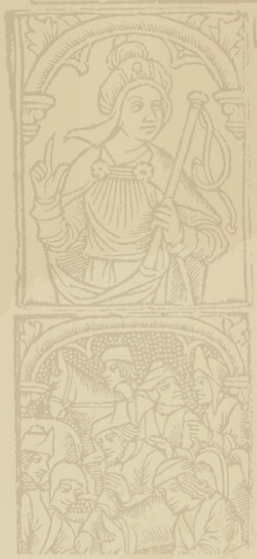
ISSN 2188-3424



慶應義塾大学DMC紀要
第 6 号

DMC Review

Keio University, Vol.6, No.1

目　次

巻頭言 …… 4
重野 寛（慶應義塾大学 DMC 研究センター所長／理工学部教授）

特集　DMC 研究センターシンポジウム　第 8 回
「デジタル知の文化的普及と深化に向けて：メタデータ再考」

講演

「学術データの共有と利活用のための工夫」 …… 6
原 正一郎（京都大学東南アジア地域研究研究所教授）

「コンテンツネットワーク形成におけるメタデータの限界」 …… 29
金子 晋丈（慶應義塾大学理工学部専任講師／DMC 研究センター研究員）

話題提供

「書誌学者の立場から見たボーダレスなデータ利活用のための情報組織化」 ‥ 47
安形 麻理（慶應義塾大学文学部准教授）
「パースペクティヴとメタデータ」 …… 53
久保 仁志（慶應義塾大学アート・センター所員）
「「MoSaIC」におけるデジタルコンテンツの情報組織化」 …… 59
石川 尋代（慶應義塾大学 DMC 研究センター特任講師）

パネルディスカッション

「ボーダレスなデータ利活用のための情報組織化とは」 …… 63
（パネリスト）
原 正一郎（京都大学東南アジア地域研究研究所教授）
金子 晋丈（慶應義塾大学理工学部専任講師／DMC 研究センター研究員）
安形 麻理（慶應義塾大学文学部准教授）
久保 仁志（慶應義塾大学アート・センター所員）
石川 尋代（慶應義塾大学 DMC 研究センター特任講師）
（モデレーター）
安藤 広道（慶應義塾大学文学部教授／DMC 研究センター副所長）

研究ノート

「インターネット前提時代におけるパブリック・ヒューマニティーズへの挑戦」 …… 82
宮北 剛己（慶應義塾大学 DMC 研究センター研究員）

記録 …… 93

編集後記 …… 102
安藤 広道（慶應義塾大学 DMC 研究センター副所長／文学部教授）

巻頭言

重野　寛
慶應義塾大学 DMC 研究センター所長
理工学部教授

『慶應義塾大学 DMC 紀要』第 6 号をお届けいたします。本号には、2018 年秋の DMC 研究センターシンポジウム第 8 回「デジタル知の文化的普及と深化に向けて：メタデータ再考」における講演やパネル・ディスカッションをはじめとして、この 1 年間の活動報告、所員の研究成果などが掲載されています。

当センターは 2004 年にデジタルメディア・コンテンツ統合研究機構としてスタートし、2010 年にその活動を現在の研究センターに引き継ぎました。以来、文理にとらわれない学問分野の融合を目指し、新しい知の創造や流通、あるいは教育、文化、芸術等のさまざまな領域への貢献を考えながら、活動を続けてまいりました。

当センターでは、幅広い研究プロジェクトを展開しています。そのひとつにデジタル・アナログ融合型のユニバーシティ・ミュージアムの実現を目指す Museum of Shared and Interactive Cataloguing（MoSaIC）プロジェクトがあります。MoSaIC プロジェクトでは、文化資源の周辺や背後にある「物の関係性」をコンテクストと呼び、このコンセプトを進化させる方向で、デジタルミュージアムの技術開発と実践を重ねてきました。一方で、文化資源のデジタル情報の活用においてはメタデータの利活用の重要性が認識されています。メタデータのあり方は我々の考える文化資源のコンテクストと常に深い関係にあります。今回のシンポジムは、デジタル知のボーダレスな利活用をさらに促進する観点から、メタデータについて再度考えるとともに、私どもの立ち位置を確認する意味でも大変貴重な機会となりました。

当センターが推進するもうひとつのプロジェクトとして、FutureLearn におけるオンライン講義配信があります。FutureLearn は英国ロンドンに本部を置く、オンライン講義の配信事業体で、これに慶應義塾として参画しています。2018 年 8 月の段階で、慶應義塾大学からは 6 本の講義を配信しており、学習の登録者は全体ですでに 5 万人を数えるというところまで成長し、まだまだ増えていくと見込まれています。

『DMC 紀要』第 6 号を通じ、より多くの研究者、関係者の皆様へ当センターの活動をお伝えし、デジタル知の利活用やそれらを取り巻く諸問題についてご検討を進める一助となりますと幸いです。

特集

DMC 研究センターシンポジウム　第８回「デジタル知の文化的普及と深化に向けて」

メタデータ再考

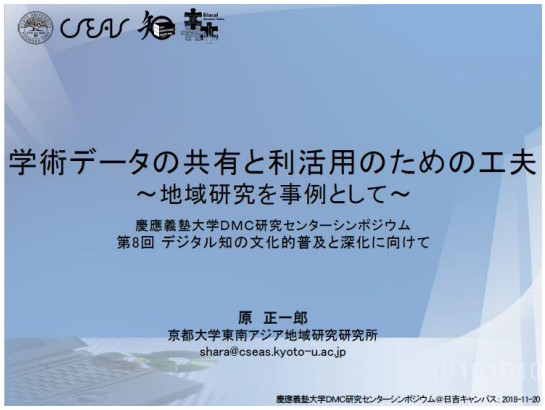
慶應義塾大学デジタルメディア・コンテンツ統合研究センター（DMC 研究センター）では、2018 年 11 月 20 日（火）に DMC 研究センターシンポジウム第８回「デジタル知の文化的普及と深化に向けて：メタデータ再考」を開催した。本シンポジウムの目的は、文化資源の利活用のボーダーを創造的に越えるために、メタデータを取り巻くフレームワークを改めて問い直すことであった。このシンポジウムより講演、パネルディスカッションを採録する。

講演

「学術データの共有と利活用のための工夫」

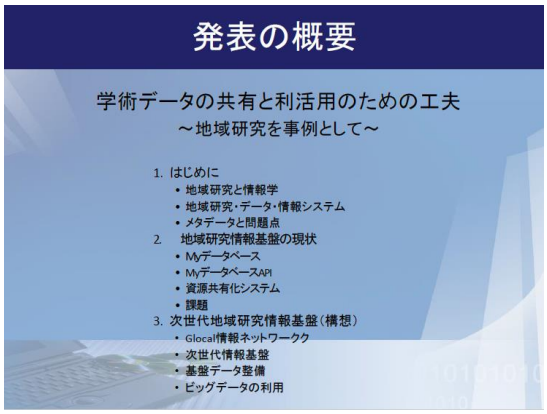
原 正一郎

（京都大学東南アジア地域研究研究所教授）



ご紹介にあずかりました、京都大学東南アジア地域研究研究所の原と申します。この発表では、東南地域研と呼ばせていただきます。

本日お話しする内容は、このようになっております。

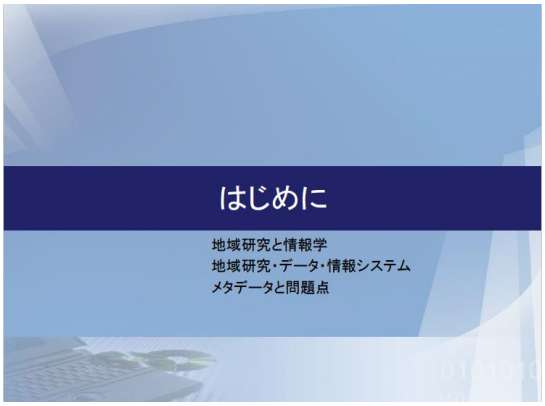


最初に、地域研究とは何か、地域研究を支援する情報システムとはどのようなものか、そこで考慮しなければならない事柄についてメタデータを中心にお話をします。この発表の基礎となる事柄です。

