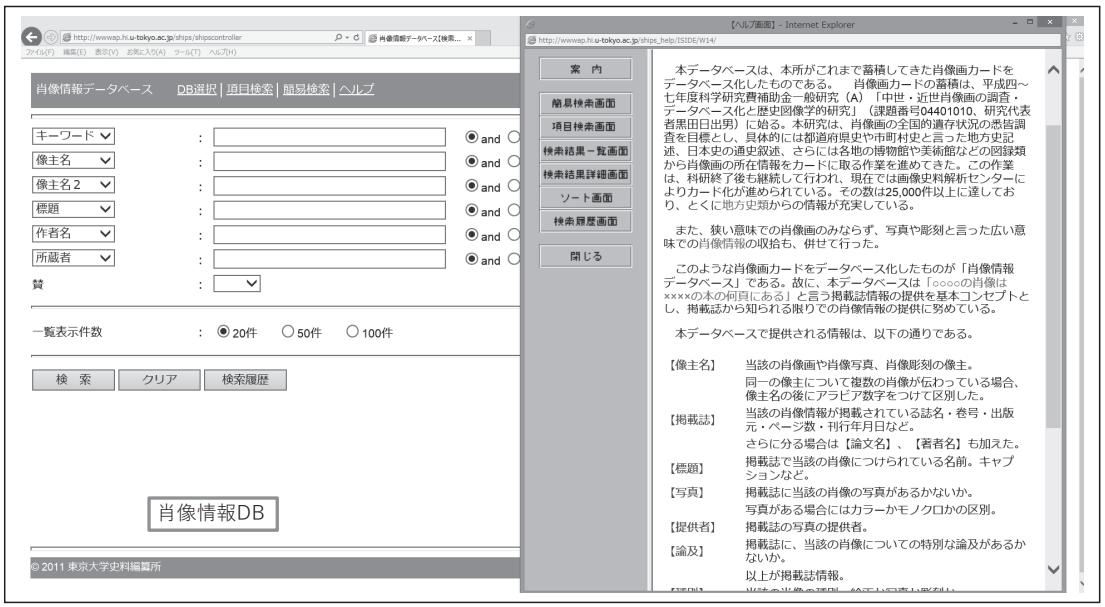
21


「払子」を持つ
肖像画
（一例の詳細画像）
肖像画模本DB

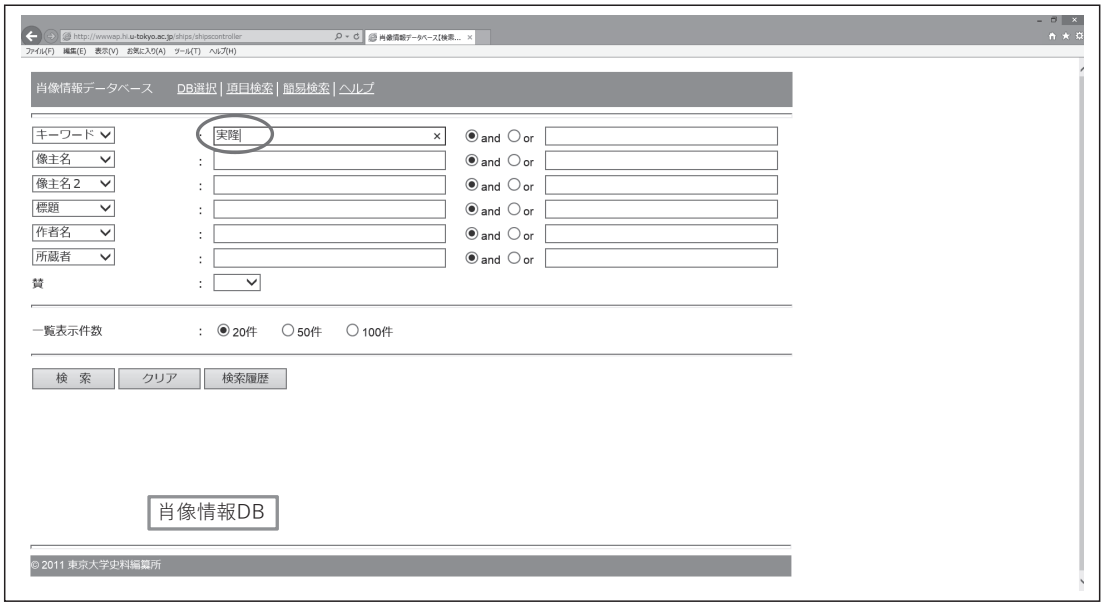
22


簡略な記述で全体をカバー
肖像画模本DB

23


肖像情報データベース　DB選択 | 項目検索 | 簡易検索 | ヘルプ
キーワード
像主名
像主名２
標題
作者名
所蔵者
賛
一覧表示件数　20件　50件　100件
検　索　クリア　検索履歴
肖像情報DB
© 2011 東京大学史料編纂所
【ヘルプ画面】- Internet Explorer
案　内
簡易検索画面
項目検索画面
検索結果一覧画面
検索結果詳細画面
ソート画面
検索履歴画面
閉じる
本データベースは、本所がこれまで蓄積してきた肖像画カードをデータベース化したものである。　肖像画カードの蓄積は、平成四～七年度科学研究費補助金一般研究（A）「中世・近世肖像画の調査・データベース化と歴史図像学的研究」（課題番号04401010、研究代表者黒田日出男）に始る。本研究は、肖像画の全国的遺存状況の悉皆調査を目標とし、具体的には都道府県史や市町村史と言った地方史記述、日本史の通史叙述、さらには各地の博物館や美術館などの図録類から肖像画の所在情報をカードに取る作業を進めてきた。この作業は、科研終了後も継続して行われ、現在では画像史料解析センターによりカード化が進められている。その数は25,000件以上に達しており、とくに地方史類からの情報が充実している。
また、狭い意味での肖像画のみならず、写真や彫刻と言った広い意味での肖像情報の収拾も、併せて行った。
このような肖像画カードをデータベース化したものが「肖像情報データベース」である。故に、本データベースは「○○○○の肖像は××××の本の何頁にある」と言う掲載誌情報の提供を基本コンセプトとし、掲載誌から知られる限りでの肖像情報の提供に努めている。
本データベースで提供される情報は、以下の通りである。
【像主名】　当該の肖像画や肖像写真、肖像彫刻の像主。
同一の像主について複数の肖像が伝わっている場合、像主名の後にアラビア数字をつけて区別した。
【掲載誌】　当該の肖像情報が掲載されている誌名・巻号・出版元・ページ数・刊行年月日など。
さらに分る場合は【論文名】、【著者名】も加えた。
【標題】　掲載誌で当該の肖像につけられている名前。キャプションなど。
【写真】　掲載誌に当該の肖像の写真があるかないか。
写真がある場合にはカラーかモノクロかの区別。
【提供者】　掲載誌の写真の提供者。
【論及】　掲載誌に、当該の肖像についての特別な論及があるかないか。
以上が掲載誌情報。

24


肖像情報データベース　DB選択 | 項目検索 | 簡易検索 | ヘルプ
キーワード
実隆
像主名
像主名2
標題
作者名
所蔵者
賛
and or
一覧表示件数
20件
50件
100件
検　索
クリア
検索履歴
肖像情報DB
© 2011 東京大学史料編纂所

25

肖像情報データベース　DB選択 | 項目検索 | 簡易検索 | ヘルプ
検索結果：12件　検索式：キーワード='(実隆)'
ソート
1-12/12件
No | 詳細 | 像主名 | 掲載誌 | 出版年月日 | 標題
1 | 詳細 | 三条西実隆（なし） | 川西町史・通史編・上巻 | 1987 | 三条西実隆像
2 | 詳細 | 三条西実隆（二尊院） | 京都府葛野郡史概要 | | 絹本着色遊遙院實隆像
3 | 詳細 | 三条西実隆（二尊院） | 史料京都見聞記・1巻 | | 西三条遊遙院尭空画像
4 | 詳細 | 三条西実隆（） | 新宮町史・5巻 | 1962/7/25 | 三条西実隆肖像
5 | 詳細 | 三条西実隆（酒井宇吉） | 伊丹市史・1巻 | 1971/3/31 | 三条西実隆
6 | 詳細 | 三条西実隆（二尊院（天台宗）） | 全国寺院名鑑 | 1983/5/1 | 絹本着色・遊遙院実隆像
7 | 詳細 | 三条西実隆（酒井家） | 宮廷画壇史の研究 | 1996/2/28 | 三条西実隆像・下絵
8 | 詳細 | 三条西実隆（二尊院） | 室町時代の狩野派 | 1996 | 三条西実隆像
9 | 詳細 | 三条西実隆（京都市・二尊院） | 丹南町史・上巻 | 1994/3/31 | 三条西実隆画像
10 | 詳細 | 三条西実隆（二尊院） | 記録の文化　－日記の世界－ | | 三条西実隆画像
11 | 詳細 | 三条西実隆（二尊院） | 日本の歴史⑨・日本国王と土民 | | 三条西実隆画像
12 | 詳細 | 三条西実隆 | 特別展　室町時代の肖像画 | 2000年9月 | 三条西実隆像
© 2011 東京大学史料編纂所
検索結果一覧
校正と増補が必要
肖像情報DB

26

8/12件
【像主名】　三条西実隆（二尊院）
【著者名】　なし
【論文名】　作品解説
【掲載誌】　室町時代の狩野派(京都国立博物館) 220頁 1996年
【請求番号】
【標題】　三条西実隆像
【写真】　有り(モノクロ、全体)
【提供者】
【論及】　有り
【種別】　絵画
【所蔵者】　二尊院
【作者】
【製作年代】
【法量】
【着彩】　なし
【材質】
【賛】　有り
© 2011 東京大学史料編纂所
検索結果　詳細表示
元はカードで採取
→直接PCで蓄積
肖像情報DB

27

画像検索：図解トレース図

と

キャプション検索
：工具書の横断索引

との二本立て

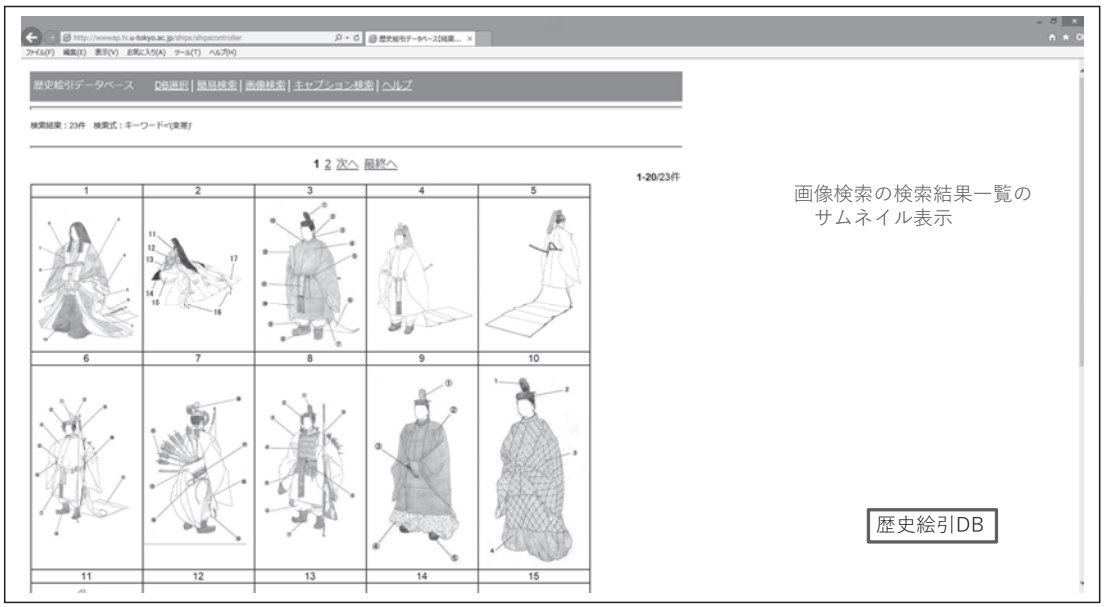
歷史絵引DB

28

画像検索と
キャプション検索の
横断検索結果一覧

歷史絵引DB

29



画像検索の検索結果一覧の
サムネイル表示

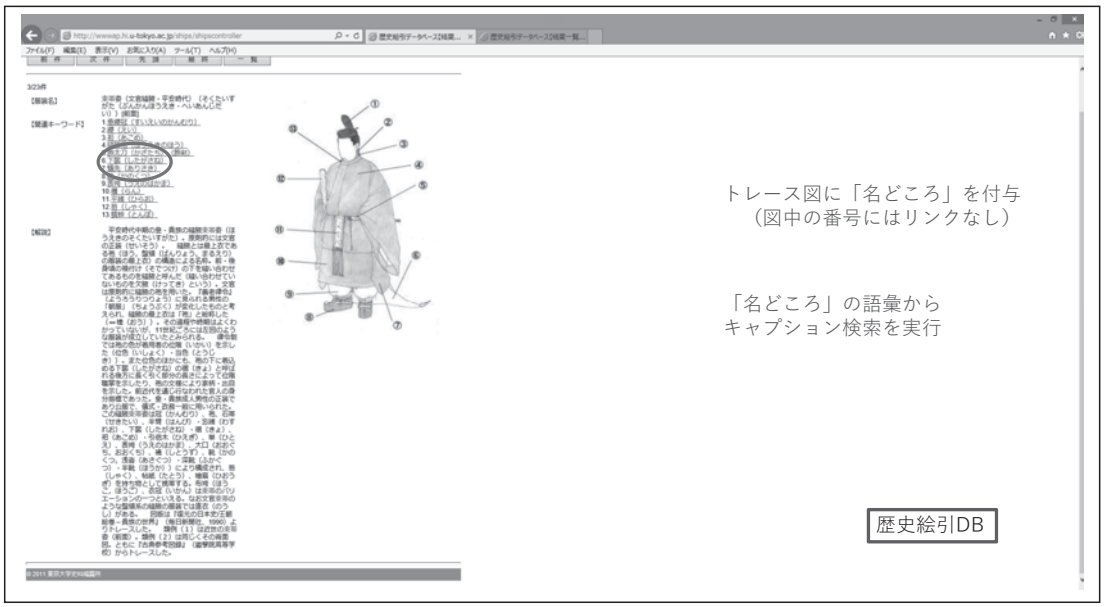
歷史絵引DB

30

画像検索の検索結果
詳細内容

歷史絵引DB

31



トレース図に「名どころ」を付与
（図中の番号にはリンクなし）

「名どころ」の語彙から
キャプション検索を実行

歷史絵引DB

32

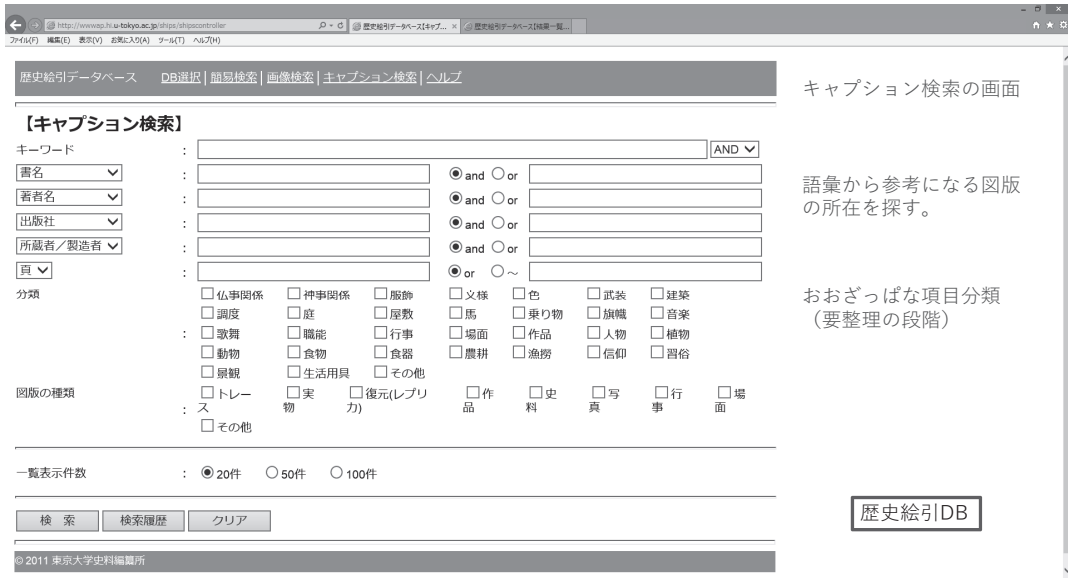
No	詳細	キャプション	図中名所	書名	頁
1	詳細	209　御下襲　夏の料		『服飾史図絵』上	193
2	詳細	228　下襲　白浮線綾丸文綾　冬の料		『服飾史図絵』上	197
3	詳細	229　下襲の別裾　浮線綾丸文綾		『服飾史図絵』上	197
4	詳細	230　下襲　濃綠の殺襪　夏の料		『服飾史図絵』上	197
5	詳細	270　衣冠姿　石帯,半臂,下襲をのぞき,指貫,靴をはく（春日権現験記絵）		『服飾史図絵』上	203
6	詳細	385　褐衣と下襲の裾とを重ねる（加茂祭礼図巻）		『服飾史図絵』上	219
7	詳細	646　舞楽装束　平装束　袍、半臂、下襲、裾、袴（東京国立博物館）		『服飾史図絵』上	252
8	詳細	655　綾切　平装束　右肩を脱ぎ半臂、下襲をあらわしている（年中行事絵巻）		『服飾史図絵』上	253
9	詳細	686　装束の紋様　小榮<白小榮は天皇御下襲>綾		『服飾史図絵』上	258
10	詳細	687　装束の紋様　臥蝶の丸<親王、摂関の下襲>綾		『服飾史図絵』上	258
11	詳細	795　宮廷着付（男装）　下襲を着せる		『服飾史図絵』上	269
12	詳細	装束　束帯　正面（234図参照）	緌，袍，袍の袖，奥袖，鰭袖，襴，蟻先，半臂，下襲の袖，下襲の裾，表袴，表袴のおめり，平緒，浅沓	『服飾史図絵解説』	97
13	詳細	装束　束帯　後姿（235図参照）	冠，緌，袍，襴，蟻先，下襲の裾，表袴，笏，石帯，平緒，飾剡，浅沓	『服飾史図絵解説』	97
14	詳細	装束一覧　束帯　下襲		『服飾史図絵解説』	103
15	詳細	続きの下襲		『有識故実大辞典』	347