続いて、東南地域研が公開している 3 つの情報ツールについて、具体的に説明します。この発表の中心です。

最後に、オープンデータやビッグデータに対応した、新しい情報プラットフォームの開発についてのお話をさせていただきたいと思っております。

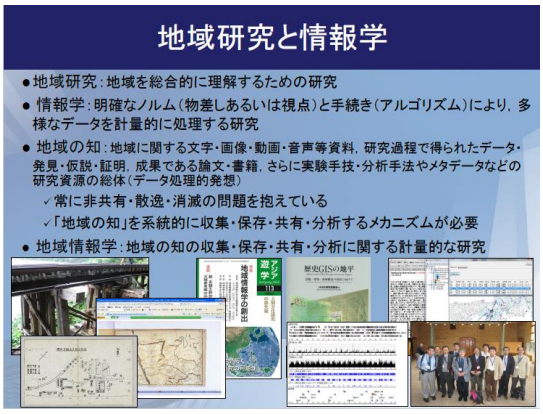
と言いましても、規模の小さい人文系の研究所の、その中の小さな情報センターの話なので、それほど大がかりなことはできません。ですが、手を広げ過ぎていて、どれもこれもが中途半端な状態になっております。



まず、「地域研究」や「報学」や「知」について、整理しておこうと思います。

言うまでもないことですが、地域研究は複合領域です。実際、東南地域研の構成員は、歴史・文化人類・経済・医学・農学・生物などを専門とする研究者です。ですから、地域研究の定義は研究者ごとに違って

います。ここに示したのは、あくまでも私の定義ですが、「地域を総合的に理解する」研究領域としています。もちろん、「地域」とは何処かあるいは何か、「総合的に」とはどのようなことか、どうしたら「理解」したと言えるのかなど、曖昧この上ないのですが、研究領域を厳密に定義する意義があるとは思えないので、ざっくばらんにこの程度に考えています。



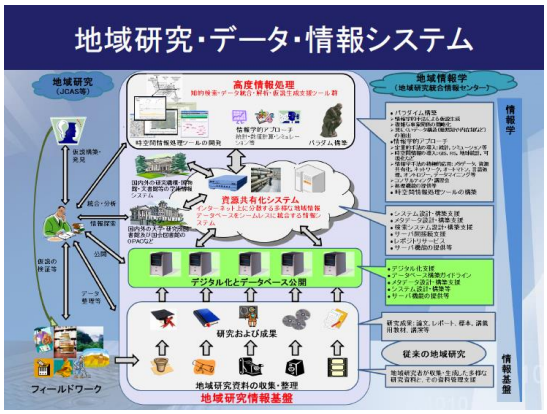
一方、情報学ですが、「明確なノルム（物差しあるいは視点）と手続き（アルゴリズム）により、多様なデータを計量的に処理する研究」と定義しています。これも、なかなか突っ込みがいのあるところですが、地域研究と同じく、情報とはなにかを突き詰めていくと分からなくなってしまうので、ステレオタイプ的ですが、人文社会科学の定性的アプローチに対比するという意味で、計量性と再現性を強調しています。つまり、同じデータと同じ方法を使えば、誰でも同じ結論を再現できる、と言うことです。もちろん、情報学で扱うデータが全て計量的とは言いませんが。

地域の知ですが、これはもっと曖昧です。ここでは「地域に関する文字・画像・動画・音声などの資料、研究過程で得られたデータ・発見・仮説・証明、成果である論文・書籍、さらに実験手技・分析手法やメタデータなどの研究資源の総体」を知としています。データ処理的発想とでもいいましょうか。コンピュータを前提とすれば研究資源はデジタルデータとなりますが、ここにはアナログデータが入っても良いかなと考えています。

最後に、地域情報学ですが、地域研究に関する知を組織化して利活用を支援する取り組みと考えております。たとえば、左下の写真は「戦場にかける橋」の場所ですが、このようなフィールドに足を運んだり、その右の写真は地図ですがさまざまな資料を集めたり、それらをデータベース化したり、分析するためのツールを作ったり、右側の写真のように成果を国際学会で発表したりする活動をしています。

これは、地域研究を支援する地域情報学

を説明するために、いつも使っている図です。地域の知を情報の流れとして、東南地域研で開発した情報ツールと必要な情報技術を整理したものです。



地域研究ですから、全てはフィールドワークから始まります。都市や村落や森林などいろいろなフィールドに滞在して、観察したり、インタビューしたり、資料を収集したりします。それらの研究資源を研究室に持ち帰って、整理・分析して、論文や書籍を執筆します。ここで研究は終了しますが、私はこれを「従来型の地域研究」と呼んでおります。

一方で、収集した研究資源、例えば写真や動画などは、再現のきかない、その時・その場所でしか収集できない資料なので、その意味で貴重です。さらに、地域で集めた資料や研究成果は地域や一般に還元すべき、広く学術コミュニティーで共有すべき、あるいは執筆した論文や書籍の証拠として保存すべきなどといった、社会的・学術的な要請が強くなっています。そのため、研究資源のデジタル化とデータベース公開は、地域研究においても必要不可欠な要件となってきました。これが、東南地域研でもデータベースの構築と公開を積極的に進めている理由で、そのために開発したものが「地域研究情報基盤」です。

詳細は後にしますが、「地域研究情報基盤」において、研究資源のデジタル化と公開の機能を担う情報ツールが、Ｍｙデータベースです。東南地域研のほとんどのデータベースは、Ｍｙデータベースにより構築されています。

ところが、データベースの数が増えてくると、どの資料がどのデータベースに入っているか分からない、一つずつデータベースを検索するのは面倒だということで、データベースを統合する仕掛けが必要になってきました。そのために開発した情報ツールが「資源共有化システム」です。これも後で詳しく説明します。

さらに、統合されたデータの高度情報処理を進めるために、いくつかの分析ツールを試作しています。

このようにして得られた研究成果は、地域にフィードバックされます。これで地域の知が一巡します。この地域の知の循環のプロセスを支援するＭｙデータベース、資源共有化システム、分析ツールなどの情報ツールの総体を地域研究情報基盤、その開発に関わる情報学を地域情報学と呼んでい

ます。こういうとカッコは良いのですが、どれも完成の域には達しませんし、新しい技術が次から次へと現れてくるので、どの技術を導入するか、どの段階で古い技術を入れ替えるかなど、戸惑いが多いのが実際のところです。

さてデータベースを構築するうえでメタデータは欠かせません。このメタデータですが、「あるデータにアクセスするための要約というか Proxy となるデータ」と言われています。自分なりに翻訳すると、ある対象物や現象の特徴を何らかの視点あるいは基準で計測して定型的に要約したデータとします。たとえば、人物が対象なら、生年月日や住所や地位などのデータの集まりになりますし、同じ人物でも健康が対象であれば、血糖値やコレステロール値などのデータの集まりになります。つまり物理的には同じ対象物や現象であっても、視点や切り口が違えば、違うメタデータとなります。

メタデータについて、もう少し掘り下げたいと思います。筑波大学の杉本先生の説明法が面白いので、それを拝借します。

さて、お茶の自動販売機のディスプレイが左側のボルトの写真のようだとします。ここでボトルの中身をデータ、自動販売機をデータベースと考えます。このお茶は何でしょうか？　私は「おーいお茶　濃い茶」が欲しいのですが、右側のようなラベルがなければ正しいボトルを選ぶことができません。メタデータとは、このラベルのように中身について記述した事柄です。

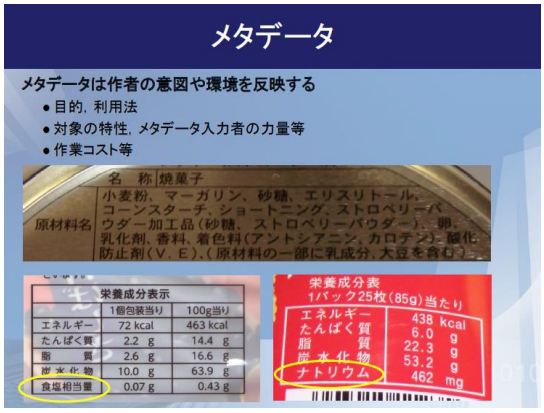
つまり、ラベル「おーいお茶　濃い茶」がなければ、ボトル「おーいお茶　濃い茶」を選ぶことができない。言い換えると、ラベルというメタデータがなければ、正しいデータにアクセルでないことになります。メタデータの書き方は、今のところは問わないけれど、これがなければデータベース検索はできませんということです。

このメタデータですが、作者の意図や環境を反映したものとなります。先ほどの人物データベースと健康データベースは、対象物は同じヒトであっても、目的によってメタデータの内容が異なるという例です。この視点は、学術データベースにおいて重要です。後ほど説明いたします。

それから、あまりいい言い方ではないですが、メタデータの質は作成者の力量に左右されます。どんなに精緻なメタデータを設計しても、必要な情報を読み取る目がなければ、正しいメタデータを作成できないからです。たとえば、医療データを検索す

るためのメタデータがあったとしても、必要な語彙や医学知識あるいは症状を正しく認識できる技術などがなければ、使えるメタデータを作成することはできません。

それから、作業コストにもメタデータ作成において重要な要素です。たとえば完璧なメタデータを 1 件作るのに 1 万円かかるとします。この場合、メタデータの質は高いもののデータ量の少ないデータベースで我慢するか、メタデータの質は多少劣るけれどもデータ量の多いデータベースとするかなど、悩ましい決断が必要になります。



もう少しメタデータの中身を見てみましょう。先ほどの、「作者の意図を反映する」ことに関係しています。これらの食品表示ラベルもメタデータの例です。このメタデータを例にした理由ですが、私にとっては死活問題だからです。私は大きな病気をしまして、食塩の摂取量が 1 日 6 グラムに制限されています。食塩 6 グラムはどのぐらいか、カップラーメン 1 個で 5 グラムといったら、一食あたりの食塩量がどれだけ少ないか想像つくと思います。なので、スーパーなどで食材を買う時に、「食塩相当量」や「ナトリウム」は、私にとってはとても重要なデータなのです。ですから、下の 2 つの栄養成分表は私の意図に合ったメタデータですが、上側の食品表示ラベル役に立たないメタデータということになります。

次に「メタデータの書き方は、今のところは問わないが」に関係しますが、下の 2 つの栄養成分表を改めてご覧下さい。これらは「内容的」には同じですが、記述法が違っています。まず左側は一包あたりあるいは 100g あたりの値で、右側は 85 g あたりの値です。また左側は食塩相当量ですが、右側はナトリルウム量です。メタデータの記述法が多様だということがお分かりと思います。

実際のデータベースを例にします。この上下 2 つの仮想データベースの内容は同じですが、項目名、データ型、記述規則、粒度が異なっています。

まずデータ項目ですが、上が日本語であるのに、下は英語となっています。これが同じであることはヒトにとっては当然ですが、コンピュータにとっては別物です。

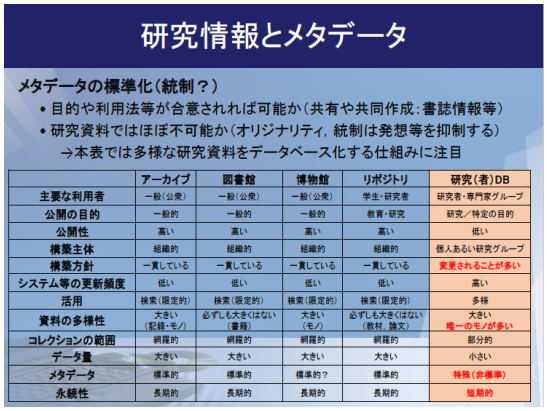
データ型について、上の識別では「0001」ですが、下の ID では「1」です。「イチ」ということでは同じかもしれませんが、「0001」は文字型、「1」はおそらく整数型で、違うデータ型です。

上の性別の「M」と下の SEX の「1」は記述規則の違いです。ちなみに、JIS（JIS X0303（性別コード））では男は「1」、女は「2」となっています。上の誕生日の「昭和 32 年 10 月 11 日」と下の birthday の「1957/10/11」も記述規則の違いです。

上の氏名は「大学　太郎」で、下は surname が「ダイガク」で forename が「タロウ」で、粒度が異なっています。これらのように、内容的の同じメタデータでも、記述法は違っていることが分かります。

では、異なるメタデータをどうしたら良いでしょうか。一つの手段はメタデータの統制あるいはメタデータの標準化です。みんなで同じメタデータを使いましょう、ということです。データサービス機関の間で標準についての合意が形成できるなら、良い解決法だと思います。図書館の機械可読目録（MARC :MAchine-Readable Cataloging）はその典型例です。日本では、国立国会図書館の JAPAN MARC や国立情報学研究所の NACSIS-CAT などがこれに該当します。

一方、研究分野におけるメタデータの統制あるいは標準化は、必ずしも容易ではないし、良い方法とも思えません。なぜかというと、分野が違えば語彙が異なります。あるいは、同じ語彙であっても意味が違うこともあります。それから、研究ではオリジナリティーを重視しますから、対象物が同じであっても、メタデータが異なるのは当たり前です。標準化や統制は、発想の自由を阻害します。新しい発想のもとで作られるメタデータは、それまでとは全く違う構造となる可能性があります。そのような理由で、研究分野のメタデータの統制や標準化は不可能であり有用でもないと考えています。

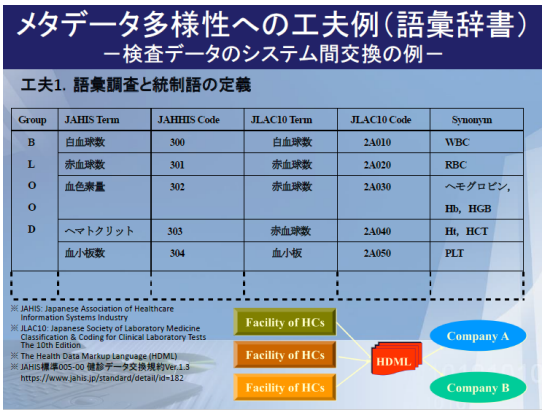


ここまで述べたことを表にまとめてみました。図書館のメタデータは標準化が進んでいます。アーカイブ・博物館・レポジトリについても、標準規約の制定は進みつつありますが､導入は進んでいないようです。これらのいわゆる機関データベースに比べると、「研究（者）DB」の様相はかなり異なっています。主要な利用者は研究者あるいは研究グループなどの個人あるいは小規模なグループです。そのためもあり、公開性は必ずしも高くはありません。個人データなどが含まれている場合は、もちろん公開できません。目的も研究の遂行に特化していますから、メタデータの構造はデータベースごとに異なっていて、これをメタデータの heterogeneity と呼ぶことがあります。

データ収集法やデータ構造や利用法についても、図書館や博物館は一貫していますが、研究者は変化します。今日、私はデータサービス側の立場で話をしているので、メタデータやアプリケーションの変更はありがたくないのですが、反対にデータを利用する研究者側の立場のとき、朝令暮改はよくあることです。

このような多様なデータをどうやって管理・運営するかというのが、これまでの課題でした。ついでですが、もう一つ、今、直面してる問題は、多様なデータの長期的な保存と利活用をどのように実現するかということです。