「名どころ」の語彙から
キャプション検索を実行

検索結果の一覧

歷史絵引DB

33



キャプション検索の画面

語彙から参考になる図版の所在を探す。

おおざっぱな項目分類（要整理の段階）

歷史絵引DB

34

キャプション検索
検索結果一覧

歷史絵引DB

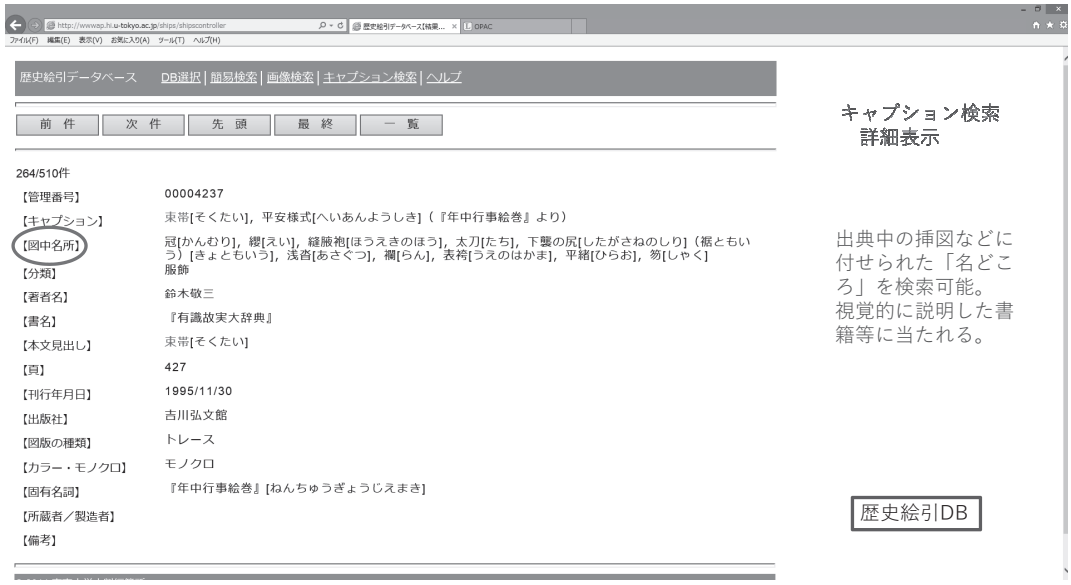
35

キャプション検索
詳細表示

［よみがな］は検索便宜のために入力してあるが、必ずしも正確ではなく、典拠として参照される可能性があり、問題。

歷史絵引DB

36



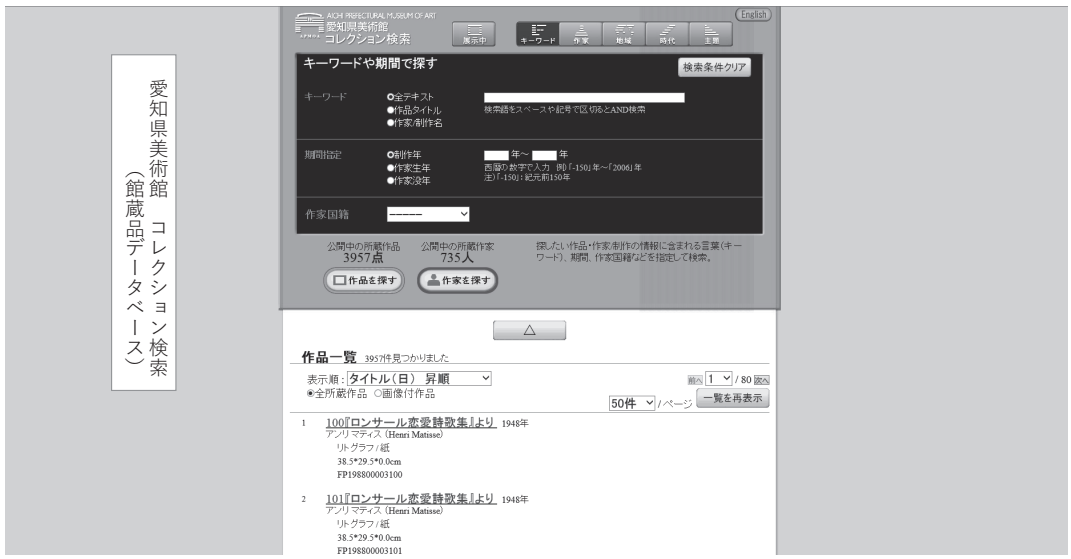
キャプション検索
詳細表示

出典中の挿図などに付せられた「名どころ」を検索可能。
視覚的に説明した書籍等に当たれる。

歴史絵引DB

37

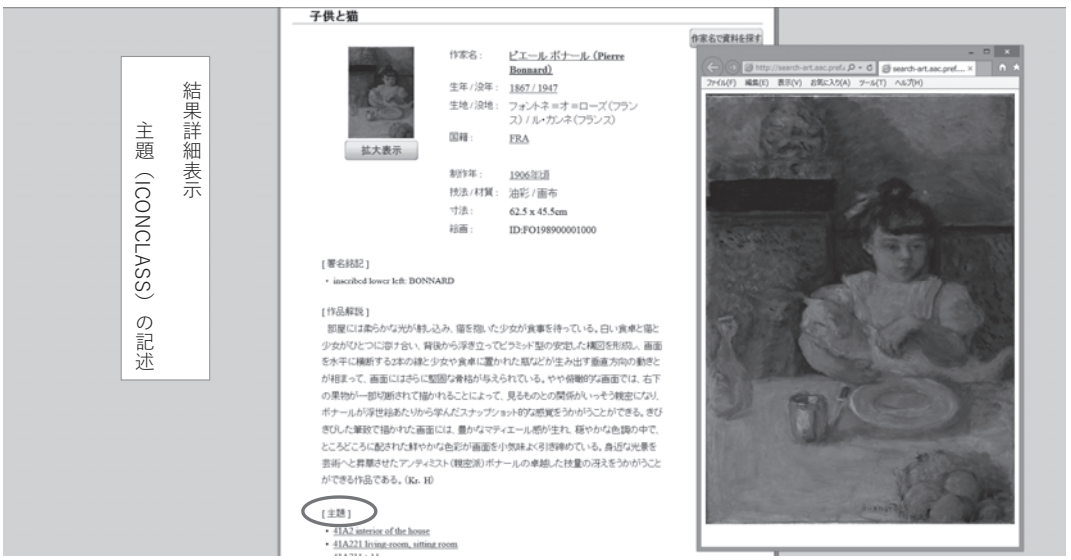
愛知県美術館　コレクション検索（館蔵品データベース）



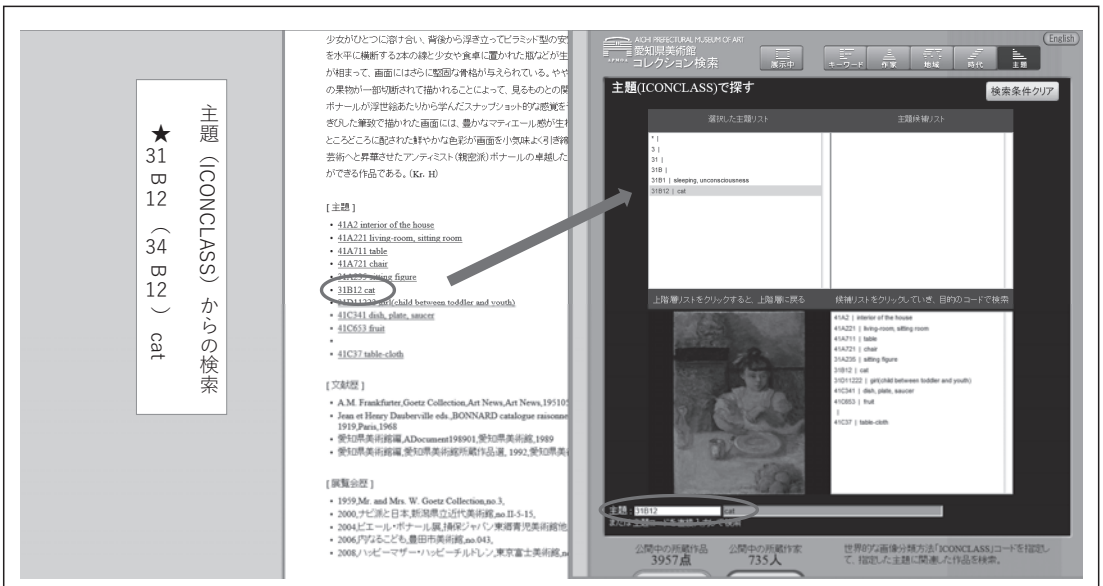
38

キーワード検索「猫」

39


結果詳細表示
主題（ICONCLASS）の記述
子供と猫
作家名：
ピエール ボナール（Pierre Bonnard）
生年／没年：
1867 / 1947
生地／没地：
フォントネ＝オ＝ローズ（フランス）／ル・カンネ（フランス）
国籍：
FRA
制作年：
1906年頃
技法／材質：
油彩／画布
寸法：
62.5 x 45.5cm
絵画：
ID:FO198900001000
拡大表示
作家名で資料を探す
[署名銘記]
・inscribed lower left: BONNARD
[作品解説]
[主題]
・41A2 interior of the house
・41A221 living-room, sitting room

40


主題（ICONCLASS）からの検索
★31B12（34B12）cat
[主題]
・41A2 interior of the house
・41A221 living-room, sitting room
・41A711 table
・41A721 chair
・31B12 cat
・41C341 dish, plate, saucer
・41C653 fruit
・41C37 table-cloth
[文献歴]
[展覧会歴]
ACHI PREFECTURAL MUSEUM OF ART
愛知県美術館
コレクション検索
English
主題(ICONCLASS)で探す
検索条件クリア
選択した主題リスト
主題候補リスト
上階層リストをクリックすると、上階層に戻る
候補リストをクリックしていき、目的のコードで検索
主題：31B12 cat
公開中の所蔵作品 3957点
公開中の所蔵作家 735人
世界的な画像分類方法「ICONCLASS」コードを指定して、指定した主題に関連した作品を検索。

41

主題（ICONCLASS）からの検索
※主題の記述が未入力もあるが、描かれた内容検索。
主題：31B12 cat
公開中の所蔵作品 3957点
公開中の所蔵作家 735人
世界的な画像分類方法「ICONCLASS」コードを指定して、指定した主題に関連した作品を検索。
作品を探す
作家を探す
作品一覧 3件見つかりました
表示順：タイトル（日） 昇順
全所蔵作品 画像付作品
10件 / ページ
一覧を再表示
仔猫 1940年
熊谷守一（くまがい もりかず）
墨画淡彩／紙本
40.6 x 56.9cm 135.4 x 67.0cm
JJ200300018000
子供と猫 1906年頃
ピエール ボナール（Pierre Bonnard）
油彩／画布
62.5 x 45.5cm
FO198900001000
猫 1964年
熊谷守一（くまがい もりかず）
墨画／紙本
143.1 x 71.1cm 42.4 x 55.6cm
JJ200300002000
関連サイト
ICONCLASS

42

國立故宮博物院
NATIONAL PALACE MUSEUM
請輸入關鍵字
搜尋
熱門 ： 故宮週末夜· 衍生性商品· 網路商城· 山水合璧· 翠玉白菜· 繪畫· 書法· 展覽
網站導覽 常見問答 網路會員中心 行動版 聯絡我們 語言選擇
一般參觀者 學校師生 研究人員 媒體記者 合作廠商
參觀故宮 典藏資源 學習資源 博物館行政 認識故宮
典藏精選
主題網站
文物3D賞析
典藏資料庫系統
2015.12.28
故宮南院
璀璨登場
The Grand Opening of the Southern Branch of the National Palace Museum
南部院區網站
最新消息
活動行事曆
友站資訊
RSS訂閱區
行政服務
最新消息
參觀及預約資訊，請連結至<南部院區網站>查詢。
最新消息
國立故宮博物院團體導覽解說服務培訓課程及認證實施計畫
友站資訊 2009/01/09 行政院公報資訊網
最新消息 2016/02/03 105年度1月份拾得遺失物處理情形追蹤表及寄物櫃退幣盒拾獲零錢登記表
最新消息 2016/02/03 「天國的寶藏－教廷文物特展」展覽說明書提供下載
友站資訊 2016/02/02 行政院時政講義－「全球招商，臺灣不敗」
最新消息 2016/02/01 105年2月4日博物館商店提早於晚間6:30結束營業
新聞專區 2016/02/01 國立故宮博物院南部院區與策展人有約活動 歡迎各位媒體朋友參加
教育活動預告 2016/02/01 105年「天國的寶藏－教廷文物特展」專題演講暨教師研習（報名時間: 2016/2/1-3/9)
更多
1 2 3 4 5 6 7 8 9 10

43

國立故宮博物院
NATIONAL PALACE MUSEUM
訊息專區 參觀故宮 典藏資源 學習資源 博物館行政 認識故宮
典藏資源
典藏精選
主題網站
文物3D賞析
典藏資料庫系統
加值服務區
故宮電子報
網路商城
電子賀卡
政府資訊公開
回首頁 > 典藏資源 > 典藏資料庫系統
字級設定：
典藏資料庫系統
器物典藏資料檢索系統
書畫典藏資料檢索系統
先秦銅器紋飾資料檢索系統

44

書畫典藏資料檢索系統
典藏檢索
品名
猿
作者
朝代
類型
主題
作品號
檢索
清除畫面
瀏覽檢索.....

45

書畫典藏資料檢索系統
搜尋條件： （書畫） 品名：猿
List 條列式
Graphic 圖文式
5 10 15 20 25 30
共計17筆
唐韓幹猿馬 軸
宋人枇杷猿戲圖 軸
元顏輝畫猿 軸
明宣宗戲猿圖 軸
清郎世寧畫白猿 軸

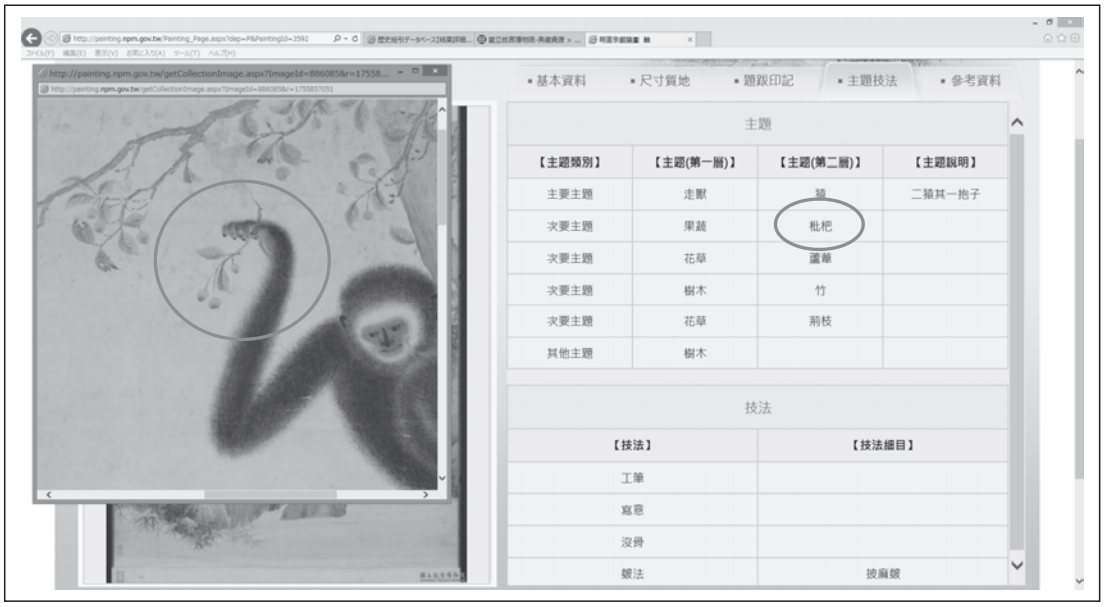
46

基本資料
尺寸質地
題跋印記
主題技法
參考資料
藏品類型 繪畫
作品號 故-畫-000883-00000
品名 明宣宗戲猿圖 軸
作者 朱瞻基,明宣宗,Emperor Xuanzong,Ming Dynasty
書體
色彩 設色
裝裱形式 軸
創作時間 明宣宗宣德二年（1427）
數量 一軸
作品語文
集藏號 件： 開： 幅： 單件
釋文

47

基本資料
尺寸質地
題跋印記
主題技法
參考資料
主題
【主題類別】 【主題(第一層)】 【主題(第二層)】 【主題說明】
主要主題 走獸 猿 二猿其一抱子
次要主題 果蔬 枇杷
次要主題 花草 蘆葦
次要主題 樹木 竹
次要主題 花草 荊枝
其他主題 樹木
技法
【技法】 【技法細目】
工筆
寫意
沒骨
皴法 披麻皴

48


▪ 基本資料
▪ 尺寸質地
▪ 題跋印記
▪ 主題技法
▪ 參考資料
主題
【主題類別】
【主題(第一層)】
【主題(第二層)】
【主題說明】
主要主題
走獸
猿
二猿其一抱子
次要主題
果蔬
枇杷
次要主題
花草
蘆葦
次要主題
樹木
竹
次要主題
花草
荊枝
其他主題
樹木
技法
【技法】
【技法細目】
工筆
寫意
沒骨
皴法
披麻皴

49

網頁展示說明
joy in the relationships found in a family of gibbons. Squatting on a rock, the mother gibbon clasps her baby, the father having plucked loquats for them. The youth has its left arm around its mother's neck and reaches out for the fruit teasingly presented by its father on the other bank of the stream. The varied poses of the gibbons with their animated expressions and actions are complemented by the expressive texture of fur rendered ingeniously with light and dark as well as wet and dry applications of ink. The lines for the thorny shrubs, bamboo, reeds, and water ripples in the setting are also natural and fluid.(20120407)
網頁展示說明
明宣德帝（1399 - 1435）、朱瞻基、詩文や書画を好んだ・落款によれば、本作は宣德2年（1427）、宣德帝29歲の詩の作品である・猿一家の和やかな様子が溫かな筆致で生き生きと描かれている・子猿を抱いた母猿が石の上にうずくまっている・まだ幼い子猿は母猿の首に左手を回してしがみつき、せせらぎを隔てた樹上の父猿に向かって右腕を伸ばしている・父猿は手に果実を持って子猿の気を引いている・3匹の猿の姿はそれぞれ異なり、表情も生き生きとしている・猿の毛並みは墨の濃淡乾溼を巧みに生かして質感が表現されている・猿たちを引き立てるイバラや笹の葉、アシ、波紋などの線もごく自然に描かれている・(20120407)
1/2
上一件
下一件
回上頁

50

画像内容に基づく検索技術に対する期待と現実

国立情報学研究所　北　本　朝　展

画像検索とは、画像データベースの中から利用者が求める画像データに素早く到達するための技術を指す。画像検索はいくつかの技術要素から構成される。まず入力方法としてキーワード（タグ）、テキスト、画像などを選び、検索技術としてインデックス、特徴表現、類似度（距離）定義、フィードバックなどの要素に技術的な工夫を加え、検索結果として画像やテキストなどを出力する。以上の技術的要素を様々に組み合わせて用途に合わせた画像検索を実現することになるが、中でも以下の５つの組み合わせがよく用いられる。(1) テキストを入力し、画像に付与された統制語やタグなどを検索する。(2) テキストを入力し、画像に関連付けられたテキストを検索する。(3) 画像を入力し、画像から抽出したエンティティを検索する。(4) 画像を入力し、類似した画像を検索する。(5) テキストを入力し、画像の自動アノテーションで得られたテキストを検索する。従来型のデジタルアーカイブでは (1) の方式が多く、Google 画像検索のようなウェブ画像検索では (2) の方式が多い。これらも画像検索の一種ではあるが、あくまで画像と関連付けられたテキストの検索であって、画像内容が主な分析対象ではない点に注意しておきたい。

これに対して画像内容に基づく検索、すなわち画像に付与された付加情報ではなく画像に含まれる内容そのものの分析に基づく検索手法にも長い研究の歴史がある。この検索手法は主に３種類に分類できる。(1) 同じもの（エンティティ）を探す。(2) 同じ種類のもの（カテゴリ）を探す、(3) 似ているもの（距離、類似度）を探す。(1) の方法は顔認識を用いた人物検索などが代表的な例であるが、物販（ファッション等）や広告などの分野を中心に商品やブランドの検索などへの応用も進められている。(2) の方法は画像への自動タグ付与に相当するものであり、写真アルバムやストックフォトの整理に使えることが期待されている。一方 (3) の方法は「あんな画像を検索したい」という漠然としたニーズに対応できることが期待されているものの、ニーズが漠然としていることもあって、それほど目覚ましい進展はないというのが現状である。

画像検索の研究はこれまでも地道に積み重ねられてきたものの、その進展はテキスト検索などに比べるとゆっくりとしたものであった。その流れを一変させたのが深層学習（ディープラーニング）の登場である。これはニューラルネットワークを用いた機械学習の一種であり、その原理は1980年代から知られていたものの、ここ５年ほどの間に技術が急速に発展したことで、画像検索の分野にも非連続的な進化が生じることになった。例えば画像の自動タグ付けは人間の精度に匹敵するほど向上し、1000種類以上のタグを自動付与することさえ可能となった。また画像の自動キャプション付与では、画像を与えるだけでそれを説明する文章が生成できるようになった。このように、深層学習の登場以前はしばらく先の技術と思われていた自動生成タグやテキストを用いた画像検索についても、今や実用化へ向けての期待が急速に高まりつつある。

とはいえ、これらの技術が現実的にどれほど有用かは冷静に見極める必要がある。深層学習で大きな成功を収めているのは「教師あり学習」、つまり教師データに合うような出力を生成できるアルゴリズムである。このときアルゴリズムが「正解」を出力できる理由は、概念を理解できたからではなく、関連するデータをたくさん学習したからである。逆に言えば、深層学習モデルに概念を明示的に教えることはできないため、大量の良質な教師データを確保することが精度向上には不可欠となる。こうした良質のデータは、Google や Facebook などの巨大企業であれば自社サービス経由で収集できるが、一般のプレイヤーがそれに匹敵するデータを確保することは至難の業である。そこで、クラウドソーシングやヒューマンコンピュテーションなど、インターネット上の不特定多数の力を集約して生み出された大規模な教師データを研究者コミュニティで共有して利用し、それを共通の基準として研究を比較するアプローチが広がりつつある。

ただし人文学において期待される専門的な情報の検索は、大規模教師データを深層学習に適用するというアプローチとは相性が悪い面もあるため、このアプローチが専門性の高い分野でも有用かどうかは検証する必要がある。第一に、専門的な情報を正確にタグ付けするには専門知識が必要となるため、インターネット上の不特定多数を対象としてさえ適任者の絶対数が少ないという問題がある。第二に、そもそも適任者が少ないため収集できる教師データのサイズも小規模となり、教師あり学習の精度が上がりづらくなる問題がある。第三に、古典籍に現れる線描の絵画のように抽象度が高いデータでは、単純に見えるデータに高度な意味が埋め込まれているため、概念を理解しない機械学習では少数の教師データを有効に活用できない可能性がある。

このように概念という「意味」の扱いは依然として難題であるため、画像検索の研究戦略においては「意味」という問題をできるだけ避けて、物理や統計などの扱いやすい方向からアプローチすることが望ましい。例えば深層学習では、画像の局所的な特徴から画像の大局的な特徴までを多層で表現するモデルを用いるが、そこでは「意味」はニューラルネットワーク上で分散的に表現されているものとされる。ニューラルネットワークが正解を出力するというゴールに至る中間表現においては、意味を明示的に意識して操作する必要は特にないとも言える。

そうは言っても「意味」を避けるだけではいずれ限界に到達する。その限界を越えるには、人間と機械の協調というアプローチ、すなわち人間と機械がチームを組んでお互いが得意なタスクを担当することで、チームとしてのパフォーマンスを最大化することが必要となる。その過程では、大量・高精度な作業を高速化・自動化するタスクを中心に、一部のタスクで人間から機械への置換が進んでいくだろう。しかし高度な判断を下す部分については依然として人間に優位性があるため、少数かつ高度な判断を要する部分では人間を支援して効率化することが機械の役割となる。こうした人間と機械のチーム内分担をどうデザインするかは今後の重要な研究課題である。そして、デジタル人文学でも大規模画像データベースの重要性が増しつつある現在、画像検索技術の継続的な改善には情報学と人文学の研究者が協力するチーム作りも重要な課題となる。

画像内容に基づく検索技術に対する期待と現実

北本 朝展（きたもと あさのぶ）
国立情報学研究所／総合研究大学院大学
http://agora.ex.nii.ac.jp/~kitamoto/

画像検索とは？

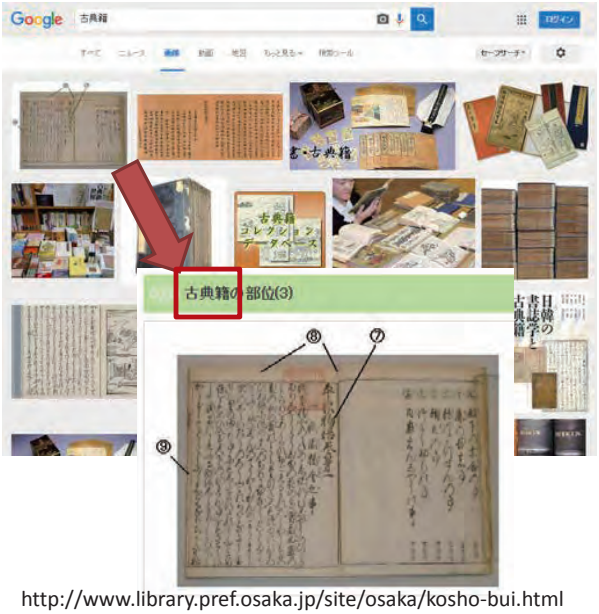
- **画像データベースの中から、利用者が求める画像データに到達するための技術。**
- **入力**：キーワード（タグ）、テキスト、画像
- **技術**：インデックス、特徴表現、類似度（距離）定義、フィードバック…
- **出力**：画像、テキスト
- **網羅的に見るのと比べ、十分効率的か？**

画像検索の主な方式

1. **テキストを入力**し、画像に付与された統制語やタグなどを検索。
2. **テキストを入力**し、画像に関連付けられたテキストを検索。
3. **画像を入力**し、画像から抽出したエンティティを検索。
4. **画像を入力**し、類似した画像を検索。
5. **テキストを入力**し、画像の自動アノテーションで得られたテキストを検索。

ウェブ画像検索

Google画像検索（標準）



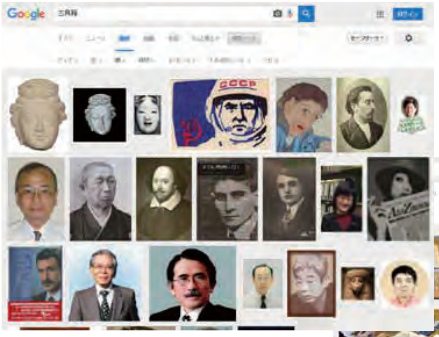
- 画像の周辺に存在するテキストを検索する。
- **画像そのものを分析しているわけではない。**
- 簡単な画像解析オプションはある。

http://www.library.pref.osaka.jp/site/osaka/kosho-bui.html

Google画像検索（色）

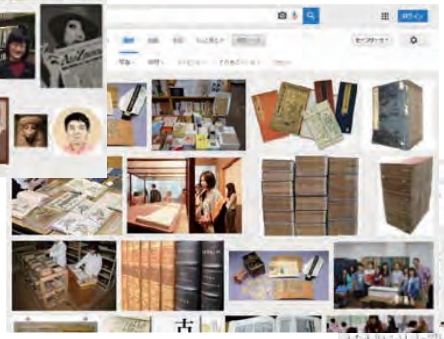
検索オプションに色を指定すると、色に応じて検索結果が変化する。

Google画像検索（種類）

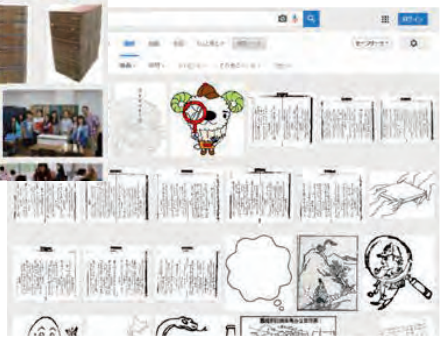


顔

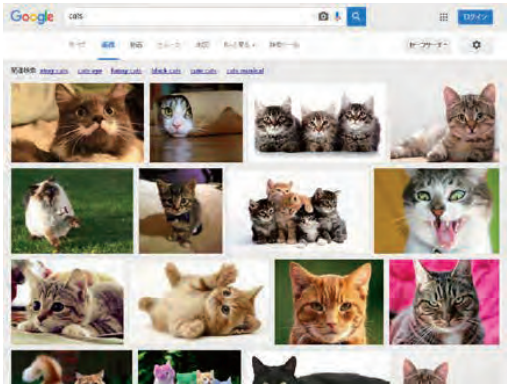
写真

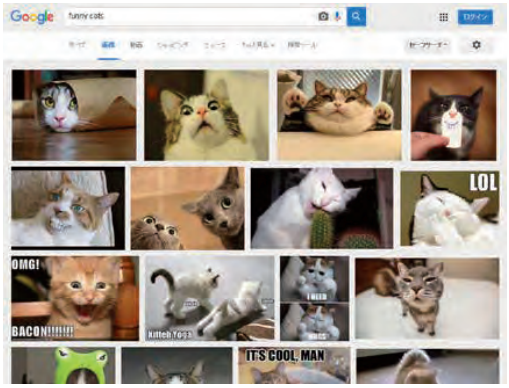


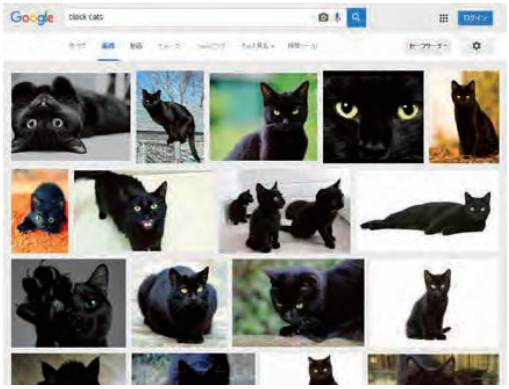
線画

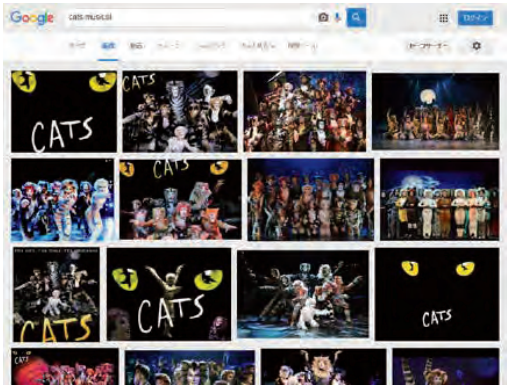


検索オプションに種類を指定すると、種類に応じて検索結果が変化する。









画像内容に基づく検索

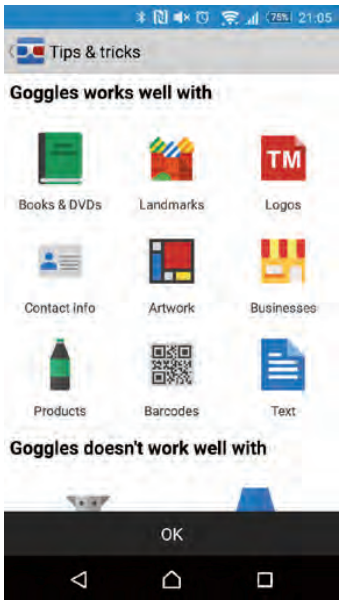
画像内容に基づく検索

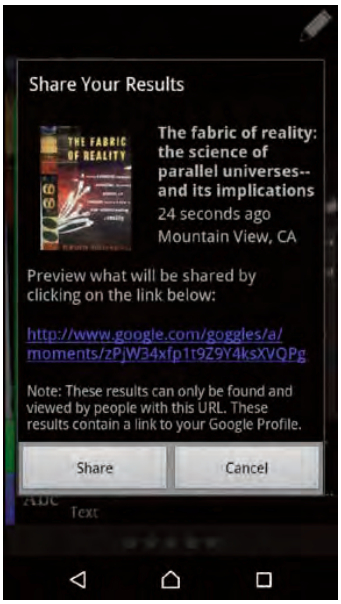
- **画像外のメタデータではなく、画像内に含まれる内容分析を重視する手法。**

1. 同じものを探す（エンティティ）。
2. 同じ種類のものを探す（カテゴリ）。
3. 似ているものを探す（距離、類似度）。

- 1と2はある程度使われているが、3は期待の割にはまだブレークしていない。
- タグ付けによらない検索機能を実現する。

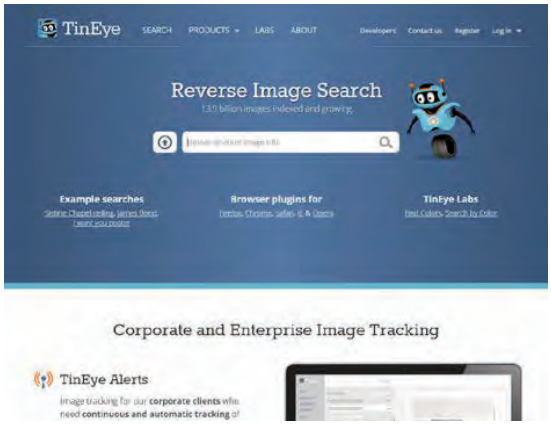
同じものを探す

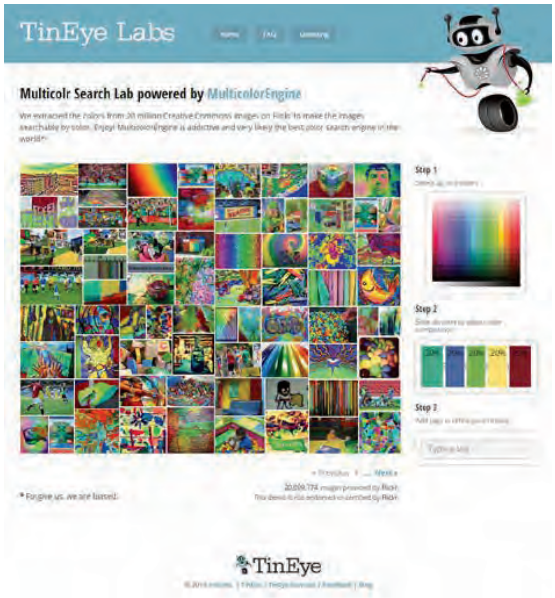




Google Goggles

同じ画像・似た画像を探す



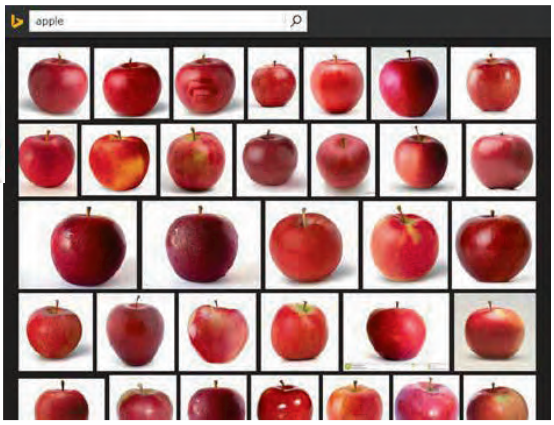


http://www.tineye.com/

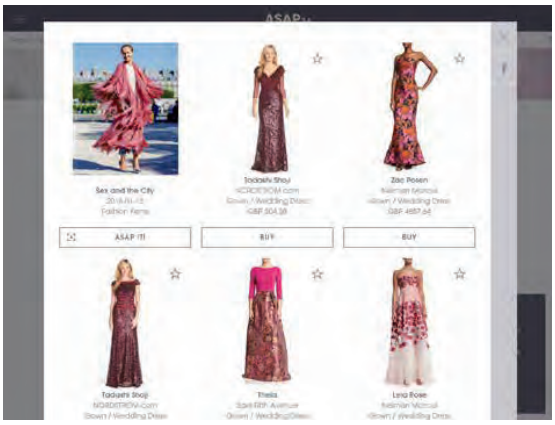
画像の類似性

一枚の画像を選んで類似画像を検索すると、色が考慮されるため、（赤い）りんごだけになる。

Appleで検索した結果。りんごのAppleとAppleのロゴが混ざっている。単語だけでは区別できない。



ショッピングへの応用



町で素敵な人を見つけたら、スマホで撮影し、似た服や靴を検索し、オンラインショッピングへ。

色や形で検索すると、キーワードでは表現できない要素で探せる。

https://www.asap54.com/

ディジタル・シルクロード・プロジェクトでの取り組み

東洋文庫貴重書デジタルアーカイブ

http://dsr.nii.ac.jp/toyobunko/

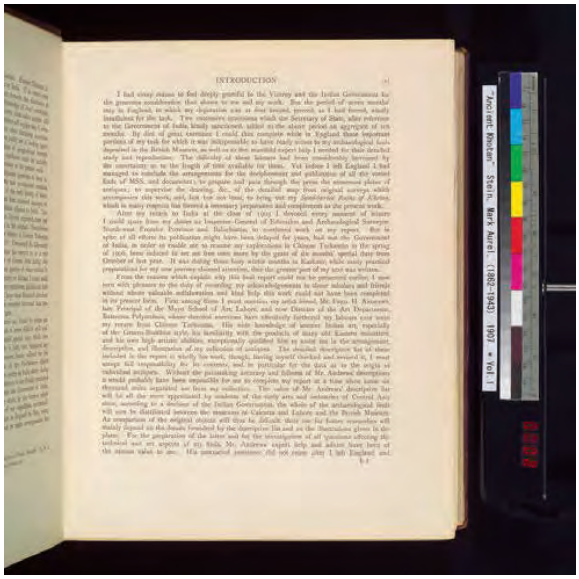
- 237冊の書籍70,898ページをデジタル化。
- 研究コミュニティで必須の書籍を選定。
- キャプションや目次などは人手で入力。
- 全文テキストOCR入力（誤りは未訂正）。

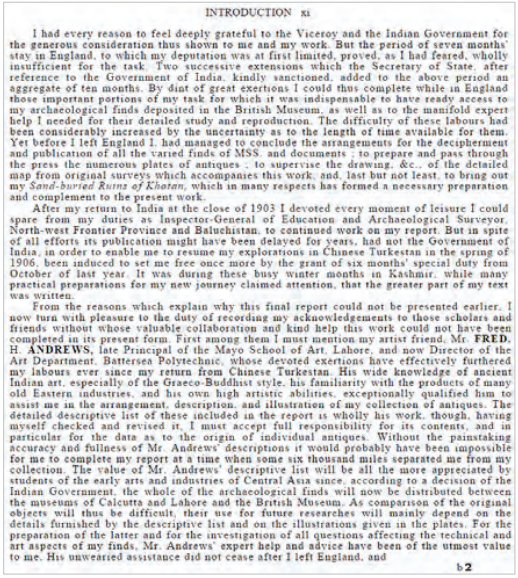
画像切り出し＋画像類似検索

図の部分の自動切り出し

色が類似した画像を検索

文字認識





INTRODUCTION xi

I had every reason to feel deeply grateful to the Viceroy and the Indian Government for the generous consideration thus shown to me and my work. But the period of seven months' stay in England, to which my deputation was at first limited, proved, as I had feared, wholly insufficient for the task. Two successive extensions which the Secretary of State, after reference to the Government of India, kindly sanctioned, added to the above period an aggregate of ten months. By dint of great exertions I could thus complete while in England those important portions of my task for which it was indispensable to have ready access to my archaeological finds deposited in the British Museum, as well as to the manifold expert help I needed for their detailed study and reproduction. The difficulty of these labours had been considerably increased by the uncertainty as to the length of time available for them. Yet before I left England I had managed to conclude the arrangements for the decipherment and publication of all the varied finds of MSS. and documents, to prepare and pass through the press the numerous plates of antiques, to supervise the drawing, &c., of the detailed map from original surveys which accompanies this work, and, last but not least, to bring out my *Sand-buried Ruins of Khotan*, which in many respects has formed a necessary preparation and complement to the present work.

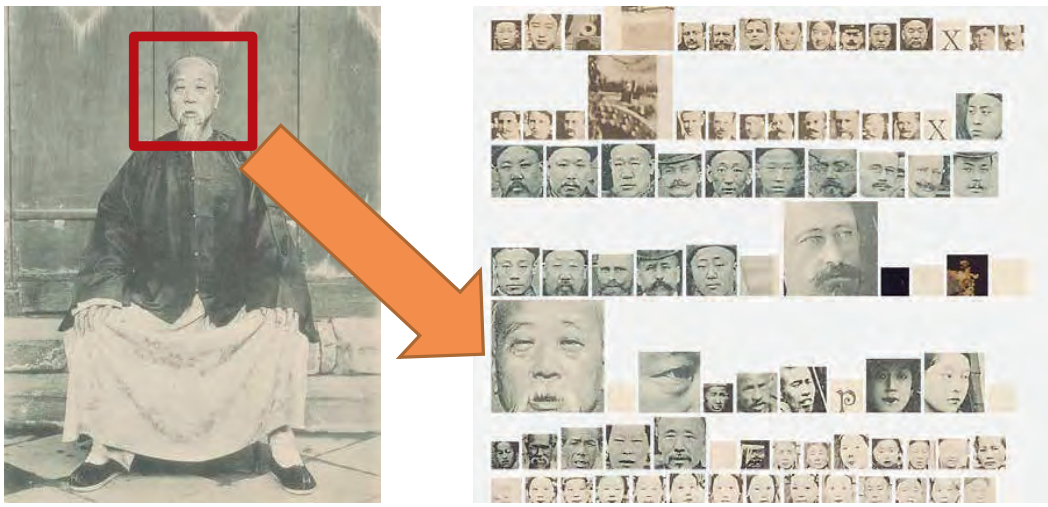
After my return to India at the close of 1903 I devoted every moment of leisure I could spare from my duties as Inspector-General of Education and Archaeological Surveyor, North-west Frontier Province and Baluchistan, to continued work on my report. But in spite of all efforts its publication might have been delayed for years, had not the Government of India, in order to enable me to resume my explorations in Chinese Turkestan in the spring of 1906, been induced to set me free once more by the grant of six months' special duty from October of last year. It was during these busy winter months in Kashmir, while many practical preparations for my new journey claimed attention, that the greater part of my text was written.

From the reasons which explain why this final report could not be presented earlier, I now turn with pleasure to the duty of recording my acknowledgements to those scholars and friends without whose valuable collaboration and kind help this work could not have been completed in its present form. First among them I must mention my artist friend, Mr. **FRED. H. ANDREWS**, late Principal of the Mayo School of Art, Lahore, and now Director of the Art Department, Battersea Polytechnic, whose devoted exertions have effectively furthered my labours ever since my return from Chinese Turkestan. His wide knowledge of ancient Indian art, especially of the Graeco-Buddhist style, his familiarity with the products of many old Eastern industries, and his own high artistic abilities, exceptionally qualified him to assist me in the arrangement, description, and illustration of my collection of antiques. The detailed descriptive list of these included in the report is wholly his work, though, having myself checked and revised it, I must accept full responsibility for its contents, and in particular for the data as to the origin of individual antiques. Without the painstaking accuracy and fullness of Mr. Andrews' descriptions it would probably have been impossible for me to complete my report at a time when some six thousand miles separated me from my collection. The value of Mr. Andrews' descriptive list will be all the more appreciated by students of the early arts and industries of Central Asia since, according to a decision of the Indian Government, the whole of the archaeological finds will now be distributed between the museums of Calcutta and Lahore and the British Museum. As comparison of the original objects will thus be difficult, their use for future researches will mainly depend on the details furnished by the descriptive list and on the illustrations given in the plates. For the preparation of the latter and for the investigation of all questions affecting the technical and art aspects of my finds, Mr. Andrews' expert help and advice have been of the utmost value to me. His unwearied assistance did not cease after I left England, and

b 2

画像分析に基づくが、画像検索とは言わない。

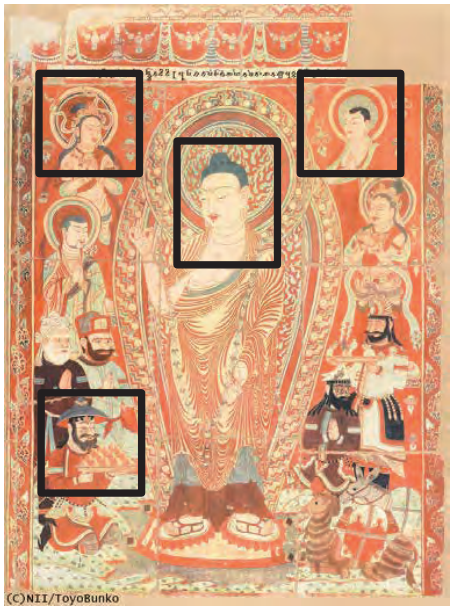
（人間の）顔認識



画像処理ライブラリOpenCVの顔認識アルゴリズムは、古写真でもある程度は使える。

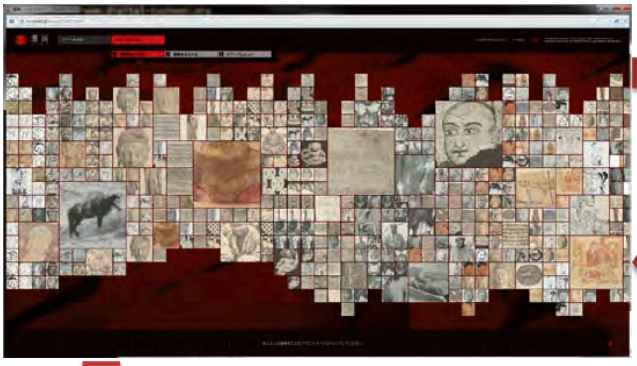
遷画：作品の断片化と再構成

http://dsr.nii.ac.jp/senga/



- 東洋文庫貴重書デジタルアーカイブのデジタル書籍の利活用。
- 画像の一部を断片化し、メタデータを付与して素材画像化。
- 素材画像を利用者が任意に再構成し、新たな作品として公開。

「遷画」の基本構造

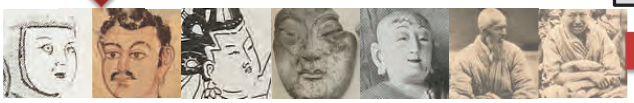


収集支援機能

テーマ検索
エリア検索
類似画像検索
協調画像検索

収集（Collect）

並べ替え（Order）



共有

東洋文庫ミュージアム
ポストカード

21

22

画像の意味の文脈依存性

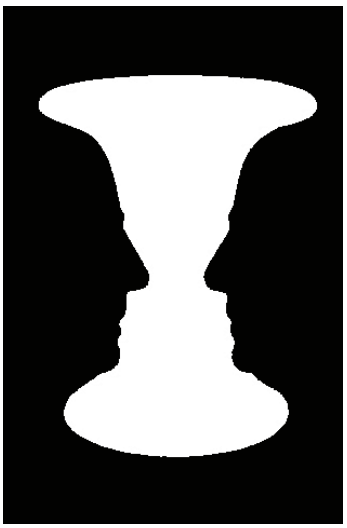
アニマル三昧

仲良し二人組み

飛び出し注意

敦煌 鳥

画像の多義性



ルビンの壺

Wikipediaより

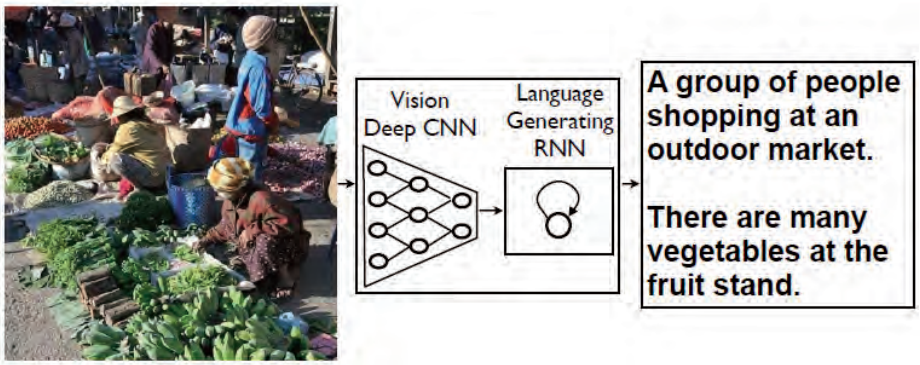
自動画像アノテーションとDEEP LEARNING（深層学習）

自動画像アノテーション

Vinyals, et.al., Show and Tell: A Neural Image Caption Generator, CVPR 2015

ディープラーニング

- 画像とテキストを同じ空間に埋め込む深層学習（Deep Learning）を使って、画像キャプションを自動的に生成。

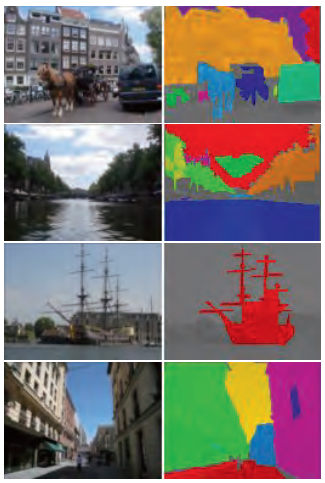


Vinyals, et.al., Show and Tell: A Neural Image Caption Generator, CVPR 2015

深層学習の大流行

- 画像分類問題では、人間に匹敵する分類精度を記録（SuperVision, ILSVRC 2012）
- 教師なし学習では、猫の顔に相当するパターンを自動獲得（Google, ICML 2012）
- 画像学習結果を流用し、「悪夢のような」アート作品を生成（DeepDream）
- 膨大な学習データと莫大な計算機パワーが必要。使うのはそう簡単ではない。

クラウドソーシング



http://labelme.csail.mit.edu/Release3.0/browserTools/php/mechanical_turk.php

- 勝負の鍵：（高品質）学習データが多ければ性能が向上する（特に深層学習）。
- 通常：ネット上の不特定多数の人間に、タグ付け等の単純作業を低額で発注。
- 巨人：GoogleやFacebookは、無料自社サイトに良質な写真データを大量に集積。

ヒューマンコンピュテーション

https://www.google.com/recaptcha/intro/index.html

- **CAPTCHA**：HCの代表的な例。機械と人間の能力差をテスト。
- **reCAPTCHA**：OCR誤りという、機械に難しい問題を人間に解かせる。
- **くずし字版CAPTCHA**：一般人の能力が機械以下だと意味ないが…

ゲームの活用（ESP Game）

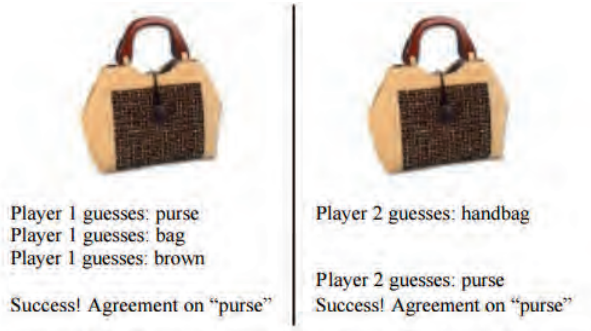


Figure 1. Partners agreeing on an image. Neither of them can see the other's guesses.