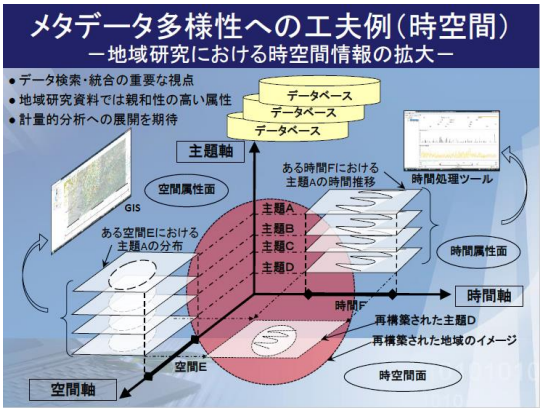
研究メタデータの特徴を明らかにしたところで、異なるメタデータを統合するあるいは繋ぐにはどうしたら良いかという話に進みたいと思います。



皆さんが健康診断を受けると、血糖値は幾つなど書かれた検査結果票を受け取ると思います。検査項目名には、いろいろな書き方があります。「血色素量､ヘモグロビン、Hb、HGB」といった具合です。これらが同じであることを専門家は知っていますが、コンピュータはそうはいきませんから、このままでは、情報システム間でデータを統合したり分析したりすることはできません。そこで、検査項目語彙を辞書にまとめて利用しようとしたのが、この研究です。

ここでは、主要な検査ラボにアンケートを送って、利用している語彙や単位などの関連情報を集めて同義語集を作り、同義語に代表語と ID を付けました。こうすることで、情報システムの間で交換する検査データの項目名については、対応できるようになりました。後で述べる「資源共有化システム」は、この方法を応用しています。

では、検査項目名が交換できれば問題が解消されるのかと言えば、そうは問屋が卸さないわけです。情報システム間で交換された検査値を利用する場合、先ほど述べた整数や文字といったデータ型以外にも、単位や計測法や記述規則などの情報が必要となります。このように検査値を記述する上で考慮しなければならない事柄、データの意味ということが出来ると思いますが、それらを属性としてまとめたのがこの表です。この研究では、これらの属性を検査値と一緒に交換することで、異なる情報システム間でのデータ利用の可能性を試みました。



異なるデータを統合する方法として、データ項目名とデータ属性の交換以外に、時空間データの利用可能性についての研究を進めています。例えば書籍資料のデータベースであれば、資料名や著者名や出版社名などに注目した検索やデータ統合が可能です。しかし、遺物や発掘された壷のようなものには、書名や著者名などはありません。一方、これらの対象物に関するデータベースの検索や統合を行う上で、壺などが利用されたと想定される時期や発掘された場所といった時空間データは有効です。多様な史資料を対象とする地域研究において、このような時空間データは親和性の高いものと言えます。また、時空間情報をグレゴリオ暦や緯度・経度のように数値化できれば、計量的な取り扱いも容易になります。

この図は、時間と空間を考慮した、我々の情報モデルです。主題軸は書誌データベースの書名や著者名や主題に相当する文字的な情報です。同じ主題を持つ史資料を検索するあるいは集める際に利用します。

それに対して、空間軸は対象物の場所に関連する情報です。緯度経度や地名などによって、史資料を検索する、集める、あるいは空間的関係を分析する際などに利用できると考えています。時間軸は空間軸と同じ機能を時間で実現しようとするものです。

このモデルを地域研究的に解釈すると、この 3 次元的な仮想空間内のデータ点の塊が、ある地域の全体像に対応すると見なせます。また、この塊を主題・空間・時間あるいはそれらを組み合わせた面で、ちょうど CT のように輪切りにして、可視化・分析することは、地域を多様な視点で観察することになると考えています。これは、情報学による地域研究という新しい可能性を示すものと期待しています。あるいは、そのようなことができたら面白いと考えています。

以上が前触れで、ここからは、東南地域研の情報基盤の紹介です。大きく分けて 4 つの機能を実現または実現しつつあります。データベースを「作る」機能、「使う」機能、「共有する」機能、「使う」機能です。

データベースがなければ、データ公開も、共有も、使うこともできませんから、「作る」機能は最優先です。そのために My データベースという情報ツールを構築しました。

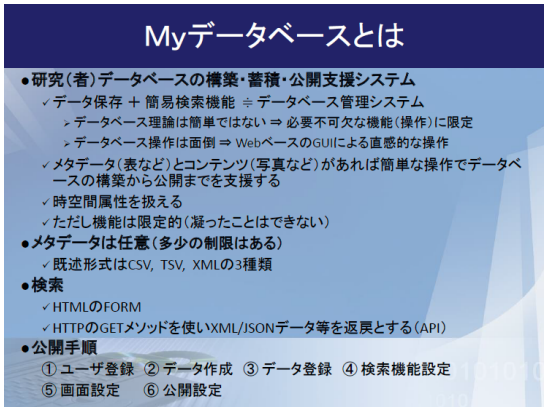
ところで、ユーザーインターフェースや可視化アプリケーションなど、データベースを「使う」機能も重要です。Ｍｙデータベースは誰もが簡単に使える情報ツールですが、機能は制限されていて凝った使い方は出来ません。そこで Application Programming Interface あるいは API とも呼ばれる機能を公開しています。API はプログラムでＭｙデータベースを使うための仕掛けです。研究に適したアプリケーションは、自分で作って下さいということです。

次に「共有する」機能です。Ｍｙデータベースで多くのデータベースが作られていますが、どの資料がどのデータベースに入っているかは、検索して見なければ分かりません。といって、一つ一つ検索するのも面倒なので、データベースを共有化して一

括検索できる情報ツールも構築しました。それが「共有する」機能です。

さらに「使う」機能として GIS ツールなども構築していますが、今日、この話はいたしません。

それでは、「作る」機能、「使う」機能、「共有する」について、詳しく説明します。



Ｍｙデータベースはデータベースの公開を支援するツールです。もしデータベースを自力で構築するなら、サーバーを購入して、OS やセキュリティ環境などを整えてから、データベースソフトウェアの設定や、アプリケーションを作成する必要があります。また、基本的なデータベース理論や検索言語の習得なども必要です。残念ながら、データベースの構築は特に人文社会学研究者にとっては、敷居の高い作業です。

しかしながら、人文社会学においても、データは EXCEL などのツール使って作ることが当たり前になっています。そこで、EXCEL レベルのデジタルデータからデータベースを作れるように工夫した情報ツールがＭｙデータベースです。

技術的な説明はしませんが、ID を必要としない、データ型やデータサイズを気にしない、データ項目名の記述制限を緩やかにしているなど、データ作成の制約を減らす工夫をしています。また、データベース作成の作業は、コマンドではなく、GUI によって直感的にできるようになっています。さらに、地域研究用なので、地図あるいはタイムラインを使ったデータ検索機能や表示機能をはじめから用意しています。

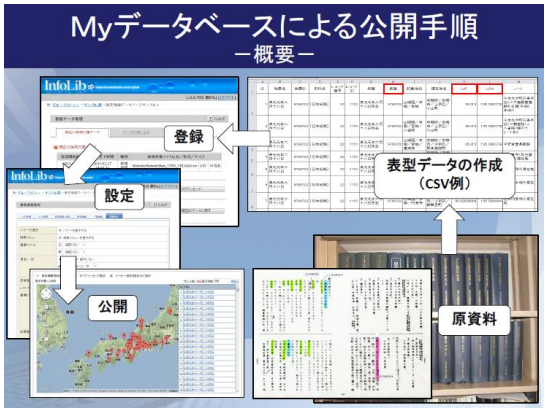
ということで、Ｍｙデータベースのメタデータスキーマは任意です。データフォーマットは、CSV、TSV、XML の 3 種類です。スキーマは任意と言っても、多少の制限はあります。たとえば XML の場合、データ構造は入れ子になっている、つまり整形式でなければなりません。また CSV あるは TSV の場合、下部構造を持たない、つまり第一正規形を満たしていることなどです。

余談ですが、EXCEL と ACCESS の違いを理解している人文社会学研究者は多くないようで、かなり手を入れないと使えない EXCEL データを持ち込まれる方もいらっしゃいます。