Luis von Ahn, et.al. Labeling Images with a Computer Game, Proc. CHI, 2004

- 他者のタグと一致したら成功。他者を想像させ「合意できるタグ付け」を実現。

31

古典籍の画像検索

32

古典籍画像検索の可能性

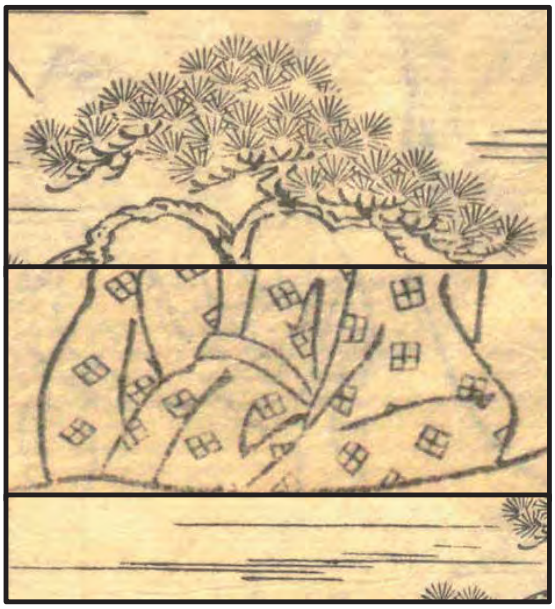
- **タグ検索**：既入力データを学習データとして活用することで、タグ付け作業を支援しスピードアップ。
- **パターン検索**：文様などテキスト化しにくいパターンを自動的に抽出して検索。
- **イラスト（線画）検索**：自然画像（写真）に比べて抽象度が高い（詳細度が低い）ため、意味付けがより難しい。

絵本磯馴松　http://www2.dhii.jp/nijl/NIJL0007/049-0245/049-0245-00010.jpg

抽象化された表現

- 山、川、滝、木。細部の省略による現実世界の高度な抽象化表現。
- 抽象度が高い概念は、自動タグ付けが非常に困難。

規則的なパターン



- 松の葉、着物の文様、奥行き感を出す水平線など。
- 抽象化と同時に、規則性が読み取れるパターンである。
- 検索したいパターンを特定し、データから学習させる。

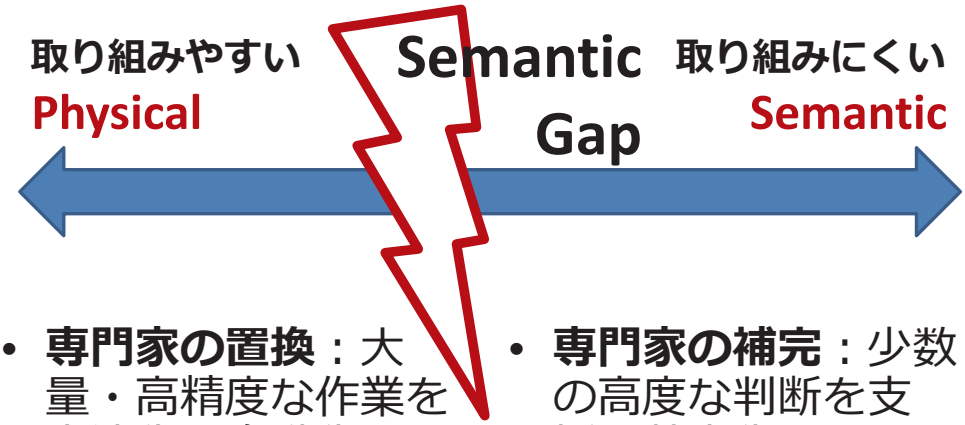
まとめ

画像分析の難易度を考える

- この画像とあの画像は同じものか？
- 同じ本の異なる版はどこ（の画素）が異なるか？
- ある形状と類似した形状はあるか？

- この画像は何を描いたものか？
- この画像と類似した構図や印象の画像はあるか？
- ある概念はどのように描かれているか？

2方向の画像分析戦略



- **専門家の置換**：大量・高精度な作業を高速化・自動化する。
- 例：写本の同一性判定や特徴の分類。

- **専門家の補完**：少数の高度な判断を支援・効率化する。
- 例：画像のフィルタリングやタグ付け。

関連情報

- ディジタル・シルクロード
  - http://dsr.nii.ac.jp/

「デジタルデータ・ツールを活用した人文学研究」に関心があれば、**人文情報学**や**デジタル・ヒューマニティーズ**等の研究会にもご参加下さい。

- 人文科学とコンピュータ研究会
  - http://www.jinmoncom.jp/
- 日本デジタル・ヒューマニティーズ学会
  - http://www.jadh.org/

人文科学におけるオープンサイエンスの課題

第12回人間文化研究情報資源共有化研究会

○主　催　大学共同利用機関法人人間文化研究機構
総合情報発信センター高度連携情報技術委員会
○日　時　平成29年２月３日㈮　13時30分～17時
○開　場　愛知工業大学本山キャンパス３Ｆ大学院講義室１（13：00開場）

○開催趣旨

世界の科学分野においては、論文の根拠となるデータをインターネット上で公開・共有化するオープンデータ、オープンサイエンスの動きが急速に始まっています。データの捏造問題が後を絶たない中、第三者による検証がより容易に可能となるよう、データの共有化を求める動きが進むことは、必然的なものといえるでしょう。また、情報公開の流れのなかで、行政側が主導して、研究機関を含む公的機関が保有するさまざまな情報のオープンサイエンス化を推進する動きもあります。他方、人文科学の分野においては、この問題に対する反応はあまり迅速ではないように思われます。

本年度の資源共有化研究会では、国の施策のなかでのオープンサイエンスをめぐる最新の状況についての認識を深めるとともに、人文科学の分野でオープンサイエンス化を進めるうえで、どのようなことが課題となるのか、また研究の進展や研究成果の社会化のためにはどのようなオープン化が必要であり、望ましいのか、といった問題について意見をかわす場としたいと思います。

○プログラム

13：30～13：35　問題提起　人間文化研究機構理事　榎原　雅治
13：35～14：20　「オープンサイエンスが生まれた背景と最近の政策の動き」
文部科学省科学技術・学術政策研究所　林　和弘
14：20～14：50　質疑
15：00～15：25　「歴史資料のオープンデータ化に関する現在と未来
――歴博の総合資料学の取り組みを通じて――」
国立歴史民俗博物館研究部　後藤　真
15：25～15：50　「歴史的典籍NW事業におけるオープンデータ
――その戦略と課題――」
国文学研資料館古典籍共同研究事業センター　山本　和明
15：50～16：15　「言語研究と「オープン」データ」
国立国語研究所コーパス開発センター　前川　喜久雄
16：15～17：00　総合討論

1

自己紹介
ORCID ID: 0000-0003-1996-4259

- 1995年頃の電子ジャーナル化が本格化する頃より、ドメイン（東京大学・理学部（有機合成化学））を持ちながら学術情報流通の変革に実地で参画（化合物データ管理）
- 日本化学会にて、電子投稿査読開発、電子ジャーナル化、世界最速レベルの出版体制構築、 ビジネスモデルの確立、オープンアクセス対応などをこなす（電子付録（データ）対応）
- 学術情報流通を俯瞰する過程で化学に限らない学術情報流通の将来と研究活動基盤自体の変革に興味を持つ
- 2012年より科学技術・学術政策研究所で、科学技術予測調査の傍ら、オープンアクセス、オープンサイエンス政策などの調査研究と実装に取り組む（内閣府、RDA、G7）

2

多様な“帽子”

- [学術出版界、産業界]
  - ALPSP理事 (2011)
  - J-STAGE & JaLC (2001-)
- [図書館とオープンアクセス]
  - SPARC Japan Steering Committee (2007-)
  - Open Access Week International Adviser Meeting (2014-)
- [科学コミュニティ]
  - 日本学術会議特任連携会員 (2010-2014)
  - IUPAC 国際純粋応用化学連合　会員 (201)
- [政策と行政]
  - NISTEP, 文部科学省、内閣府 (2006-)

The Association of Learned & Professional Society Publishers

J-STAGE

Japan Link Center

SPARC Japan
SPARC : Scholarly Publishing and Academic Resources Coalition

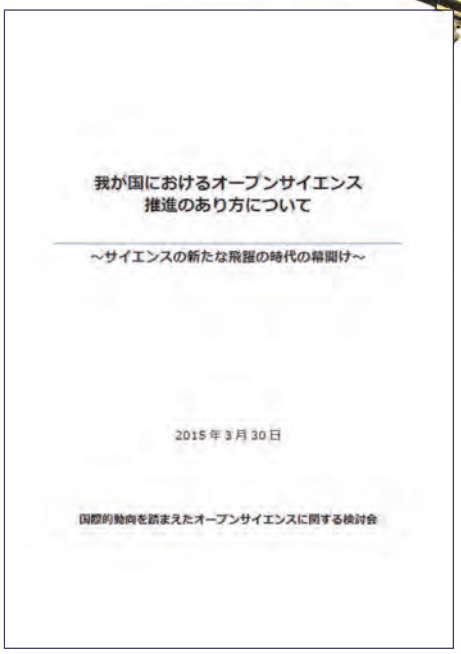
SPARC

日本学術会議
SCIENCE COUNCIL OF JAPAN

IUPAC

3

国際的動向を踏まえたオープンサイエンスに関する検討会 (内閣府)



4

4

0 本日の構成

オープンアクセス、オープンサイエンスの展望

1. 学術雑誌の電子化と論文のオープンアクセスから研究のオープンサイエンスへ
2. オープンサイエンスの多様性

研究データ利活用促進

3 科学技術・学術にとっての研究データの利活用
4. 最近の政策の動向
5. 人文系研究に与える示唆

5

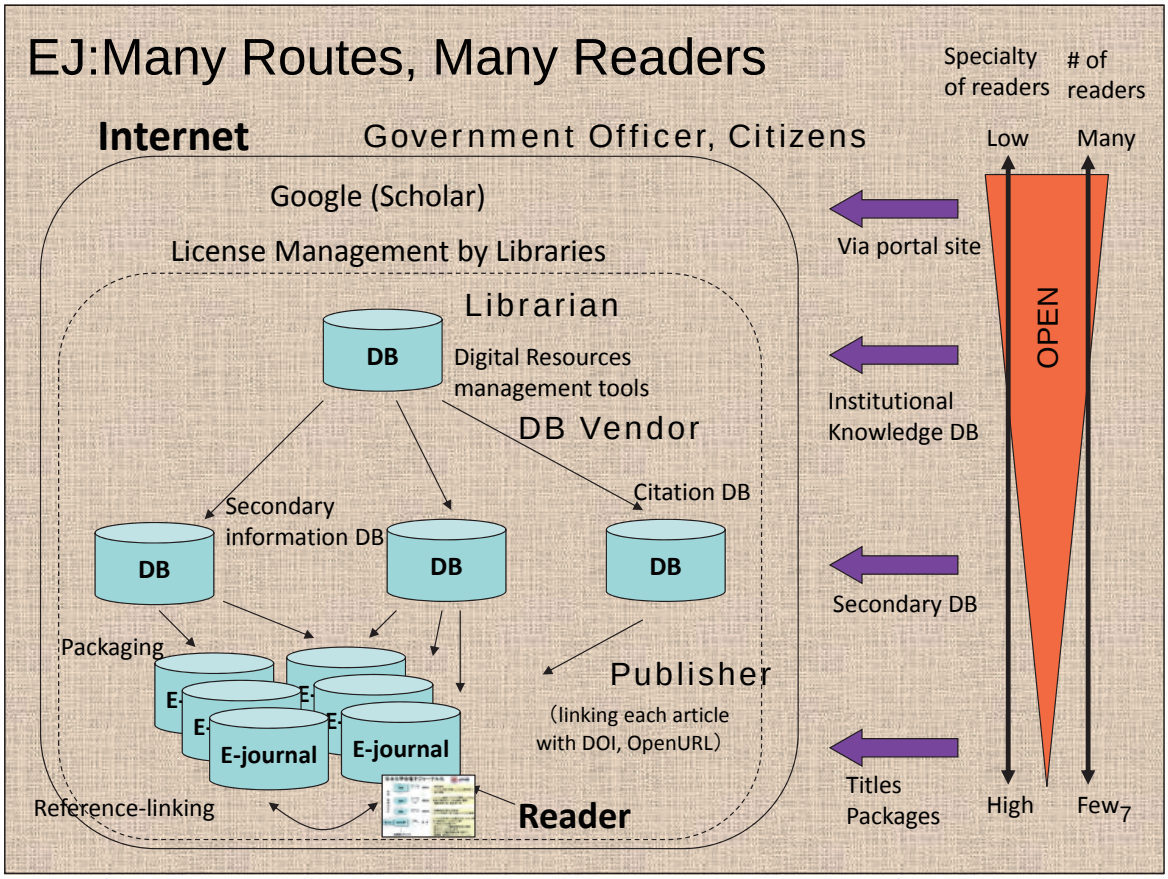
5

1 論文のオープンアクセスから研究のオープンサイエンスへ

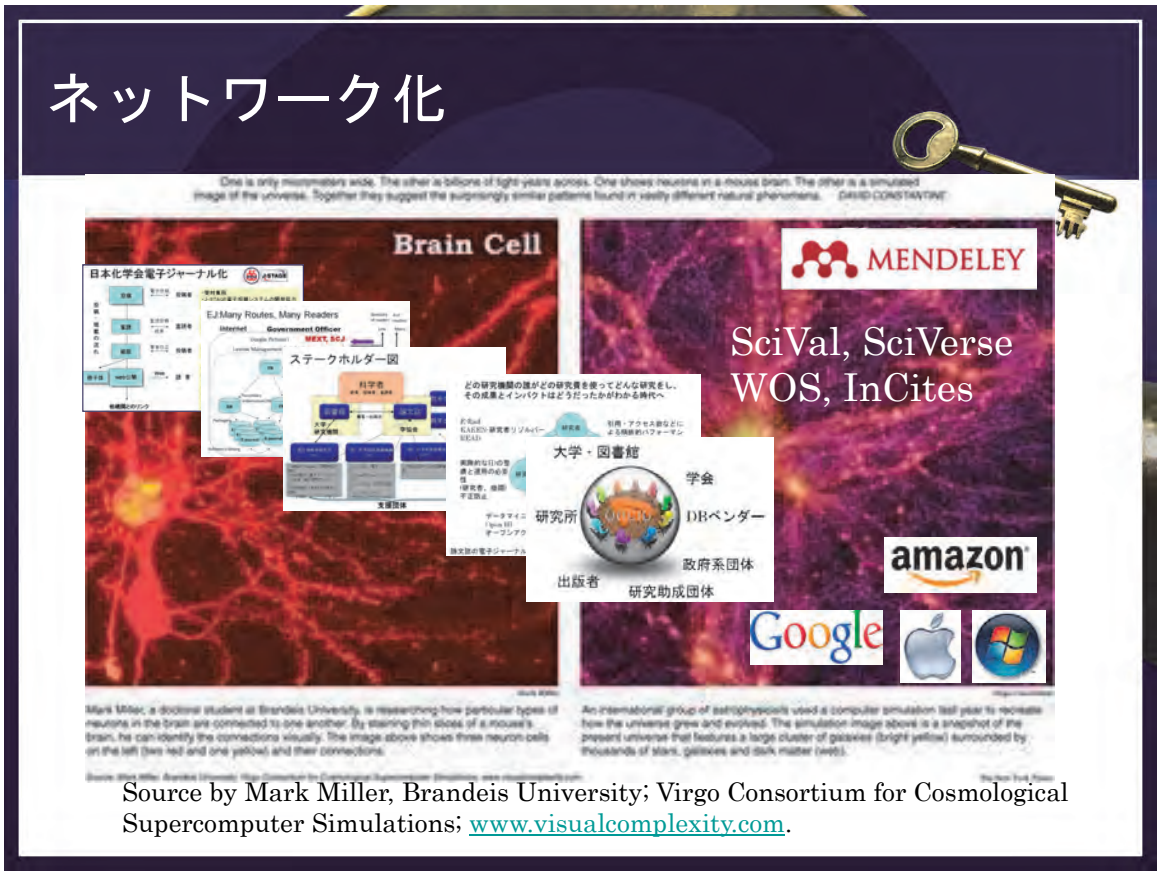
1−1　論文の「電子化」からネットワーク化
1−2　アクセスから再利用、出版から共有へ
1−3　過渡期と変化の加速、非連続変化
1−4　研究活動のオープン化と研究「成果」の共有・利活用

6

6



7



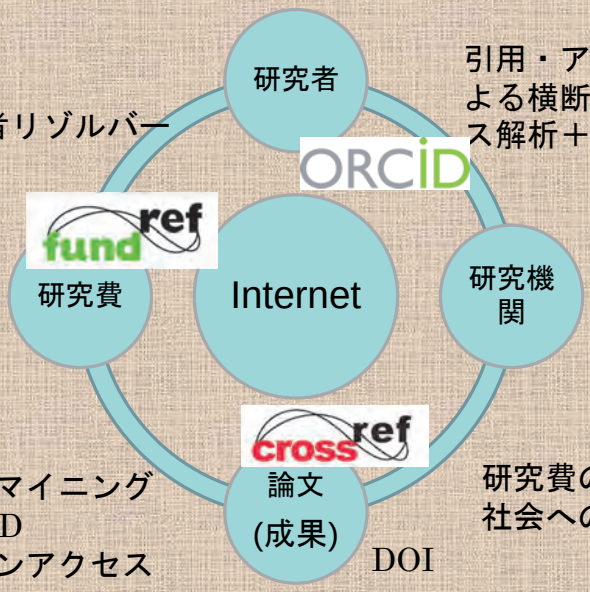
Source by Mark Miller, Brandeis University; Virgo Consortium for Cosmological Supercomputer Simulations; www.visualcomplexity.com.

8

ID（識別子）の重要性

どの研究機関の誰がどの研究費を使ってどんな研究をし、その成果とインパクトはどうだったかがわかる時代へ

ORCID
E-Rad
KAKEN-研究者リゾルバー
READ



引用・アクセス数などによる横断的パフォーマンス解析＋Altmetrics

SciVal
InCites™
Calibrate Your Strategic Research Vision

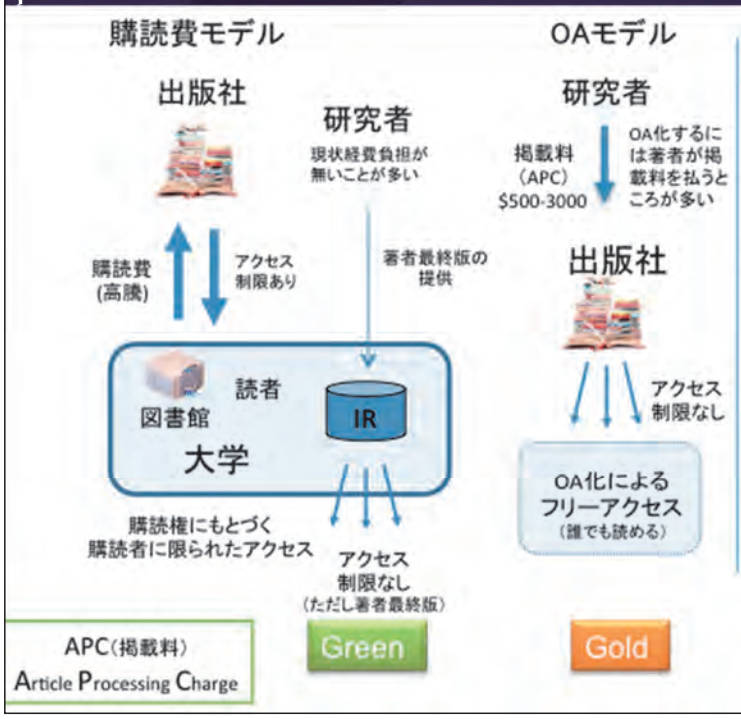
データマイニング
Open ID
オープンアクセス

研究費の透明性の確保
社会への説明

論文誌の電子ジャーナルをめぐる最近の動き，科学技術動向, 2009/7, 100, 10-18.
(一部改変)

9

オープンアクセス（OA）と手段　Green & Gold



読み手が払うか
（購読費モデル）
書き手が払うか
（Gold OA モデル）

読めない人への別の手段を提供するか
（購読費ジャーナルに対するGreen OAの付加）

科学技術動向2013

10

潮目の変化

旧来の図書館と出版者の枠組みを超えた動き

2010年頃から

- Gold OA journal “Rush”
- OA mega-journal
- Mendeley, ResearchGate
- Altmetrics
- Rubriq
- Data journal
- figshare

SPRINGER NATURE

11

11

歴史に習えば

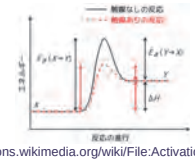
- **ポストグーテンベルグの過渡期に居る我々**

Print based dissemination

Letter based dissemination

Past Design

Transition State



http://commons.wikimedia.org/wiki/File:Activation_energy_ja.svg

Web native dissemination

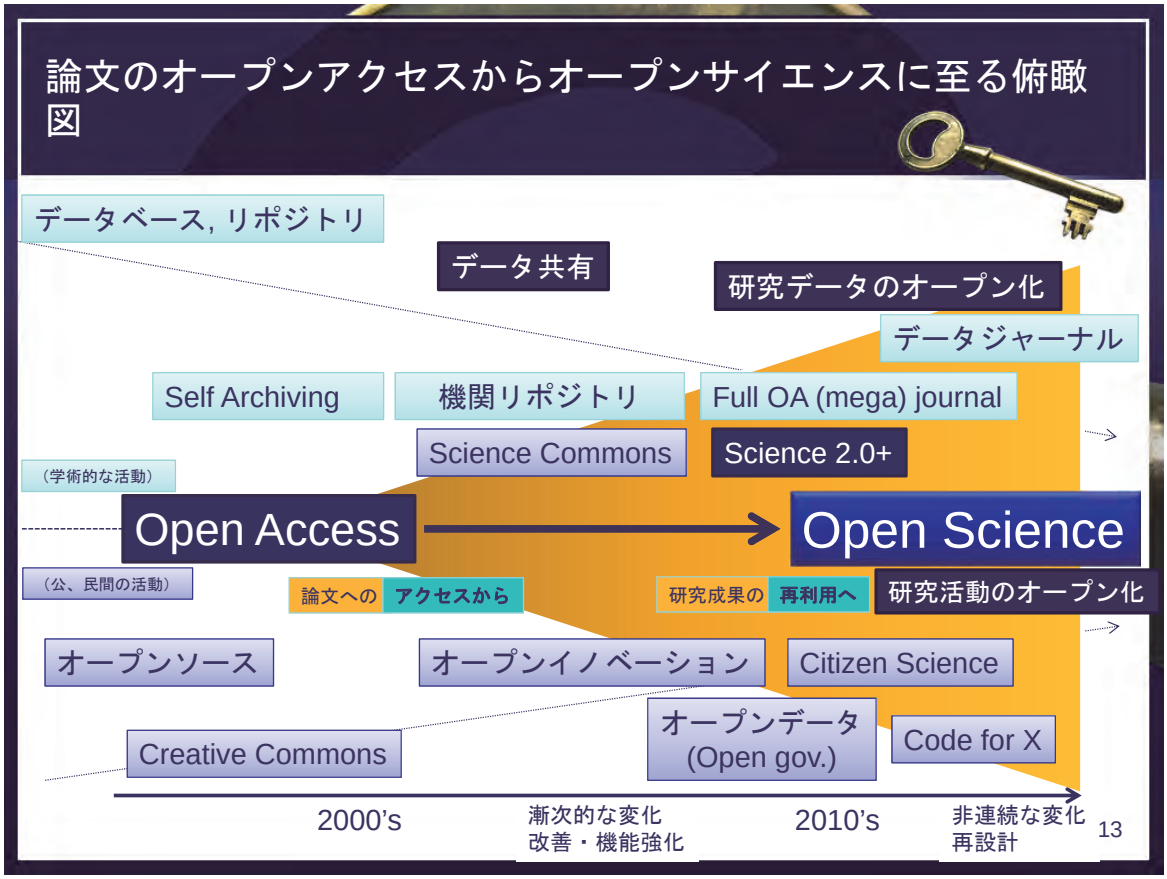
Future Design

学術情報流通の再発明

K. Hayashi, “Current States of Impact Assessment of Research Outputs in Japan and Some Challenges to Measure New Impacts for Japan’s Stakeholders,” OECD-ESTONIA WORKSHOP ON IMPACT ASSESSMENT: PRACTICES, TECHNIQUES AND POLICY CHALLENGES, May 15-16 2014, Estonia. (revised)

12

12



13

学術情報流通を取り巻くアイテム、サービスの変革

	基準	変革第初段階	変革次段階	不連続変革
アイテム	対象	対象の電子化	新しい価値の付加	別業種、新規ステークホルダーの参入、異なる視点からの価値の付与、サービスの実装
ジャーナル	冊子体	PDF	Xhtml データベースとの連携 動画ジャーナル	（データ出版）
査読	Peer Review	電子査読システム	Open Peer Review OAメガジャーナル用簡易Review	Altmetrics等を利用した事後レビュー
文献管理	ファイリング	EndNote（初期）	RefWorks	Zotero, Mendeley、ReadCube
購読・配信	発送ベースの購読管理	IP、ID管理	パッケージとビッグディール	オープンアクセス
書籍	紙の書籍	PDF	ePuB(eBook)、独自フォーマット	
蔵書管理	目録	OPAC	WebCat, World Cat	カーリル、ディスカバリーサービス、Amazon
授業	プリント授業	ppt利用	OCW(Open Course Ware)	MOOC
板書	黒板	電子黒板	インタラクティブホワイトボード	MOOC上のスクリーン
目的	紙、物流、郵送ベースの仕組みで目的を達成する手段	アイテムのデジタル化、WWW対応	前段階をベースにインクリメンタルに革新することが繰り返される	アイテムの本来の目的に（結果的に）立ち返り、別の手段、パラダイムで目的を実現する

*あくまで例示であり、各要素、サービスごとに、1つの見方を切り取って紹介している場合もある

14

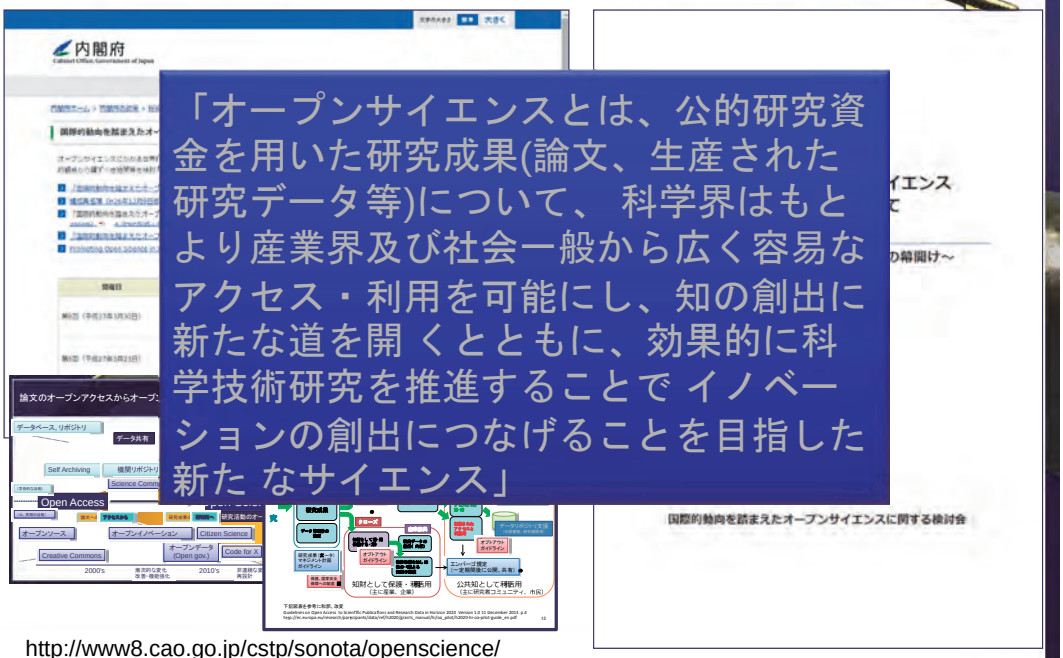
２オープンサイエンスを推進する意味

2-1　科学を変えるオープンサイエンス
2-2　産業を変えるオープンサイエンス
2-3　市民を変えるオープンサイエンス
2-4　科学技術・学術の再構成、社会変容と政策
　　（オープン［デジタル］サイエンス）
2-5　研究公正との親和性

15

15

内閣府：我が国におけるオープンサイエンス推進のあり方について



「オープンサイエンスとは、公的研究資金を用いた研究成果(論文、生産された研究データ等)について、 科学界はもとより産業界及び社会一般から広く容易なアクセス・利用を可能にし、知の創出に新たな道を開 くとともに、効果的に科学技術研究を推進することで イノベーションの創出につなげることを目指した新た なサイエンス」

http://www8.cao.go.jp/cstp/sonota/openscience/

16

オープンサイエンスの主な要素

- Science 2.0
  - 科学そのものが変わる
    - Data Driven Science
    - Collaborative Team Science on a platform
- Open Innovation
  - 技術、産業、知財の在り方が変わる
    - Industry 4.0
    - 著作権や特許の制度疲弊
- Citizen Science
  - 市民の科学技術への関与が変わる
    - 科学者の範囲が広がる
    - 市民の科学技術政策への参画

図書館とは何か？

大学とは何か？

17

17

OAの潜在的便益（政策的観点）

- 研究を加速し成果を見つけやすくすることで<u>研究開発投資の費用対効果</u>を上げる
- 同じ研究を繰り返すこと避け、研究開発コストを抑える
- 境界領域や多領域にまたがる研究の機会を増やし、<u>多分野の協調</u>を促す
- 研究結果の商業化を早く広い観点から行い、公共研究開発投資の効果を上げ、科学情報を基にした<u>新しい産業</u>を生み出す

Fact sheet: Open Access in Horizon 2020

https://ec.europa.eu/programmes/horizon2020/sites/horizon2020/files/FactSheet_Open_Access.pdf

18

18

G8 2013 Science Ministers' Agreement of Open Research Data

G8 Science Ministers Statement London UK, 12

Introduction

We, the G8 Science Ministers met in London on Wednesd
of our respective national science academies, as part of th
this unique meeting we discussed how our nations could le
transparency, coherence and coordination of the global sc
in order to address global challenges and maximise the so
of research.

3. Open Scientific Research Data

Open enquiry is at the heart of scientific endeavour, and rapid technological change has profound implications for the way that science is both conducted and its results communicated. It can provide society with the necessary information to solve global challenges. We are committed to openness in scientific research data to speed up the progress of scientific discovery, create innovation, ensure that the results of

4. Expanding Access to Scientific Research Results

G8 Open Data Charter will 'increase transparency' and 'fuel innovation'

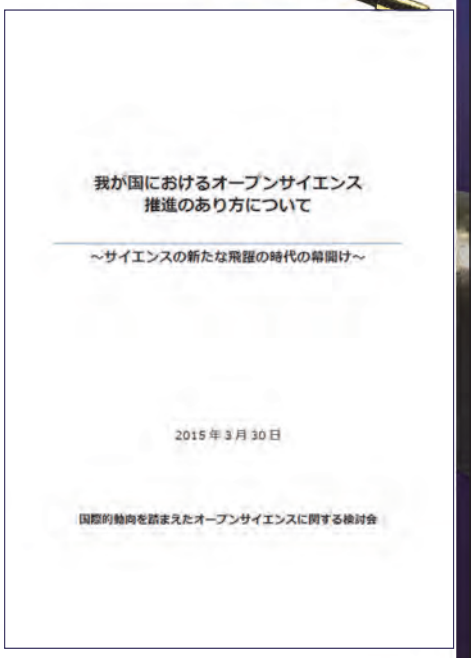
Five key principles outlines how governments should release datasets for economic and social benefits

"Open Government Data"

Y. Murayama

19

内閣府：我が国におけるオープンサイエンス推進のあり方について



http://www8.cao.go.jp/cstp/sonota/openscience/

20

日本の政策の動き

第4章「科学技術イノベーションの基盤的な力の強化」

・今後起こり得る様々な変化に対して柔軟かつ的確に対応するため、若手人材の育成・活躍促進と大学の改革・機能強化を中心に、基盤的な力の抜本的強化に向けた取組を進める。

（2）「知の基盤の強化」③「オープンサイエンスの推進」

・オープンサイエンスの推進体制を構築し、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。

第4章2節3項「オープンサイエンスの推進」

21

21

G7 科学技術大臣会合　つくば　2016.5

G7 Science and Technology Ministers' Meeting
Tsukuba, Ibaraki

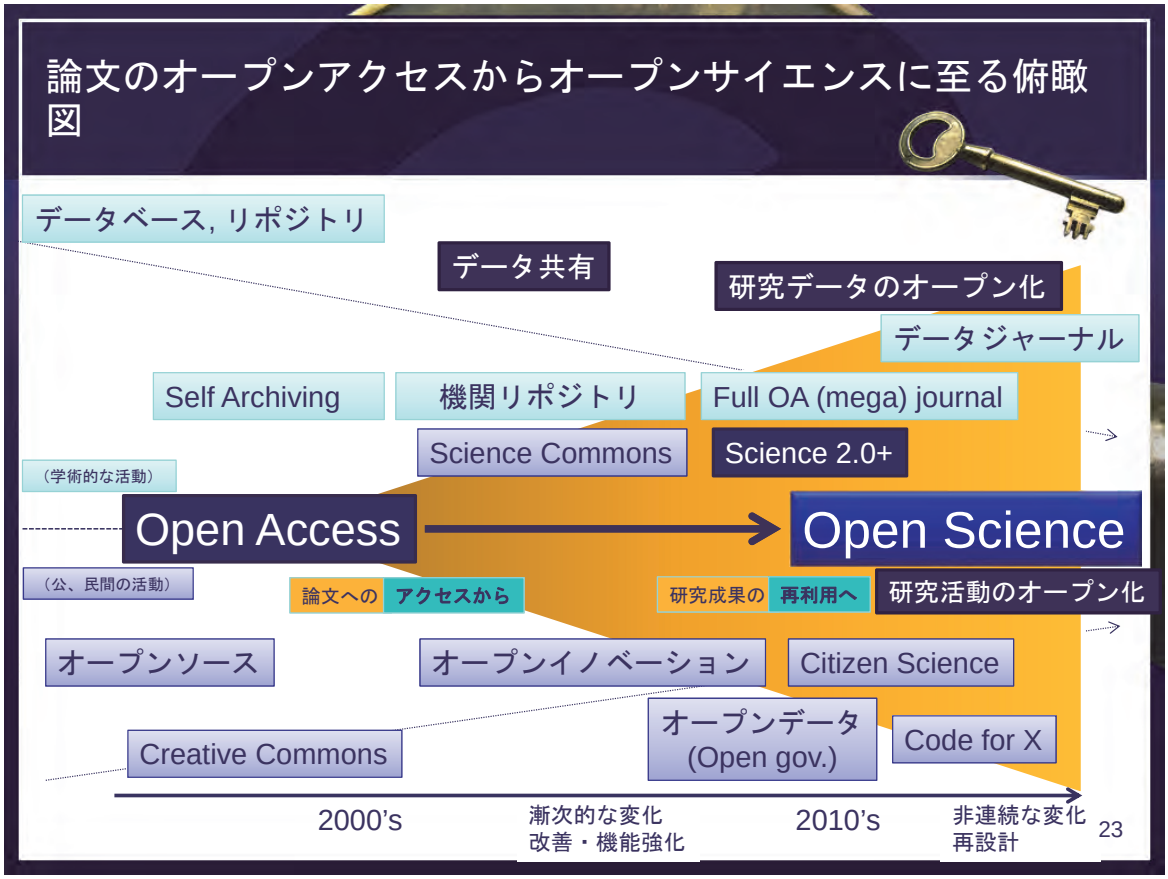
6: Open Science - Entering into a New Era for Science:

Putting into Practice New Framework of Research and Knowledge Discovery, Sharing, and Utilization

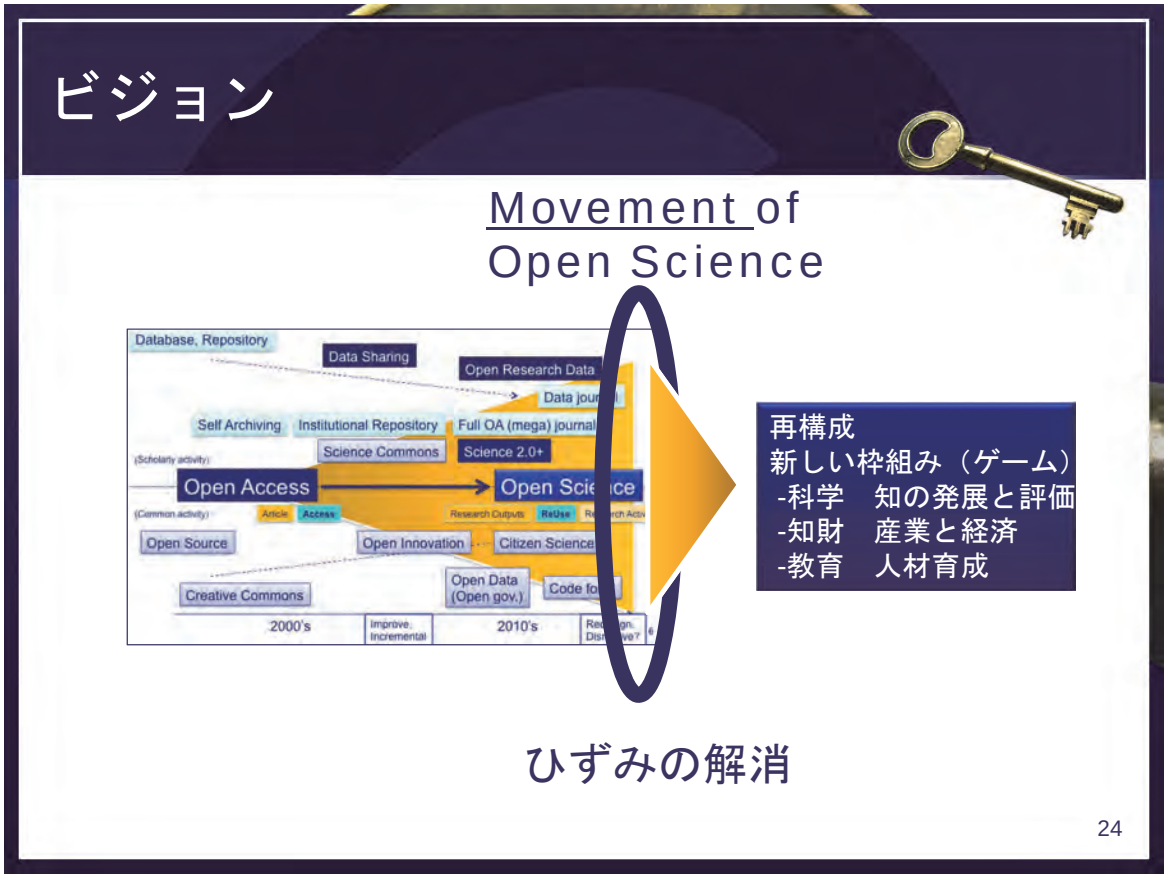
i. オープンサイエンスに関する作業部会を設置して、OECD といった国際機関との連携 を視野に入れたオープンサイエンスのポリシーの共有、インセンティブの仕組みの検 討、公的資金による研究成果の利用促進のためのグッドプラクティスの特定を行うこ と。

ii. オープンサイエンスが有効に活用され、すべての人がメリットを享受できるようにする ために、国際的な協調や連携を推進して、デジタルネットワークの整備、人材の確保など、適切な技術やインフラを整備すること。

22


論文のオープンアクセスからオープンサイエンスに至る俯瞰図
データベース, リポジトリ
データ共有
研究データのオープン化
データジャーナル
Self Archiving
機関リポジトリ
Full OA (mega) journal
Science Commons
Science 2.0+
（学術的な活動）
Open Access
Open Science
（公、民間の活動）
論文への アクセスから
研究成果の 再利用へ
研究活動のオープン化
オープンソース
オープンイノベーション
Citizen Science
オープンデータ (Open gov.)
Code for X
Creative Commons
2000's
漸次的な変化 改善・機能強化
2010's
非連続な変化 再設計
23

23


ビジョン
Movement of Open Science
Database, Repository
Data Sharing
Open Research Data
Self Archiving
Institutional Repository
Full OA (mega) journal
Science Commons
Science 2.0+
Open Access
Open Science
Open Source
Open Innovation
Citizen Science
Creative Commons
Open Data (Open gov.)
2000's
2010's
再構成
新しい枠組み（ゲーム）
-科学　知の発展と評価
-知財　産業と経済
-教育　人材育成
ひずみの解消
24

24

ゲームチェンジ

SNS上で論文を（勝手に）シェアすることが議論を呼んでいる

この月刊誌、週刊誌双方の雑誌の2ケタマイナスは、出版状況がもはや臨界点にまで至ったことを告げている。

返品率は書籍が40.5%、雑誌は42.9%で、こちらも同様だといっていい。

本クロニクルなどで繰り返し記してきたように、近代出版流通システムは雑誌をベースとして構築され、それに書籍が相乗りするようなかたちで営まれてきた。そのビジネスモデルがついに崩壊しようとしている。

http://jp.wsj.com/articles/SB10519349150193173538704581499900801192030

http://d.hatena.ne.jp/OdaMitsuo/20151201/1448895608

25

25

ゲームチェンジの兆し

ABOUT

カザフスタンの科学者アレクサンドラ・エルバキアンのつくった論文交換サイト

Sci-Hub

the first pirate website in the world to provide mass and public access to tens of millions of research papers

A research paper is a special publication written by scientists to be read by other researchers. Papers are *primary sources* neccessary for research – for example, they contain detailed description of new results and experiments.