Ｍｙデータベースの利用法は、HTML フォーム、つまり通常の Web 検索ページによるか、API を使ったアプリケーションによるかの 2 種です。

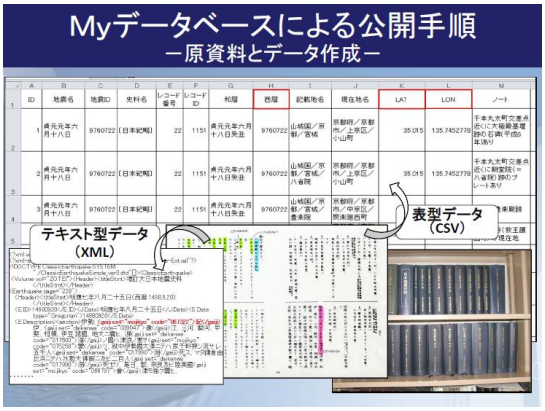
Ｍｙデータベースを利用する前に、ユー

ザー登録の申請をしてもらいます。なお、ここからの説明では、Ｍｙデータベースを作成する研究者をユーザーとします。管理者は、Ｍｙデータベースのシステム上に一定の作業領域を確保した上で、ユーザーがデータベースへアクセスするためのURIとIDとパスワードを発行します。

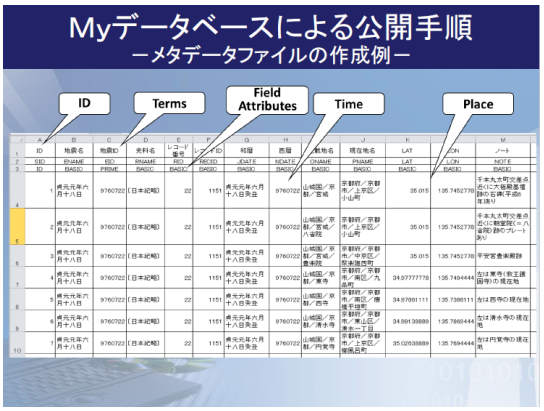


その後は、この図に示す手順でデータベースを作成します。

ユーザーは資料に基づいてメタデータを作りますが、多くの場合はテーブル形式のファイルとなっています。ここまで出来ていれば、発行されたURIを使ってＭｙデータベースを呼び出し、メタデータや画像データなどをアップロードします。さらに幾つかの設定､例えば検索するデータ項目や、表示するデータ項目、データベースの利用権限などを設定したら、データベース作成の準備完了です。データが少なくてエラーがなければ､「構築ボタン」を押して数分以内にデータベースの構築が終了して、直ちに公開できます。



もう少し詳しく説明します。これは文字資料の例です。ユーザーは、図の上側のような内容に関するメタデータをテーブルとして整理したり、左下のようにXMLを使って本文を翻刻したりします。また、この例では内容に関連する場所と時間も推定しているので、それらはLAT、LONと西暦としてメタデータに追加されています。なおＭｙデータベースにおいて、緯度と経度の表記はWGS84測地系に基づく十進表示、時間はグレゴリオ歴によるyyyymmddとなっています。



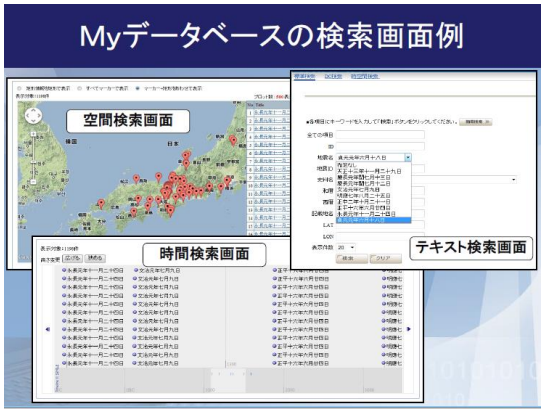
これは、Ｍｙデータベースに登録したメタデータの例です。上の3行分はＭｙデータベース独特の書き方で、1 行目はデータ

項目名、2 行目は他言語によるデータ項目名、3 行目は次に説明するデータ属性です。メタデータをＭｙデータベースにアップロードする段階で、属性の行は不要です。

データ項目名を二つ用意した理由は、例えばメタデータがタイ語で記述されているとデータ項目の意味が分かりにくいので、それを日本語や英語などの別の言語で記述する必要があったためです。地域研究データベースならではの工夫です。

メタデータをアップロードしたら、データ項目ごとに属性を設定します。データ属性ですが、どの項目が検索対象なのか、どの項目が表示対象なのか、どの項目が緯度経度なのかなどを設定します。少し高度な設定としては、データの値を参照してプルダウンメニューやスイッチボックスを作成したり、階層的な選択メニューを作成したりもできます。

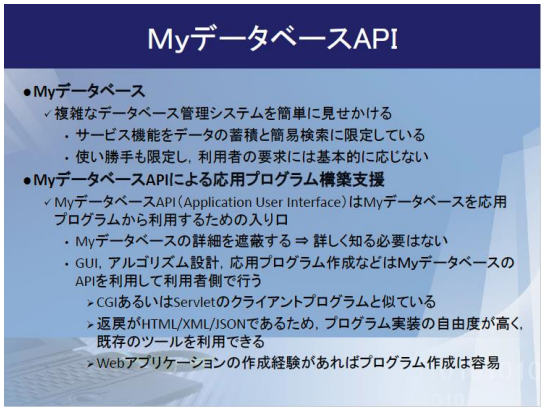
さらに、これとは別の画面ですが、データベースへのアクセス制限を設定したりします。一通りの設定が終了したら、データベースの構築を開始します。



これは、できあがったデータベースの検索画面の例を示しています。右上は通常のテキストによる検索画面ですが、一部ではプルダウンメニューを使っています。空間情報があれば、左上のように、地図画面を使った検索もできます。さらに、時間情報があれば、下側のようなタイムラインを使った検索もできます。

現在、50 個ほどのデータベースを公開していて、構築中や実験を含めると 100 個ほどのデータベースがＭｙデータベースにより構築されています。

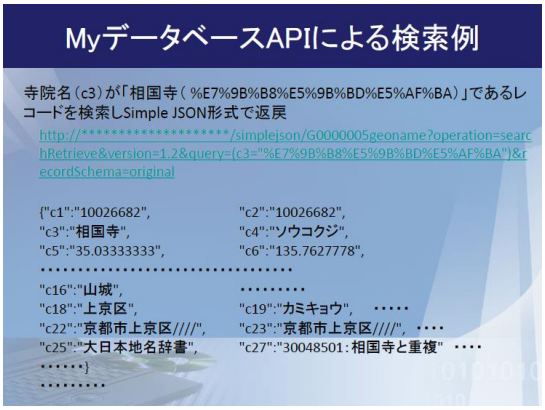
これらは、公開されているデータベースの一例です。



このように、Ｍｙデータベースを使うのは簡単なのですが、システム自体はけっこう巨大で複雑です。それを、情報系の教員 2 名だけで管理・運営しています。そのため、ユーザーからの個別の要求にはとても対応できません。というわけで、よほどの理由がない限り、Ｍｙデータベースの機能拡張に応じることはしていません。また、ユーザーのアプリケーションをＭｙデータベースに組み込むことも認めていません。システム更新によるＯＳやミドルウェアのバージョンアップへの対応や、セキュリティの責任分界などの問題があるためです。

その代わりに API を公開しているので、ユーザーは自分でアプリケーションを作ることができます。残念ながら、アプリケーションの自作は簡単ではありません。実際、少し昔であれば、Ｃ言語や、データベースシステムが提供しているライブラリや、インターネットのセッション管理などの技術や知識がなければ、アプリケーションは作れませんでしたから、専門業者に委託するしかなく、コストもかかりました。

API を使えばアプリケーションを作る負担は大幅に減ります。それでも、人文社会学研究者にはまだ敷居が高いかもしれません。ですが、Web プログラムの簡単な技術があれば、データベースシステムの中身を知らなくともアプリケーションを作ることができますから、外注してもコストはかなり低く押さえることができます。



これはＭｙデータベースの API の例で、地名辞書データベースから「相国寺」というお寺の位置情報を JASON という書き方で取り出しています。このように、API の検索式も検索結果も単純かつ定型的なので、アプリケーションプログラムの作成はかなり楽になります。

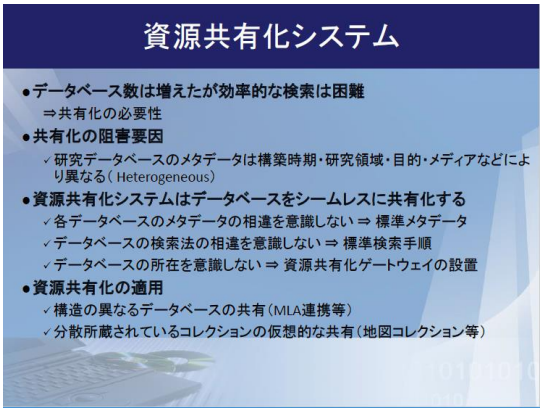
東南地域研の同僚が作った API アプリケーションの例を二つお見せします。

まず左側です。地域研究の論文などでは地図を多く使いますが、紙面に掲載できる枚数には制限があります。また、ズームインやズームアウトもできません。彼の工夫は、紙面には最小限の地図を掲載し、それ以外の地図は QR コードを使ってＭｙデータベースから参照しようとしたことです。QR コードにスマホをかざすと、京大出版会のサーバーが My データベースを検索して地図データをスマホに表示します。表示する枚数に制限はないし、デジタルデータなのでズームインやズームアウトも容易です。

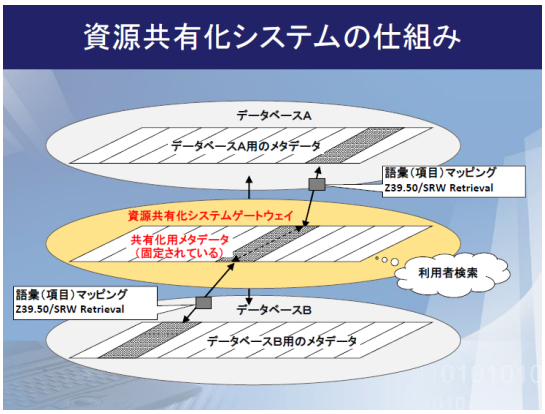
このアイデアの応用は広くて、別の研究書の例では、舞踏の写真の脇に QR コードをつけて、ここにスマホかざすと、動画や音楽が出てくるようになっています。

右側の例は、いわゆるマルチメディアデータベースです。右側はマレー語原本の画像で、画像データベースからの出力です。左側はそれを電子翻刻したデータで、テキストデータベースからの出力です。API を使って別々のデータベースを検索して、Web 上で合成しています。

これら以外にも GIS と組み合わせたアプリケーションなども作成されており、ユニークなものが、これからも作成されると期待しています。



次に資源共有化システムです。前にも述べましたが、Ｍｙデータベースで作られたデータベースは、ユーザーの研究目的が異なるので、同じメタデータ構造を持ったものはありません。一方、Ｍｙデータベースの検索者にとっては、どのデータベースに何が入っているのか分かりません。かといって、一つ一つ検索するのは面倒です。資源共有化システムは、Ｍｙデータベースで作られたメタデータ構造の異なるデータベースを統合検索する仕掛けです。実際は、東南地域研以外のデータベースも統合検索できるようになっています。



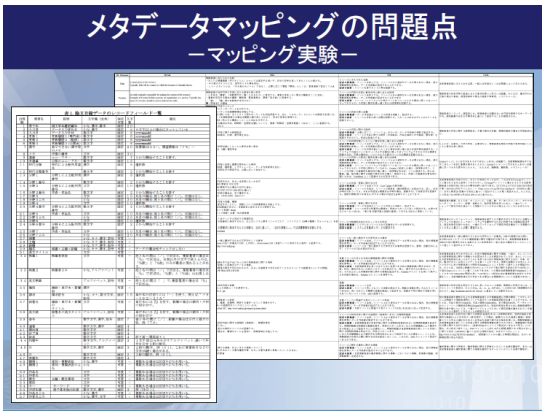
資源共有化システムの仕掛けは単純で、個別データベースから独立した中立的なメ

タデータを利用します。これを共有化メタデータと呼んでいます。現在の共有化メタデータは、Durbin Core と Metadata Object Description Schema（MODS）の 2 種類です。

Ｍｙデータベースで作ったデータベース、図ではデータベース A と B を、資源共有化システムに登録する際に、各メタデータのデータ項目と、共有化メタデータのデータ項目の関連を定義します。これをマッピングと呼んでします。

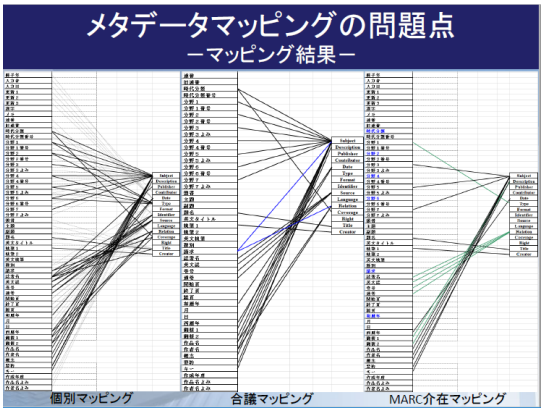
検索者は、個別のデータベースではなく資源共有化システム、図の資源共有化システムゲートウェイにアクセスします。検索には共有化メタデータの項目を使うので、検索者は個別のメタデータを気にする必要はありません。

資源共有化システムゲートウェイは検索命令を各データベースに渡しますが、その際に対象データベース用の検索命令に変換します。資源共有化システムゲートウェイと各データベースは、Z39.50 または SRW という情報検索プロトコルで自動的に通信します。そのため、検索者は個別データベースの所在や検索式を気にする必要はありません。



言うまでもないことですが、資源共有化システムの検索性能は、各メタデータと共有化メタデータのマッピングに異存します。

そこでマッピングの実験をしました。この図の左は、私が以前に勤務していた国文学研究資料館の論文目録データベースのデータ項目の説明文、右側は Dublin Core の説明文です。この説明に基づいて、論文目録から Dublin Core へのマッピングを試みました。被験者は、国文学研究資料館の目録担当図書館員と教員の 5 名です。



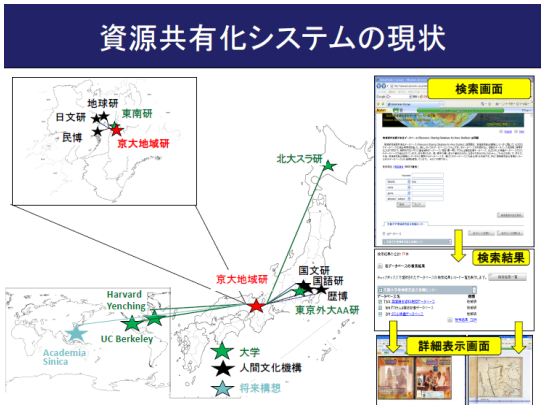
左側は各被験者のマッピング結果をまとめたものです。できる限り全項目をマッピングしようとしていた者もいれば、できるところだけをマッピングした者もいるなど、マッピングの程度はさまざまでしたが、そ

の結果もバラバラでした。

中央は、個別マッピングの結果を受けて、5 人で検討したマッピング結果です。「これでなければならない」ではなく「これなら妥協してもよい」というレベルの合意です。個別マッピングよりも収斂した印象を受けますが、論文目録から Dublin Core への対応が 1 対多になっている項目があります。