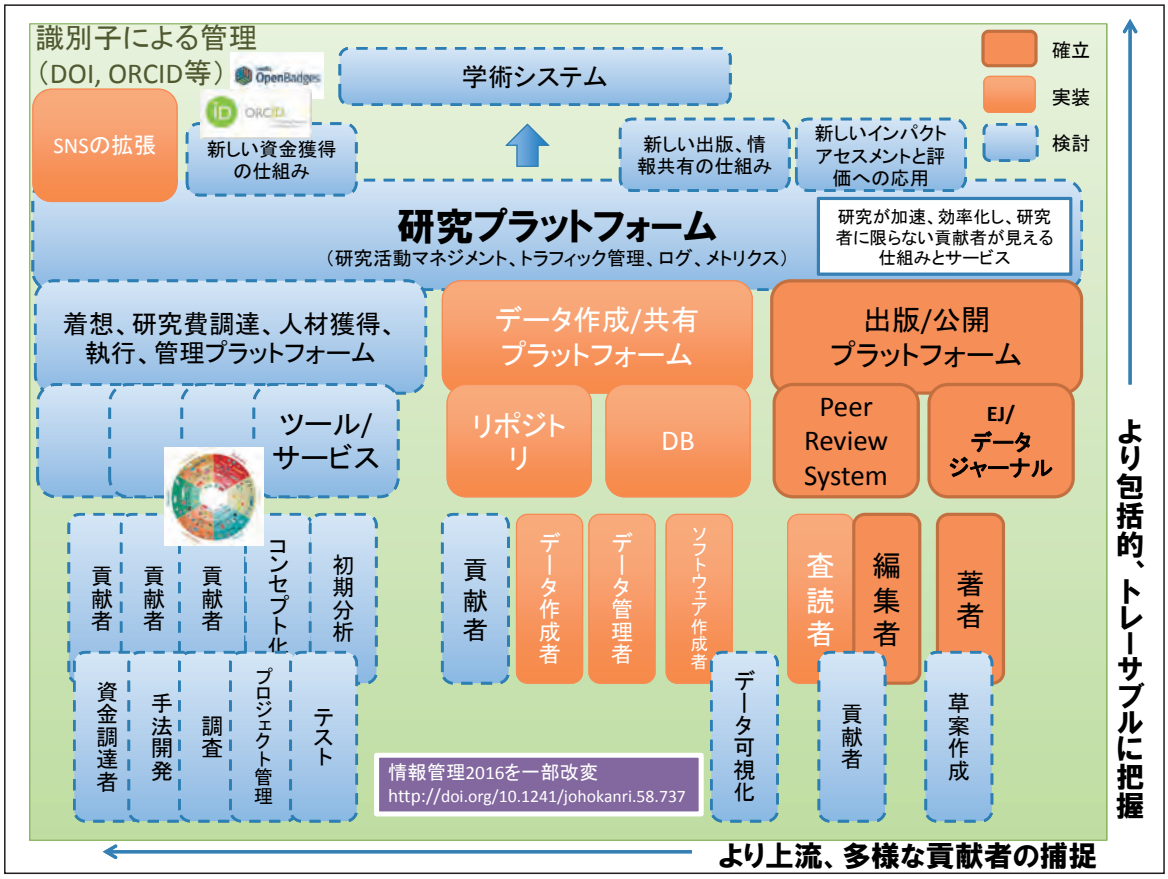
「科学の発展」のために5000万件以上の研究論文を無料で読めるようにしている海賊版（義賊版?）サイト「Sci-Hub」

http://sci-hub.cc/26

26

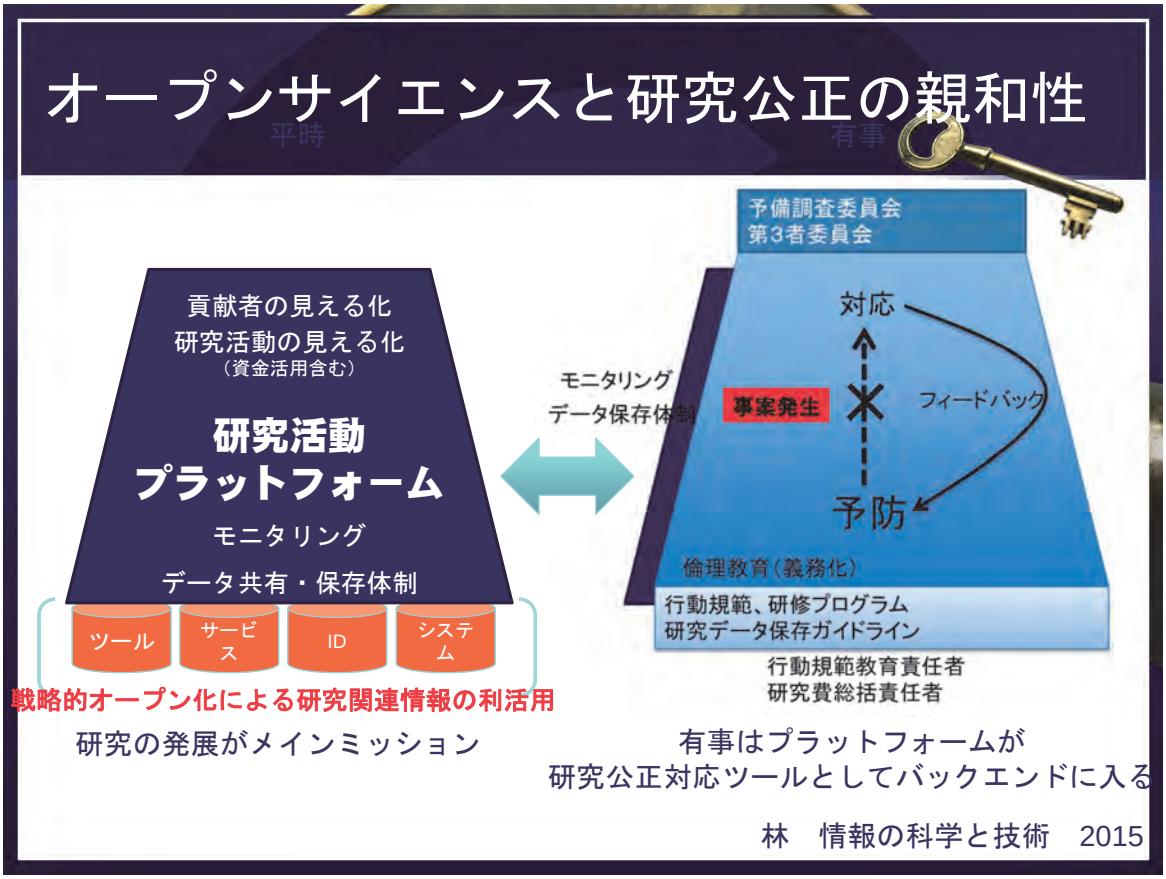
確立
実装
検討
データ作成/共有
プラットフォーム
出版/公開
プラットフォーム
リポジトリ
DB
Peer Review System
EJ/データジャーナル
貢献者
データ作成者
データ管理者
ソフトウェア作成者
査読者
編集者
著者
データ可視化
貢献者
草案作成
情報管理2016を一部改変
http://doi.org/10.1241/johokanri.58.737
より上流、多様な貢献者の捕捉

27


識別子による管理
（DOI, ORCID等）
OpenBadges
ORCID
学術システム
SNSの拡張
新しい資金獲得の仕組み
新しい出版、情報共有の仕組み
新しいインパクトアセスメントと評価への応用
確立
実装
検討
研究プラットフォーム
（研究活動マネジメント、トラフィック管理、ログ、メトリクス）
研究が加速、効率化し、研究者に限らない貢献者が見える仕組みとサービス
着想、研究費調達、人材獲得、執行、管理プラットフォーム
ツール/サービス
データ作成/共有
プラットフォーム
出版/公開
プラットフォーム
リポジトリ
DB
Peer Review System
EJ/データジャーナル
貢献者
貢献者
貢献者
コンセプト化
初期分析
資金調達者
手法開発
調査
プロジェクト管理
テスト
貢献者
データ作成者
データ管理者
ソフトウェア作成者
査読者
編集者
著者
データ可視化
貢献者
草案作成
情報管理2016を一部改変
http://doi.org/10.1241/johokanri.58.737
より上流、多様な貢献者の捕捉
より包括的、トレーサブルに把握

28

オープンサイエンスと研究公正の親和性



林　情報の科学と技術　2015

29

Landscape of Open Science/Research Data Sharing

2008- ICSU-World Data System

1966- CODATA (Committee on Data for Science & Technology, ICSU

2012- Research Data Alliance

Group on Earth Observation
Global Earth Observation System of Systems

International Council for Science

Future Earth (ICSU, UNESCO, UNEP, UNU, Belmont Forum,…)

European Open Science Cloud
ESFRI、EUDAT、GEANT. . .

OECD：
Projects of Intl. e-Infra./ interoperable policy

Y. Murayama

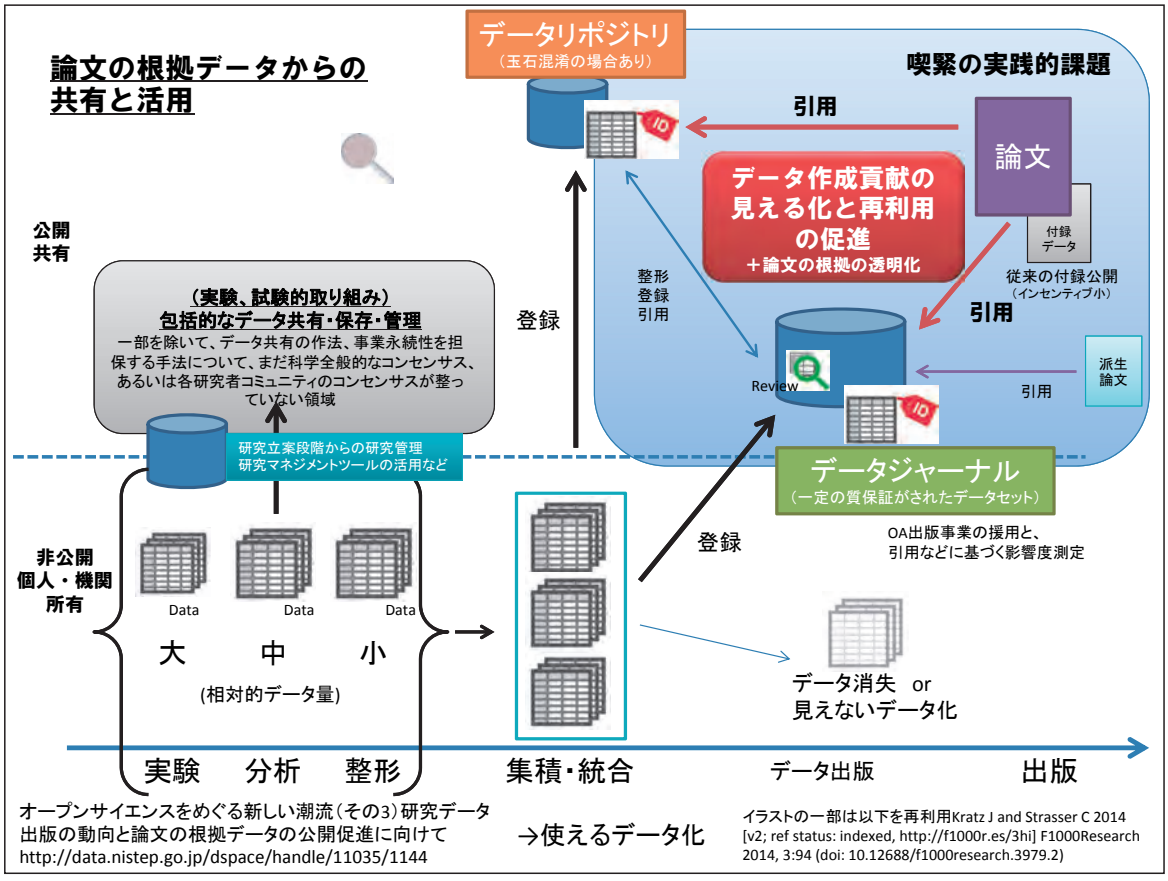
30

３科学技術・学術にとっての研究データの利活用

3−1　論文の付録からデータジャーナルへ
3−2　データリポジトリから、研究データ基盤へ
3−3　欧州サイエンスクラウドの野望と現実
3−4　国内外のオープンサイエンスの議論から見える現状

31

31



32

研究データ共有・利活用の現状

- データマネジメント
  - システム
  - プラン
  - ビジネスモデル
  - 人材
- モニタリング（観測）
  - プラットフォーム化
  - 研究活動のモデル化とシステム化
  - 多様性への対応（Herding Cats、Long Tail）
  - 研究公正対応にもうまく活用

（林、第3回オープンサイエンスデータ推進ワークショップ、京都大学, 2016.9）

33

33

欧州サイエンスクラウド（EOSC）

- 2015年5月Digital Single Market Strategy（デジタル単一市場戦略；DSM）
  - デジタル技術に基づく情報利用・サービス、ネットワークや経済の向上を実現（5億人、50兆円）
  - データ・情報通信の標準化および相互運用性（interoperability）の確保が優先事項
  - EUの試算によればEuropean Open Science Cloud(EOSC)構築に67億ユーロ、うち20億ユーロはホライゾン2020予算、残り47億ユーロは他の公的・民間資金を併用して投資するとしている

34

34

欧州サイエンスクラウド（EOSC）

- 既存の研究データ基盤構築の施策との連携調整：EUDAT、GÉANT、LIBER、OpenAIRE、EGI等
  - 分野的・地理的・施策上別箇に整備されたシステムを結合
  - 欧州全体の研究データ利活用基盤→世界的な共通基盤（“Global Open Science Cloud”）へ？
  - NIIさんのOpen AIRE連携

35

35

これまでの議論から

- **必要なところでしか研究データ共有とオープンサイエンスは進まない**
  - **研究者、研究コミュニティの実装ノウハウの積み重ねが（必要なところから）進む**
  - **義務化は（できれば）最後のひと押しであるべき**

（林、第1回オープンサイエンスデータ推進ワークショップ、京都大学, 2015.9）

36

これまでの議論から

- オープンサイエンスの本質は、実はこれまでの取り組みの中にすでにあることも多い
  - 事業モデルの構築、サービスの再構成の中で、レガシーの様々なしがらみや新しい課題にぶつかり、乗り越える必要がある。

（林、第2回オープンサイエンスデータ推進ワークショップ、京都大学, 2015.12）

37

研究者にとってのオープンサイエンス

- 研究データの「より」オープン化（Data Sharing）
- 研究活動の「より」オープン化（Collaboration）
  - Inter-disciplinary with researchers
  - Trans-disciplinary with citizens
- 新しい研究活動によるゲームチェンジ
  - 研究成果公開の新展開
  - 研究貢献者のより広い認識
  - 研究評価の新展開
- すべては科学技術・学術、産業、文化が発展し、研究者の貢献が認められることが前提

38

38

Data sharing、Open dataの意義の例

①共同解析・・・大量の実験データをグループでは解析しきれない

- ✓ LHC－ATLAS実験：約1000回／秒の陽子・反陽子衝突の即時解析のためには100万台のCPUが必要　⇒　170研究機関を通信ソフトで結び共同解析　＝　ヒッグス粒子を数十事象／110億衝突事象　を発見

②共同観測・・・観測する対象が大きすぎて1機関では扱えない

- ✓ GEOSS（全球地球観測システム）：世界全域を対象とした人工衛星や地上観測など多様な観測システムが連携した、包括的なシステムを構築

③ユニークデータの共有・・・他の研究者は再現できない

- ✓ ヒトゲノム計画：ヒトのゲノムの全塩基配列（約30億塩基対）を解析するプロジェクトの実施

④失敗データの再活用・・・ダメも多く集まれば、成功が見える

- ✓ マテリアル・インフォマティクス：調べられた化合物データの解析により新たな物質を探索　物性を予測して合成が可能な時代へ

⑤ボランティアの活躍・・・趣味が役に立つ

- ✓ Galaxy Zoo：ハッブル望遠鏡がとらえた何百万の銀河系の画像を20万人以上の市民ボランティアが分類・整理　新たな銀河発見にも繋がる

39

39

「より開かれた」科学がもたらす可能性（バイオミメティクスの例）

研究成果（当面は研究データ）の公開が公的資金を活用した研究成果を中心に進展

①分野間協働、共創・・・研究、研究者レベルでのコラボレーションを促す（interdisciplinary)

- エンジニアリングにおけるバイオミメティクスの活用
  - ハコフグの骨格から自動車の骨格を製造、蟻塚の構造を模した通気性の良いビル、アリフェロモンの働きを利用した経路探索アルゴリズム

②セクター間の協働、共創・・・研究に参画するより広い関係者のコラボレーションを促す（transdisciplinary)

- 生物模倣技術の社会実装の進展と標準化

→オープンサイエンスはバイオミメティクス研究が進展するための基盤

40

40

当面使いこなすのは

- 時間データ（歴史）
  - 古文書から大地震研究
- 地理データ（GIS）
  - 比較文化学
- 画像データ
  - 遺跡の解析
- テキストデータ「化」
  - 崩し字

41

41

歴史に習えば

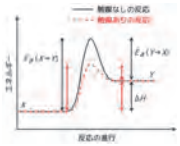
・ポストグーテンベルグの過渡期に居る我々

Print based dissemination

Letter based dissemination

Past Design

Transition State



http://commons.wikimedia.org/wiki/File:Activation_energy_ja.svg

Web native dissemination

Future Design

学術情報流通の再発明

K. Hayashi, "Current States of Impact Assessment of Research Outputs in Japan and Some Challenges to Measure New Impacts for Japan's Stakeholders," OECD-ESTONIA WORKSHOP ON IMPACT ASSESSMENT: PRACTICES, TECHNIQUES AND POLICY CHALLENGES, May 15-16 2014, Estonia. (revised)

42

42

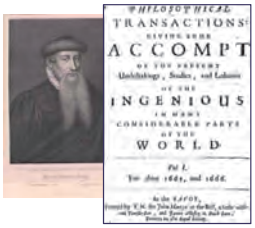
歴史は繰り返す

- グーテンベルグによるある種のオープン革命

手紙、写本
手書きベース
写本

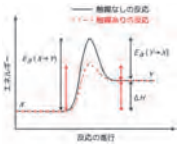
遷移状態

ジャーナル
大量印刷ベース



より
Openな
基盤

Past Design



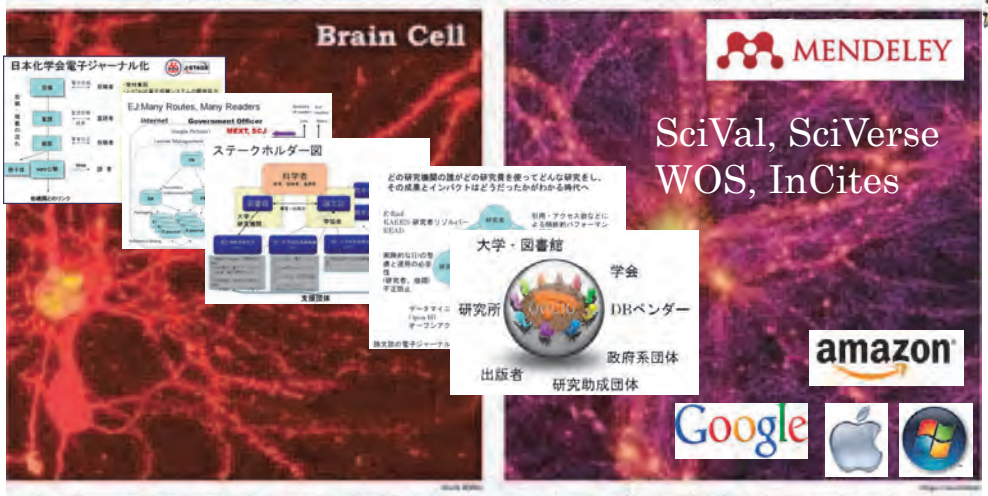
http://commons.wikimedia.org/wiki/File:Activation_energy_ja.svg

Future Design

43

43

オープン＆ネットワーク化



Source by Mark Miller, Brandeis University; Virgo Consortium for Cosmological Supercomputer Simulations; www.visualcomplexity.com.

44

ご清聴ありがとうございました。

45

45

歴史資料のオープンデータ化に関する現在と未来

—歴博の総合資料学の取り組みを通じて—

国立歴史民俗博物館　後藤真

大学共同利用機関法人 人間文化研究機構
国立歴史民俗博物館
National Museum of Japanese History

人間文化研究機構

1

アウトライン

- 歴博が進める「総合資料学の創成」事業
- 館蔵資料情報の「オープン化」の未来
- 「オープンが作る未来」

大学共同利用機関法人 人間文化研究機構
国立歴史民俗博物館
National Museum of Japanese History

人間文化研究機構

2

林先生との話を「つなぐ」

- 人間文化研究機構の論文オープン化

→人間文化研究機構リポジトリ（6機関をJAIRO Cloudで）　「基本的に」自機関の出版物を公開している

- ただし、いわゆる「オープンデータ」になっているものは少ない（ない？）
- あわせて、「評価」の問題が表に出てきている
- 一方で、人文系（を中心とする諸学問）の研究データ取得の前提となるデータをどのように出していくかが課題となっている

大学共同利用機関法人 人間文化研究機構
国立歴史民俗博物館
National Museum of Japanese History

大学共同利用機関法人
人間文化研究機構

3

総合資料学とは何か？

- 博物館資料を一つのディシプリンだけではなく、多様な分野の視点から見るためのモデルを構築する
  - デジタルデータを基盤にし、新たな研究モデルへ
- また、少しずつ異なる歴史資料の目録を、緩く統合する枠組みを構築することで、目録の違いなどでつながらなかったものをつなげていく試みを
  - そこから立ち上がる多様な研究モデルを検討したい

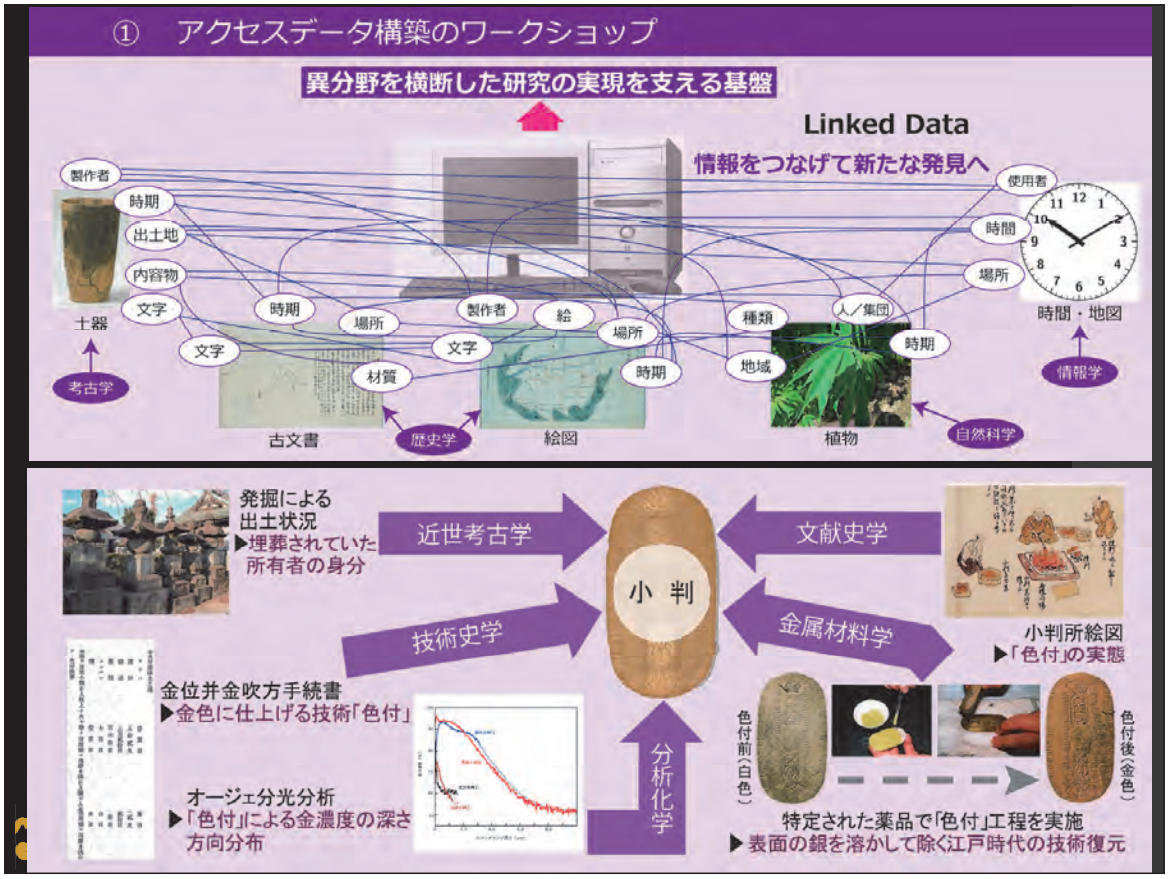
大学共同利用機関法人 人間文化研究機構
国立歴史民俗博物館
National Museum of Japanese History

大学共同利用機関法人
人間文化研究機構

4

総合資料学
研究機関の資料研究の課題を解決する
情報環境・研究モデル・社会還元の総体的研究
連携研究の必要性
歴史学をはじめ
人文系諸学
「危機」
社会的要請への対応
分析データを用いた歴史像
の複眼化
自然科学など
理系諸学
学問データへの
より多様な意義付け
記述的情報によるデータ
の複眼化
多種・多様・大量の
大学・博物館等資料
研究のための自由なアクセス
• 研究と資料情報の循環を実現
• 博物館型オープンサイエンスの構築
• 粒度の細かい地域史・地域文化研究を
実現
連携研究の実現
文理連携研究
チームによる
地域資料の総合的
調査・研究
大学・博物館等と
連携した、
地域の歴史・文化
の再構築
基盤を支える博物館等情報の災害時バックアップ
大学共同利用機関法人 人間文化研究機構
国立歴史民俗博物館
National Museum of Japanese History

5


① アクセスデータ構築のワークショップ
異分野を横断した研究の実現を支える基盤
Linked Data
情報をつなげて新たな発見へ
製作者
時期
出土地
内容物
文字
土器
考古学
時期
場所
文字
材質
古文書
歴史学
製作者
絵
文字
場所
時期
絵図
種類
人／集団
地域
時期
植物
自然科学
使用者
時間
場所
時間・地図
情報学
発掘による
出土状況
▶埋葬されていた
所有者の身分
近世考古学
技術史学
小 判
文献史学
金属材料学
小判所絵図
▶「色付」の実態
金位并金吹方手続書
▶金色に仕上げる技術「色付」
オージェ分光分析
▶「色付」による金濃度の深さ
方向分布
分析化学
色付前（白色）
色付後（金色）
特定された薬品で「色付」工程を実施
▶表面の銀を溶かして除く江戸時代の技術復元

6

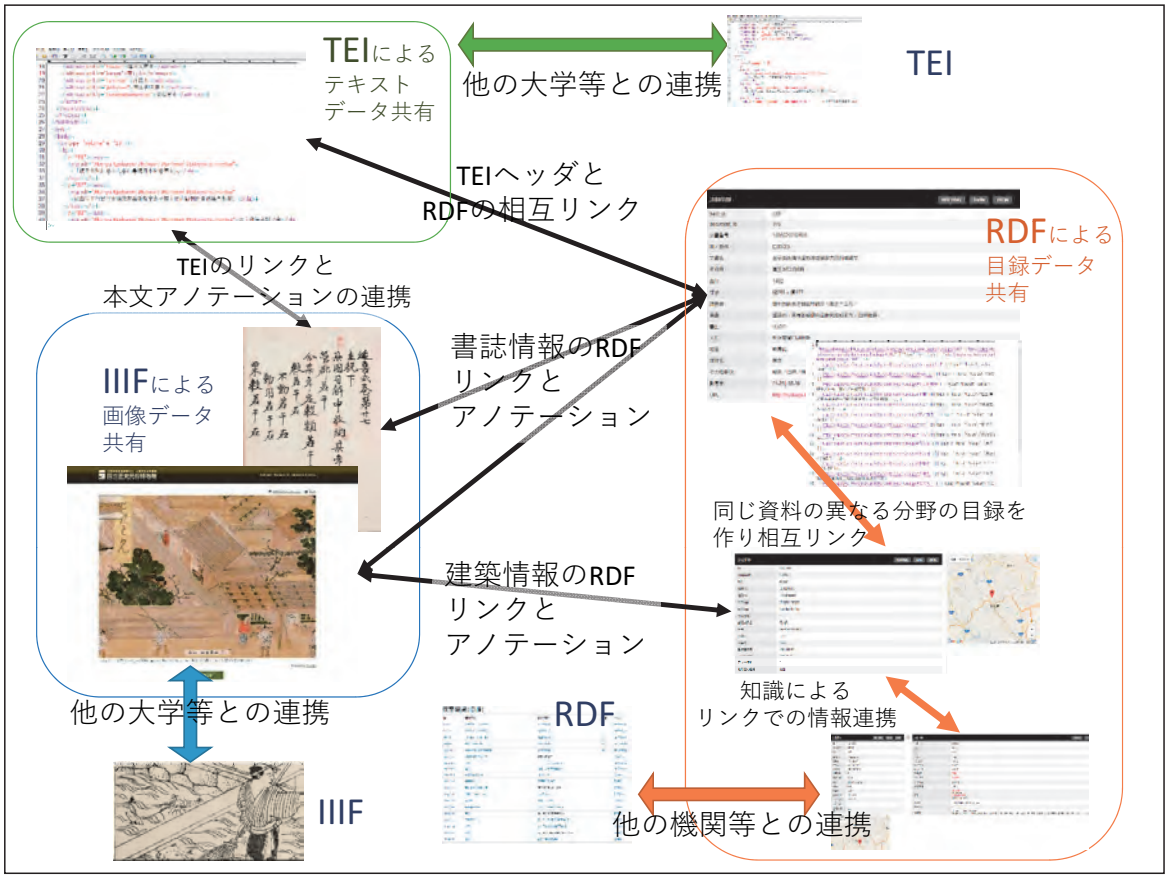
情報基盤の構築

- 特にさまざまなデータを自由に入れられるシステムを構築
- プロトタイプシステムの2系統の構築（RDFとIIIF）
- 総合書物学と共同で、新たな世界標準のテキストデータベース（TEI）構築
- これらのしくみを活用し大学や博物館などの歴史資料のデジタルネットワーク構築

大学共同利用機関法人 人間文化研究機構
国立歴史民俗博物館
National Museum of Japanese History

大学共同利用機関法人
人間文化研究機構

7



8

大学歴史資料情報化モデル

1. 大学が保持する歴史的資料
- 大学博物館・資料館等がもつもの
- 研究の成果物なども（既存のDB含）
- 歴史的なものであれば分野（文理）を問わない

2. 大学による地域歴史資料の保全活動などで把握した資料やデータ
- 歴史資料ネットワーク／（災害時に地域の資料を救出して預かるなど。資料情報はできる限りデータ化）
- 大学の地域貢献から生まれてきた歴史に関するファクトデータ

歴博および提携組織によるデジタル化

A.歴博がデータを持つ（RDF+IIIF）
- A-1:各機関でデジタル化したデータ（CSVなど）を歴博で加工して保持
- A-2:歴博で紙の目録などをデジタル化して、歴博で保持
- 歴博と提携した組織での保持というオプションも
- 保持したデータは（公開可能なものは）歴博のシステム内で大学などの機関単位の検索と横断検索の提供（複雑なUIや機能は作らず、つながる形式で機械可読可能なもので提供）

B.大学でデータを持つ（RDF+IIIFが望ましい）
- B-1:各機関で作ったデータを歴博から横断検索可能に
- B-2:各機関で作ったデータを歴博で加工して、機関に返して保持してもらう

課題
- データの安定的提供
  - DOIなどをいかにつけるか
  - 長期的データ提供
- 多様なデータの提供方式
  - 音声
  - 動画
  - 3Dなども？

9

資料保存機関に使ってもらえるデータの枠組みを考える

- 大きく4つの層を想定
  1. 積極的公開層→客寄せのためにも見てもらいたいコンテンツ　リッチコンテンツとして利用したいもの
  2. 公開層→主に所蔵目録や写真など　展示情報は1と2の中間層
  3. 関係者閲覧層→データ提供者が相互に見られる層　公開層よりはややセンシティブな情報などが含まれてもよい→資料の貸借や展示のための情報共有などの業務層
  4. 秘匿層→センシティブな情報を含むもの（寄託・取得先・取得額・評価額など）　基本的にはWebではなく、別途管理→バックアップのためのデジタル化　ダークアーカイブ

大学共同利用機関法人 人間文化研究機構
国立歴史民俗博物館
National Museum of Japanese History

大学共同利用機関法人 人間文化研究機構

10

館蔵資料情報の「オープン」化

- 本発表時点で歴博に「オープンデータ化」されている資料はない
- 「オープン（≒フリー）アクセス」にできているものは館蔵資料DB・リポジトリなど
- 再利用可能なライセンスを提供しているものは今のところない

→将来にむけては検討中（RDFやIIIFの親和性の高さとともに）

大学共同利用機関法人 人間文化研究機構
国立歴史民俗博物館
National Museum of Japanese History

大学共同利用機関法人
人間文化研究機構

11

オープンアクセス

- 館蔵資料情報にアクセスできること
- 研究成果にアクセスできること
- 展示は…？
- 館蔵資料のDOIモデルの検討など
  - 「資料が引かれている」「論文が引かれている」を可視化するためにも

大学共同利用機関法人 人間文化研究機構
国立歴史民俗博物館
National Museum of Japanese History

大学共同利用機関法人
人間文化研究機構

12

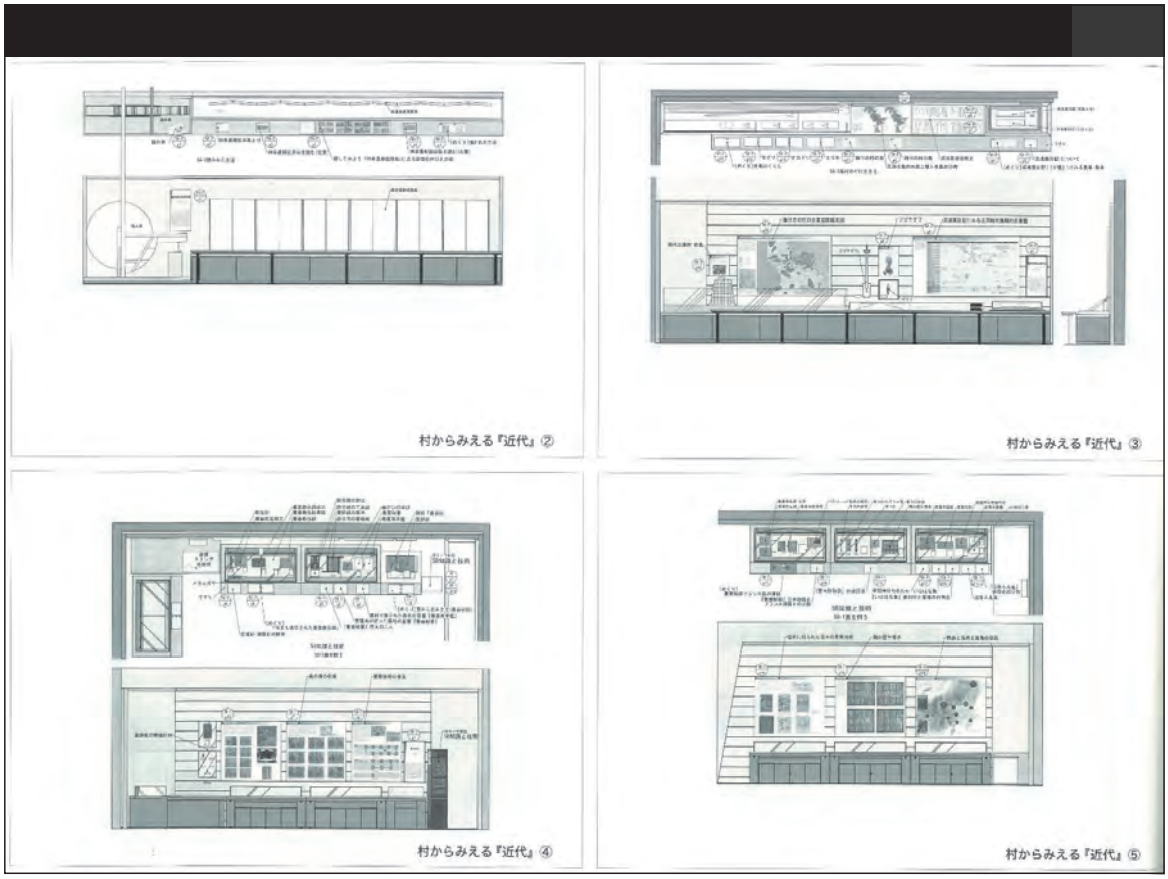
第3展示室(近世)ができるまで

国立歴史民俗博物館
総合展示リニューアルの記録

平成21年3月

大学共同利用機関法人
人間文化研究機構
国立歴史民俗博物館

13



14

オープンデータ

- （個人的意見として）館蔵資料情報から進められるとよい→特に社会の動向として
  - Cf:内閣府知財戦略本部によるガイドライン案（2016年12月現在）
    - 政府標準規約2.0→事実上のCC BY
    - ガイドラインではCC0が求められることも
  - CC0までいくかは課題が残る→利用実績の把握という「評価」の問題との関係
- ＊＊年に資料へのアクセスが××件あることと、オープンデータは運用上背反する部分がある
  →社会の中にあるべき文化資源の情報と、「評価」との関係
  →今後の社会的な意識の組み換えが必要

大学共同利用機関法人 人間文化研究機構
国立歴史民俗博物館
National Museum of Japanese History

大学共同利用機関法人
人間文化研究機構

15

みんなが使うことの意義と課題

- オープンにすることのメリット→国文研の例
- 「使いやすい」システムは外に置く？
  - 機関は最低限の仕組みとデータだけ用意してリッチなものは他機関と分業
  - 一つのかたちではある
- とりわけ人文系機関は「正しい」データの提供に特化する（ただし、機械可読性の確保は重要）戦略も取りうる
- 何を提供するべきかという、人文系機関の使命の再確認が必要に

大学共同利用機関法人 人間文化研究機構
国立歴史民俗博物館
National Museum of Japanese History

大学共同利用機関法人
人間文化研究機構

16

みんなが使うことの意義と課題

- 「みんなで翻刻」がさらにつきつけるもの
  （厳密にはオープンの議論ではないが）
  - 機関の提供するデータの「正しさ」をどこまでの精度で求めていくのか
  - 「誰が」見れば正しいのか　翻って、研究組織が作っているものは本当に正しいのか？

  →なぜ正しいのか　支えている制度は何か？（学会？クロスチェック？）
- 「正しい」を維持できるコストはあるのか？
  - データのオープン化によってますます「データそのもの」に向けられる目が変わるのかもしれない。

大学共同利用機関法人 人間文化研究機構
国立歴史民俗博物館
National Museum of Japanese History

大学共同利用機関法人
人間文化研究機構

17

大学共同利用機関として

※当然仕事がなくなるわけではない

- 中期的には
  - 「知の再配置」→私たちが培ってきた知の蓄積をどのように見せなおすことができるか
    - 囲い込まない
    - 今までの成果をよりひらいていく
    - データをただ出すだけではない→今までの人文的手法の再評価（再評価のためにも囲い込まない）
- 長期的には
  - 資料からの「知」の抽出の多様化→資料学
    - オープンになった資料をより豊かにして提供

大学共同利用機関法人 人間文化研究機構
国立歴史民俗博物館
National Museum of Japanese History

大学共同利用機関法人
人間文化研究機構

18

第12回 人間文化研究資源共有化研究会

「人文科学におけるオープンサイエンスの課題」

歴史的典籍NW事業におけるオープンデータ
―その戦略と課題―

日時：平成29年2月3日（金）15:25 ～15:50（発表20分）

会場：愛知工業大学本山キャンパス

国文学研究資料館古典籍共同研究事業センター

山本　和明

1

研究会テーマの趣旨説明文より

世界の科学分野においては、論文の根拠となるデータをインターネット上で公開・共有化するオープンデータ、オープンサイエンスの動きが急速に始まっています。

データの捏造問題が後を絶たない中、**第三者による検証**がより容易に可能となるよう、**データの共有化を求める動き**が進むことは、必然的なものといえるでしょう。また、情報公開の流れのなかで、行政側が主導して、研究機関を含む公的機関が保有するさまざまな情報のオープンサイエンス化を推進する動きもあります。

他方、人文科学の分野においては、**この問題に対する反応はあまり迅速ではない**ように思われます。

本年度の資源共有化研究会では、国の施策のなかでのオープンサイエンスをめぐる最新の状況についての認識を深めるとともに、**人文科学の分野で**オープンサイエンス化を進めるうえで、どのようなことが課題となるのか、また研究の進展や研究成果の社会化のためには**どのようなオープン化が必要であり、望ましいのか**、といった問題について意見をかわす場としたいと思います。

2

構　成

山本の立ち位置⇒ヘビーユーザーの立場で
（国文学しか分からない旧世代の人間！）

■そもそも「オープンデータ」って？
（人文科学での反応イマイチ問題）

■だれのためのオープンデータ？

□NIJL-NW プロジェクトでの取り組み
ねつ造への取り組み（検証可能性）／戦略的オープン化

□今後に必要なことは？　⇒研究成果発信のオープン

3

総務省　オープンデータ戦略の推進より

http://www.soumu.go.jp/menu_seisaku/ictseisaku/ictriyou/opendata/opendata01.html

「オープンデータ」と言えるための条件

既に各省庁のホームページ上で各種データの公開が進んでいますが、上記のような意義・目的に合致する形での**「オープンデータ」と言えるためには、**

1.機械判読に適したデータ形式で、

2.二次利用が可能な利用ルールで公開されたデータである必要があります。それにより、人手を多くかけずにデータの二次利用が可能となります。

(1) 機械判読に適したデータ形式

コンピュータが自動的にデータを再利用するためには、コンピュータが、当該データの論理的な構造を識別（判読）でき、構造中の値（表の中に入っている数値、テキスト等）が処理できるようになっていることが必要となります。機械判読が容易なデータ形式には、いくつかの段階がありますが、**画像ファイルやPDF等の形式ですと、コンピュータプログラムがその中のデータを識別することは困難となり、二次利用をするためには、人手による再入力が必要となります。**（略）

(2) 二次利用が可能な利用ルール

二次利用が可能な利用ルールについては、第三者がデータを一部改変して利用すること、すなわち**データの二次利用を、データ所有者が予め許諾していることを明示することが必要となります。**例えば、著作物には著作権が発生しますが、二次利用を広く認めるには、その著作権の不行使を予め宣言しておくことが求められます。（略）

4

（一例）　総務省　オープンデータ戦略の推進より
オープンデータの5つの段階とデータ形式

オープンデータは、機械判読の容易性、著作権等の扱いにより、その開放性の程度が異なっています。

これを、便宜的に5段階で示すと下図のようになります。

「オープンデータの5つの段階（出典：★ ）」と、データ形式

段階	公開の状態	データ形式例	参考）Linked Open Data 5star
1段階	オープンライセンスの元、データを公開	PDF、JPG	OL – Open License （計算機により参照できる（可読））
2段階	1段階に加え、コンピュータで処理可能なデータで公開	xls、doc	RE – Readable (Human & Machine) （コンピュータでデータが編集可能）
3段階	2段階に加え、オープンに利用できるフォーマットでデータ公開	XML、CSV	OF – Open Format （アプリケーションに依存しない形式）
4段階	Web標準（RDF等）のフォーマットでデータ公開	RDF、XML	URI – Universal Resource Identifier （リソースのユニーク化、Webリンク）
5段階	4段階が外部連携可能な状態でデータを公開	LoD、RDF スキーマ	LD – Linked Data （データ間の融合情報が規定。検索可能）

人が理解するための公開文書（編集不可）

公開文書（編集可）

機械判読可能な公開データ

オープンデータの5つの段階

出典：★ Open Dataのサイト（http://5stardata.info/）およびTim Berners-Lee氏のLinked Dataに関する提言ページ（http://www.w3.org/DesignIssues/LinkedData.html）を参考に作成。

5

Open Definition（http://opendefinition.org/）の定義する「オープン」

オープンとは?

The Open Definition

The Open Definition sets out principles that define "openness" in relation to data and content.

It makes precise the meaning of "open" in the terms "open data" and "open content" and thereby ensures quality and encourages compatibility between different pools of open material.

It can be summed up in the statement that:

"Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness)."