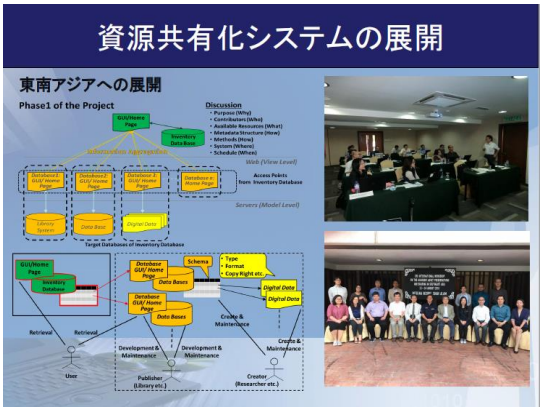
右側は MARC へのマッピングを介した結果です。目録関係者なので MARC はご存知でした。MARC から Dublin Core へのマッピングは、既存の Crosswalk を使いました。その結果、マッピングされない項目が増えた反面、マッピングはかなり整理されました。

マッピングがどれだけバラツクのかを調べる実験だったので、バラツキの違いが検索精度にどれだけの影響を与えるのかについての検討は行いませんでした。しかし、想定した以上にバラツキが大きかったので、資源共有化システムの評価は、「統合検索できないよりは良いけど、検索精度は余り良くない」というところと考えています。



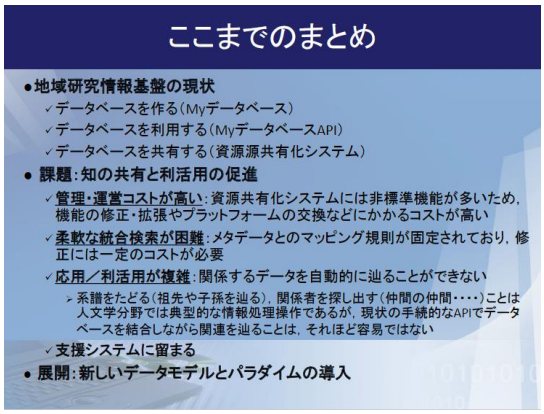
資源共有化システムの現状は、この図に示した通りです。東南地域研の外部のデータベースとして、人間文化研究機構の百数十個のデータベースのうち、地域研究に関連する国立民族学博物館、総合地球環境学研究所、北海道大学スラブ・ユーラシア研究所図書館、東京外語大学 AA 研、UC バークレー・東アジア図書館、ハーバード・イェンチンライブラリーの統合検索を実現しています（本発表後、ハーバードについては、情報システム変更により統合検索を停止中）。

このような成果を受けて、資源共有化システムの対象を東南アジアの大学図書館などに広げようとしていますが、進んでいません。東南アジア諸国の情報基盤のレベルがあまりにも大きくて、日本と遜色ないあるいは日本以上の国から、基本的な図書館情報システムすら整っていない国まで様々です。この状態で資源共有化システムを導入するのは困難と判断しました。



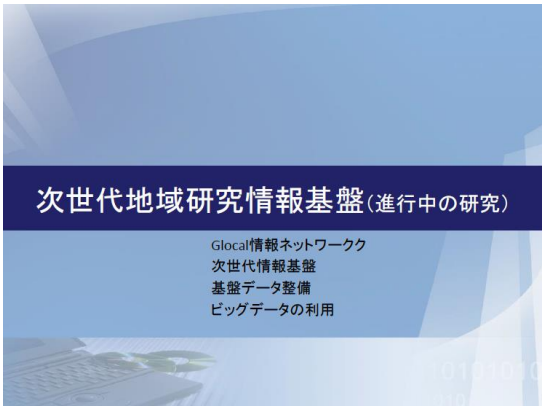
ところが、この議論を進めている中で、お互いがどのようなデジタルデータを持っ

ているのか知らないということが明らかになりました。たとえば東南地域研では三印法典などの資料価値の高いタイ語データベースを公開していますが、現地の人は知りませんでした。そこで、共有化はいったん諦めて、デジタルデータのインベントリシステムを共同で作ることにして、このスライドのような国際 Workshop などを開催しています。



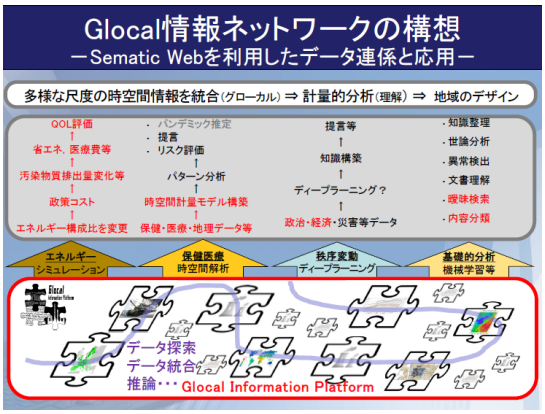
ここまでのまとめです。データベースを「作る」、「使う」、「共有する」ための地域研究情報基盤はだいたいできました。

ただし、幾つかの問題があります。というのは最初の設計が 1999 年でほぼ 20 年前です。よく 20 年も動いてきたなと思っています。ですが、古い技術や非標準の技術を多く使っているので、管理・運用コストがかかります。また、メタデータマッピングが固定されているので融通が利かない。それから、ただ検索するだけで終わってしまって、それ以上のこと、例えば関連する Web データに繋げることが難しい、といった問題があります。つまるところ、今のオープンデータやオープンサイエンスに対応することができません。



そこで、新しい地域研究情報基盤の構築に着手しました。これを Glocal 情報ネットワークと呼んでいます。Glocal は、global と local を合わせた jargon です。

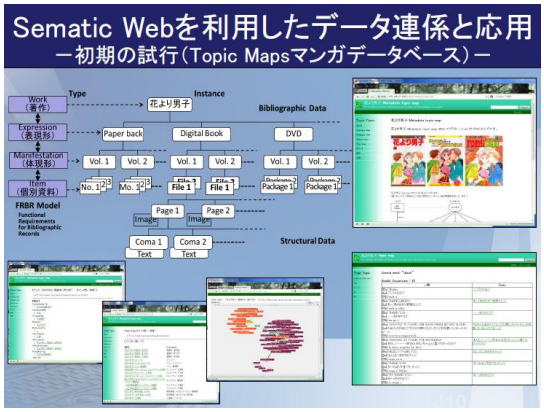
地域をいろいろな尺度、たとえば地球レベルや国のレベルで俯瞰したり、村のレベルで詳細に検討したり、時間についても長い時間で見たり、短い時間で見たり、時間の移動で見てみたりと、そういうことが柔軟にできるような情報基盤にすることを意図した名前です。



Glocal 情報ネットワークでは、大学内のデータベースとインターネット上の膨大な

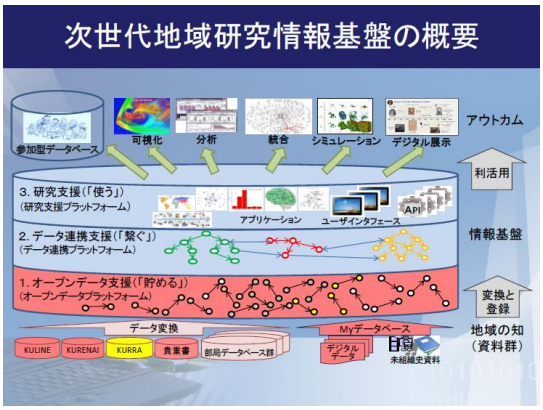
データのシームレスな結合を目指しています。多くの情報がインターネットにあって、これを使わないと地域研究も進まない状況になっているためです。たとえば巨大ハリケーンや巨大地震などの大規模災害が発生した場合、最新の情報はテレビなどのマスメディアではなくインターネット上を流通しています。大統領選挙のような政治情報や、パンデミックなどのような医療情報や、気象などの環境データでも同じです。