Put most succinctly:

"**Open data and content** can be freely used, modified, and shared by anyone for any purpose"

オープンデータとは、自由に使えて再利用もでき、かつ誰でも再配布できるようなデータのことだ。従うべき決まりは、せいぜい「作者のクレジットを残す」あるいは「同じ条件で配布する」程度である。　http://opendatahandbook.org/guide/ja/what-is-open-data/#section

「オープン」＝相互運用性を可能とするもの

相互運用性…さまざまなシステムや組織が共同で作業を進められることを意味する。つまりさまざまなデータセットを組み合わせて混ぜて使えるということを表す。

6

オープンコンテント（open content）

https://ja.wikipedia.org/より

オープンソースからの類推によって生まれた概念で、**文章、画像、音楽などの創作物などを共有した状態に置くこと**である。法律的に保護された共有と見ることができ、複製、配布や改変などについての制約がかけられていないことを指したり、またはそのような状態にある創作物を指す。

ある創作物がオープンコンテントであるということだけからは、それが具体的に**どのような利用に対してオープンであるのか**は必ずしも明らかではない。**適用されるライセンス**を確認する必要がある。

オープンコンテントは一般的に、改変するための元データ（ソースデータ）が同時に公開される。これは、結果を単に共有するだけでなく、創造の過程や、更新の過程なども共有される。

ライセンスの種類（例）

CCPL - クリエイティブ・コモンズ

GNU Free Documentation License - GNUフリー文書ライセンスなど

7

現状と課題

■「オープン」といってもわからない人文系研究者が多いという現状

データという言葉への拒否反応

研究者ですらそう、まして一般の人は…

※Web上の画像をそのままキャプチャして引用する研究者

※オープンであることを表明せずに公開している図書館

（研究者に利用してもらえれば…というけれど）

⇒人文系のオープンデータ（オープンコンテント）とは何か、何ができるか、していいかを丁寧に説明することが必要

オープンデータなのに肝心の専門領域の研究者に使われないという「悲劇」

【参考】人間文化研究機構国文学研究資料館データベース利用規程
（定義）第２条（２）「オープンデータ」とは、人間文化研究機構国文学研究資料館の画像データのオープン化に関する指針に基づき、当館データベースにおいて公開する館蔵原本資料及び別に定める資料の画像データをいう。

8

現状と課題

■誰のためのオープンデータか

NIJL-NWプロジェクトの場合
（参考）平成26年の作業部会からの留意事項①

この事業を通じて国文学分野に必ずしも精通していない異分野の研究者等の歴史的典籍の利用を一層促進することは、（中略）国文学研究資料館の評価・発展に大きく資すると思われる。

⇒誰もが使える（時には加工しやすい）形でのオープン化をすることが理想

厳密なオープンより使い勝手の良さを当初目指す

9

構　成（再掲）

□そもそもオープンデータって共通語？
（人文科学での反応イマイチ問題）

□だれのためのオープンデータ？

■NIJL-NW プロジェクトでの取り組み
ねつ造への取り組み（検証可能性）／オープン化の取り組み

■今後に必要なことは？　⇒研究成果発信のオープン

10

NIJL-NWプロジェクトでの取り組み
（公開系システムなど）

■なにをオープンにするか（目標）

当面⇒⇒古典籍の画像（TIFF保管／JPEG配信）と書誌（CSV）、時にはタグ情報、テキスト（⇒TEI）

《課題》

当館所蔵のものだけではない悩み　《交渉》

他館への働きかけが不可欠　そのための国文研蔵オープン

■なにに取り組んでいるか

□ クリエイティブコモンズ表示化

CCライセンスとはインターネット時代のための新しい著作権ルールで、作品を公開する作者が「この条件を守れば私の作品を自由に使って構いません。」という意思表示をするためのツール。

著作権の切れた古典籍⇒検証のためには所蔵先明示が重要

11

NIJL-NWプロジェクトでの取り組み
（公開系システムなど）

□検索結果の安定性：デジタルオブジェクト識別子（Digital Object Identifier＝DOI）等の導入

古典籍画像データのＤＯＩを論文等に引用記載する
⇒論文を読む人は簡単にその論拠となる作品画像にたどり着ける。

国文研は、国際的な識別子であるDOI登録機関（RA）に認定された日本で唯一の機関ジャパンリンクセンター（JaLC）の一般会員（正会員）にすでに入会済
⇒論文引用時にDOI記載を推奨していく。

Japan Link Center　DRIVEN BY doi

12

NIJL-NWプロジェクトでの取り組み
（公開系システムなど）

□画像共有の新しい標準IIIF（トリプル・アイ・エフ、International Image Interoperability Framework）の採用

《これまで》
所蔵資料のデジタル画像が組織ごとに異なる方法で提供されており、共有や再利用が難しかった。

《これから》
画像データやそれに対する注釈情報等のコンピュータ間でのやりとりの仕方を各機関の間で一元化することによって、利用者が色々な使い方を自由に選べるようにする、という状況をもたらす規格がIIIF。

⇒⇒つまり、世界中のどこのデジタルアーカイブの画像ファイルでも、その情報を自分のビューワに読み込ませるとそれを表示してくれるようになる。個別画像を引用することも、他組織が提供する画像との組み合わせといった応用も可能。

※規格外への対応⇒研究開発系でのNII林先生の研究等

13

NIJL-NWプロジェクトでの取り組み
（CODHとの協同によるデータセット公開）

利用者目線のデータセット化

３つのデータセット　リリース

11/10　研究者のためのデータセット
11/17　機械のためのデータセット
［機械学習］
11/24　**市民のためのデータセット**

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

14

NIJL-NWプロジェクトでの取り組み
（歴史的典籍を広く利活用してもらうための仕掛け／CODH）

（１）日本古典籍データセット（全冊画像のOpen Data化）

「古典籍データセット」として追加３５０点（11月10日）合計７００点

【当館所蔵本】のうち

「源氏物語」１８０点　「つれづれ草」３０点

奈良絵本２０点

「書物で見る日本古典文学史」関連古典籍１１０点など

教育現場の利活用も視野に

（２）日本古典籍字形データセットの公開（オープンデータ）

８万字（11月17日）本年度内に合計40万字以上の字形データの公開

（３）江戸料理レシピデータセットの公開（オープンデータ）

⇒クックパッドでの公開も併せて

（４）アイデアソン（Ideathon）の開催　参加型イベント

本年度１２月９日に当館にて開催（じんもんこん２０１６）

すべてメディアリリース実施

15

2017年1月9日朝日新聞

４４兆ＧＢの情報の海に
置かれたオープンデータとは

16

メディアリリースとその反響

17

最後に：今後に必要なことは？（課題）

■だれもが使えるものとしてのオープン化と関連事業（例　「古典」オーロラハンターなどの市民参加型イベント実施）

■コトガラを伝える努力と継続
（異分野融合研究などの取り組みからの学び）

■広報との連動

■データ利用のあり方を募る（アイデアソン）

■伝える努力と継続（コミュニケーション力）

■研究成果のオープン化をどのように果たすか

■画像のCC BY-SA等への表明機関の増加（オープンデータセットへの参加）

18

第12回人間文化研究資源共有化研究会「人文科学におけるオープンサイエンスの課題」
2017.02.03 於愛知工業大学

言語研究と「オープン」データ

前川 喜久雄
人間文化研究機構 国立国語研究所
音声言語研究領域・コーパス開発センター

1

最初に

- 今日は、個人として意見を述べます。国語研教授という肩書での登壇であり、国語研に関係する仕事にも触れますが、研究所としての公式見解を述べるわけではありません
- 科学におけるデータ公開の問題については若い頃から考えるところがあり、その一部については、研究所での研究活動を通じて実践してきました。その個人的体験に基づいた話をします

2

オープンサイエンスとオープンデータ

- 関連したふたつの概念
- オープンサイエンスの方はより政治的色合いが強い？
  - 科学の非制度化、民主化
  - 寡占的出版社からの離脱
  - ネット／クラウドを利用した新しい研究の進め方、新しいインフラ(Science2.0)

- オープンデータはオープンサイエンスの基盤
  - データそのものを評価（査読）の対象に
  - データ専有の阻止（公的資金によるデータ公開の義務化）
  - ひとりでは作れない規模のデータの共同開発と共有
  - 作者が想定していない利用法
  - シミュレーション科学

今日はこちらについて話します

3

「オープン」データからの連想

- **万人に公開されたデータ**
  無償かどうかは二次的な問題
- **公的資金によるデータ公開の義務化**
  データ専有の阻止、研究モラル低下の防止
- **大規模データの共同開発と共有**
  言語研究における言語資源
- **査読の対象としてのデータ**
  Natureなどの投稿規定
- **シミュレーション科学**
  天文学、化学、今後はおそらく言語学も

今日はこの辺りについて話します

4

オープンデータの必要性を感じた三つの契機

1. 音声学者のいらだち
2. 調査研究のモラル低下
3. モラル低下への反省：言語資源の整備

5

オープンデータの必要性を感じた三つの契機：その１

音声学者のいらだち

音声学は音をあつかう

音は印刷できないので論文では音声記号を使う

音声記号の正しさを保証するものは何か

例：「出雲方言のラ行子音は有声そり舌音である」
「大分県臼杵市方言には前舌円唇母音がある」

⇒　論文の中にはなにもない、研究者の評判を信頼しているだけ

⇒　音声資料を提供するのが根本的な解決策

前川「方言音声分析の問題点」日本語学, 8 (3), pp.36-46, 1989.

6

オープンデータの必要性を感じた三つの契機：その１

同じ問題は言語分析のさまざまな局面にあらわれる

研究者が自分で書いたプログラムの挙動の正しさ

日本語テキストの「単語」へのセグメンテーション

例「国立国語研究所」は何語？

「英語」「フランス語」「中国語」は？

⇒ 研究者は誠実にベストを尽くしているのだろうが、過去の研究を再現しようとすると疑問が噴出する

⇒ その結果、すべてを一からやり直すのが安全ということになりかねない

⇒ やはりプログラムやマニュアルを公開して批判してもらうことが最良の解決策

7

オープンデータの必要性を感じた三つの契機：その１のまとめ

- こうした問題は、かつては技術上の問題のために解決困難だった
- しかし現在では技術上の問題はほぼ完全に解消されている
  - ⇒ きわめて安価なストレージ
  - ⇒ ネット環境、クラウド環境の全地球的普及
- それにも関わらず問題が解決されていないとしたら、研究者コミュニティは怠慢の誹りをまぬがれない
  - ⇒ 自然科学系の雑誌では対応済みのものが多い (Nature, 医学系諸雑誌)
  - ⇒ 人文科学を例外とする合理的な根拠は私には思いあたらない

8

オープンデータの必要性を感じた三つの契機：その２

国立国語研究所における調査研究のモラル低下

- 研究所の創立(1948年）から10〜15年程は計画通りに完了した調査が多い
- やがて報告書の遅れが目立つ調査研究が現れはじめる
- 1980年代には遅れが常態化する
- ついには報告書が刊行されない調査がでてくる

9

オープンデータの必要性を感じた三つの契機：その２

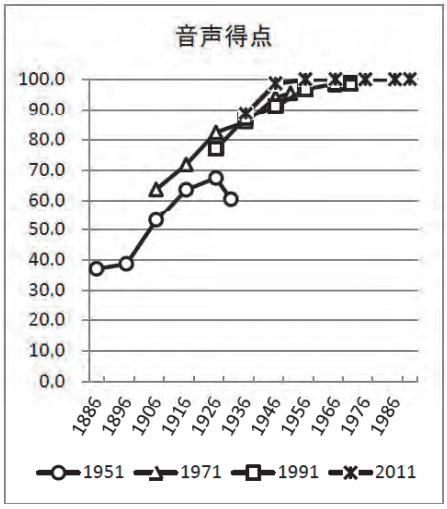
鶴岡調査：無作為抽出された市民の面接調査で共通語化の過程を実証的に分析。20年間隔で４回実施。調査開始から報告書の刊行までの年度数を比較

- **第1次調査**：1950年実施。1952年度末に報告書を刊行。3年で完了
- **第2次調査**：1971-72年に科研費で調査。1973年度末に報告書を刊行。3年で完了
- **第3次調査**：1992-93年度に科研費で調査。1994年度末に「記述調査」（サテライト調査）の報告書を刊行。2007年と2008年に「内部資料」を刊行。その後は報告書なし。16年かけて完了したかどうか不明
- **第4次調査**：統計数理研究所主体で2011年に調査実施。2014-15両年にレポート2本（PDF)を公開。5年で完了

10

調査年別・生年コウホート別共通語化率

生年コウホート	1951	1971	1991	2011
1882-1891	37.0			
1892-1901	38.7			
1902-1911	53.1	63.7		
1912-1921	63.6	71.8		
1922-1931	67.4	82.3	77.0	
1932-1941	60.4	85.8	87.1	88.7
1942-1951		94.1	91.1	98.6
1952-1961		95.4	96.8	99.9
1962-1971			98.5	100.0
1972-1981			98.6	99.9
1982-1991				99.9
1992-1996				100.0



鶴
2
• 第
• 第
3
• 第
刊
• 第
ポート2本（PDF)を公開。5年で完了

11

オープンデータの必要性を感じた三つの契機：その２

北海道調査：社会調査の手法で北海道における共通語の成立過程を分析

- **第1次調査：**1958～60年度に科研費で実施。1965年3月に報告書を刊行。7年で完了
- **第2次調査：**1986-88年度に科研費で実施。11年後の1997年3月に内部資料を作成するも、正式の報告書は今日まで公開されず

12

オープンデータの必要性を感じた三つの契機：その２

国民各層の言語生活の実態調査：1962-63年度に長岡市と松江市で、留め置き調査、面接調査、などからなる数千名規模の社会調査を実施。あわせて試験的な24時間調査（1名）も実施

- 昭和39年度(1964年度)国語研「年報」に松江調査について「(1)基礎調査と、(2)市民調査を中心とした報告書を、40年度(1965年度）じゅうに作成する予定である」(p.90)との記述があるが、結局、刊行されず。９年後の1971年3月に『待遇表現の実態―松江24時間』だけが刊行された

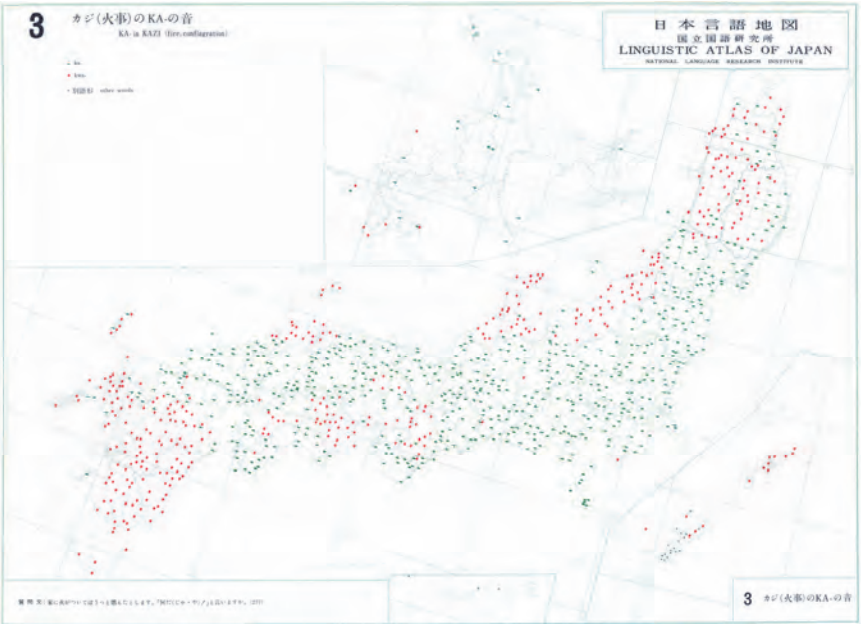
13

オープンデータの必要性を感じた三つの契機：その２

言語地図：語彙を調査した『日本言語地図』と文法・表現法を調査した『方言文法全国地図』がある

- **『日本言語地図』**（2400地点×285項目）
  1957～74年度　18年で完了
- **『方言文法全国地図』**（807地点×267項目）
  1977～2006年度　30年かけて完了

14



言語
査
法を調

- [
- [

15

オープンデータの必要性を感じた三つの契機：その２のまとめ

モラル低下の原因（いろいろあるだろうが）

- 反復調査は後になるほど新規性のある成果を出しにくい

  追加された調査項目もあるにはある

- 適正規模を超えた研究所の活動拡張

  予算を増やすために、むやみにプロジェクトを立てる
  頭数は増えないから、１プロジェクトに割ける人的コストが低下

- そもそも成果の刊行によって調査を完結させることの重要性の認識不足

  公的資金という意識の欠如
  言語研究に限った話ではないことを最近知った

16

オープンデータの必要性を感じた三つの契機：その２のまとめ

モラル低下の原因（いろいろあるだろうが）

- 反復調査は後になるほど新規性のある成果を出しにくい
  追加された調査項目も
- 適正規模を超えた研
  予算を増やすために、
  頭数は増えないから、
- そもそも成果の刊行
  公的資金という意識の
  言語研究に限った話で

ゆえに発掘ありて記録なく、記録ありてこれが刊行を怠るは、畢竟公的資料を破壊し、これを私蔵するものというべく、（中略）発掘の報告出版は、発掘事業の一部分にして、決して分離すべきにあらず、その費用と時間とは始めより見積もり置かざるべからず。これ実に考古学者の道徳的義務として、厳粛に服従せざるべからざるところなりとす。

濱田耕作『通論考古学』(1922), 岩波文庫版, p.217

17

オープンデータの必要性を感じた三つの契機：その３

モラル低下への反省：1990年代後半

⇒ さすがにちょっとまずいんじゃないかと感じ出した（人もいた）

⇒ 研究員が長期間データを抱え込んで結局凡庸な報告書を出すよりも、調査終了直後にデータを公開して多くの人に利用してもらった方が全体としては良い結果になるんじゃないか

⇒ **言語資源**(language resources)の開発

⇒ 上記の構想を試す機会が訪れたのが、科学技術振興調整費による「話し言葉工学」プロジェクト（国語研、通信総研、東工大の共同研究）。そこで国語研は『日本語話し言葉コーパス』の構築を分担した

18

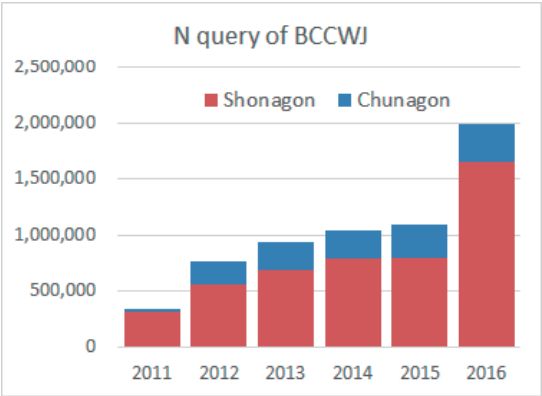
オープンデータの必要性を感じた三つの契機：その３

言語資源の開発と一般公開（私が責任者であったもの）

- **『日本語話し言葉コーパス』(CSJ)** 音声認識用自発音声コーパス, 750万語（科学技術振興調整費補助金 1999-2003年度), コーパス公開2004年5月, 報告書刊行2007年3月
- **『現代日本語書き言葉均衡コーパス』(BCCWJ)**日本語初の均衡コーパス, 1億語(科学研究費特定領域研究 2006-2010年度), コーパス公開2011年12月, 朝倉書店より講座「日本語コーパス」刊行中（8冊中6冊既刊）
- **『国語研日本語ウェブコーパス』(NWJC)**ウェブの日本語を母集団とした253億語のコーパス(運営費交付金概算要求経費 2011-2015年度), コーパス試験公開2016年8月, 一般公開2017年3月予定

19

『現代日本語書き言葉均衡コーパス』のオンライン検索数



ふたつのウェブアプリ『少納言』『中納言』で、いずれも無償公開。両者をあわせると2014年度に年間100万検索を突破。2016年度には200万回を超した

20

20

DVD版の累計頒布数

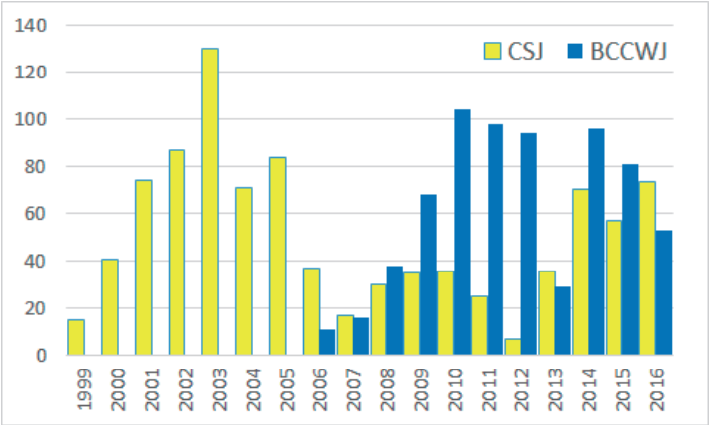
コーパス	アカデミック＋一般	商業利用	計
『日本語話し言葉コーパス』2004年5月公開	818	12	830
『現代日本語書き言葉均衡コーパス』2010年12月公開	307	20	327
		総計	1157

DVD版では形態論情報だけでなく、あらゆるデータ（音声、テキスト、書誌情報、種々のアノテーション）を利用可能
2016年3月時点の集計

21

21

論文類での被参照件数



Google Scholarのデータを集計（2016は11月段階の数字）
総計で『日本語話し言葉コーパス』（CSJ)は925件、
『現代日本語書き言葉均衡コーパス』(BCCWJ)は688件

22

22

上位5件が言語資源関係で合計870件

23

契機１～３についてのまとめ

- 言語研究は、もともとデータ公開によって補完されるべき性格のものである
- データを含め、成果公開を要請されない研究は堕落しやすい
- データ公開を主目的とする言語資源整備事業が言語研究に貢献している

24

残る問題 その１: 評価

データの作成と公開に対する評価

- 業績として認めない人がいる

  実際にコーパスを利用して論文を書いていても、本文か謝辞に書くだけで参考文献では引用しない研究者がめずらしくない

- 100本の論文で利用されたデータと著名な雑誌に掲載されたが１回も引用が確認されていない論文の価値をどう比較するのか？

  学会などできちんと検討してほしい
  そもそも従来の業績評価法が不完全

25

残る問題 その２: 公開の義務化

税金を使った研究ではデータ公開を義務化すべきか

- 倫理的には義務

  特に面接調査、アンケート調査、実験など市民に関与をもとめた研究や多額の研究費を費やした研究では強い義務

- ただし研究期間終了時にデータが公開できるとはかぎらない

  論文が採択されていない、バグとりが終わっていない等

- またデータ公開の有無を個人の業績評価と直結させることは危険

  そうすると質の悪いデータを量産して業績を稼ぐ人が必ず出てくる

  業績評価に利用するならば、長期的に引用件数なども含めた評価が必要

26

残る問題 その３: 知財関係の法整備

著作権処理が大規模データのすみやかな公開を妨げる

- 現代の著作権法は万人が権利者になるデジタル時代に適応できていない
- オプトアウト制度の法制化が望まれる
- 著作権保護期間も長すぎる

  長期の保護を必要とする人は権利物を登録するような制度が必要

  Maria Pallante（米国著作権局長）の *The Next Great Copyright Act*.

前川「コーパス構築と著作権保護」人工知能学会誌, 25(5), pp.628-632, 2010:09.

27

人間文化研究情報資源共有化研究会報告集　7

Proceedings of the Study on Information Resources of the Human Science Vol.7

平成29年３月29日発行

編集・発行　大学共同利用機関法人　人間文化研究機構
総合情報発信センター
高度連携情報技術委員会
〒105-0001　東京都港区虎ノ門４－３－13
ヒューリック神谷町ビル2F
Tel　03－6402－9200
URL：http://www.nihu.jp/

ISBN 978－4－903211－20－6

リサイクル適性Ⓐ
この印刷物は、印刷用の紙へ
リサイクルできます。