このような Web 上のビッグデータと我々のデータベースを効率的に繋げる、反対に、我々のデータベースを Web からより広く利用して貰えるためには、データ連係のための新しい技術的な枠組みが必要です。今のところ、Semantic Web 技術、特に Resource Description Framework（以下、RDF）による Linked Open Data（LOD）が回答ではないかと考えています。



少し脇道にそれますが、Semantic Web の機能を実装する枠組みとして RDF が有名ですが、最初のころは Topic Maps を使っていました。RDF よりも高度なデータの関連づけができることと、研究グループに Topic Maps の専門家いたためです。

この図は、Topic Maps を使って漫画のデータベースを構築した例です。面白いデータベースなので、これを話すだけで 1 時間はかかってしまうのですが、要は、漫画情報を本気で記述しようとするとネットワーク状になってしまうので、普通の図書目録など記述しきれず、Semantic Web の手法が必要になったということです。



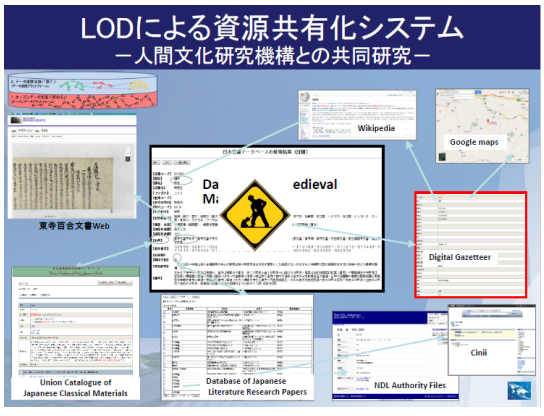
これは Glocal 情報ネットワークの構造と現状を表した図です。3 層構造となっています。

一番下の層が RDF データベースの実体です。M y データベースに蓄積されているメタデータを RDF トリプルに変換して蓄積する作業を続けています。

これらの RDF トリプルの主語と述語の語彙には、元のデータベースで使っていたデータ項目名を流用しています。語彙がバラバラなので、このままでは統合検索など

に利用できません。語彙の関係を記述したオンロジーが必要となり、それを実装するのが第 2 層です。これまでの資源共有化システムの機能を再現する場合、各メタデータのデータ項目名と共有化メタデータのデータ項目名の関係を RDF で定義します。あるいは図書館の視点でデータを統合するならば、MARC を中心したオントロジーを定義します。もちろん、独自のオントロジーを定義しても構いません。これにより、これまでの資源共有化システムよりも柔軟なデータの統合が可能となります。

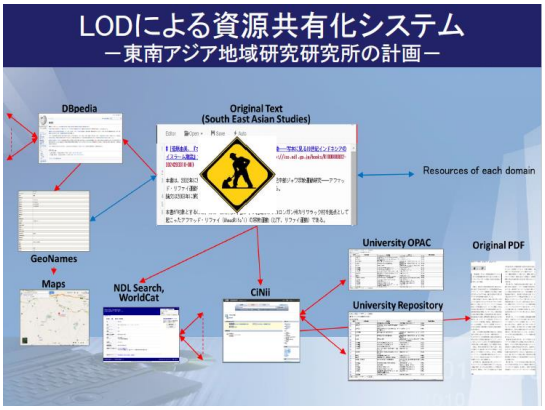
これまでに、第 1 層と第 2 層を少しずつ作り始めているところです。第 3 層は応用プログラムで、実装は進んでいません。この第 1 層と第 2 層を公開すれば、我々のデータベースをより広く利用して貰えるようになると期待しています。



その実験を人間文化研究機構と共同で行っています。これは、国立歴史民俗博物館の荘園データベースのあるレコードを起点として、関連する論文や地図データベースを辿れるように工夫した情報ツールです。例えば荘園の所在地名を使って地名辞書を検索し、地名辞書から荘園の緯度・経度を取り出して、地図上に荘園の位置を表示します。つまり、RDF で記述されている情報の所在 URI を参考にして、関連する情報を次々と自動的に繋いでいきます。

これまでキーワード検索は検索結果を表示して終わりでした。ですが、この情報ツールでは、検索をきっかけとして、関連していそうな情報をどんどん繋げていくので、新しい発見やヒントを得る上で有用な研究ツールなることが期待されます。

ただし、すべてのデータに URI を付けなければならないので、データ作成は時間と手間のかかる大変な作業になっています。



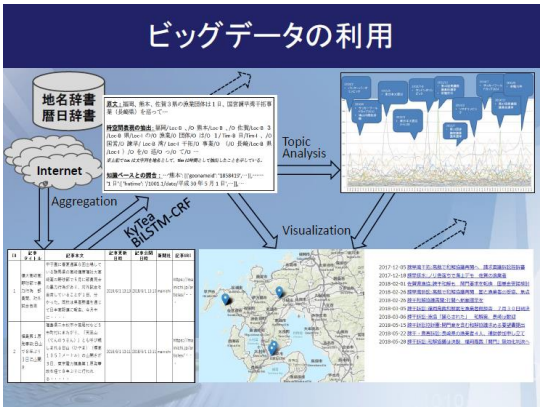
これは東南地域研で進めている別の実験です。論文検索システムは検索された論文が終点だったのですが、ここでは論文を起点にして、関連する情報へリンクします。残念ながら、これも工事中です。

先ほど地名を緯度・経度に変換すると言

いましたが、そこではデジタル地名辞書を利用します。現在の地名については国内外でフリーの Web サービスが幾つかあって、それらを使うことができますが、日本の歴史地名についてのフリーのデジタル辞書はありません。ないと困りますが、かと言って誰も作ってくれなかったので、我々で作ってみました。

幾つかの史資料を使っていますが、最初のデジタル地名辞書には大日本地名辞書を使いました。辞書の見出し語が地名なので、その内容を読んで場所を推定し、それを緯度経度に変換するという作業を繰り返しました。

この手順を応用して、延喜式の式内社名、明治につくられた旧 5 万分 1 地形図の全地名などを緯度・経度に変換した結果、これまでに約 37 万件の地名を収容したデジタル地名辞書を作りました。日本の歴史地名のフリーデータとしては、最大規模ではないかと考えています。このデータは、人間文化研究機構からダウンロードできます。



最後に、データベースとは違うのですが、進行中の 2 つの研究について手短にお話しします。

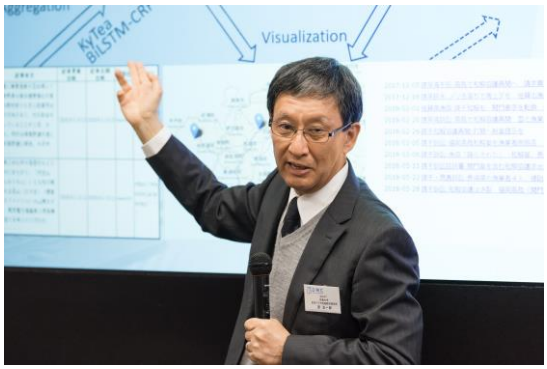
Glocal 情報ネットワークでは、大学内のデータベースとインターネット上の膨大なデータのシームレスな結合を目指すと言いましたが、これまでに説明したのは、自分達のデータベースの Linked Open Data 化についてでした。

現在進行中の研究の一つ目は、Web ビッグデータについてです。ここではソーシャルネットワークデータを使って、地域の状態を可視化することを目指しています。ソーシャルネットワークデータを使った地域研究の例はいくつもありますが、多くは、災害のような特定のデータ収集と分析が目的で、データ収集には、キーワード検索を非常に巧みに使っています。

ところが、この研究では地域の状況を俯瞰することが目的なので、特定のキーワードでデータを選別することはできません。そこで、ここではソーシャルネットワークデータを機械に分類させるという方法をと

っています。実験段階なので、ソーシャルネットワークデータと言いながら、実際は Web 上の新聞記事を利用しています。

手順は、この図に示すように、新聞データを定期的に crawling して、広告などの不要部分を取り除いて、記事テキストデータを作成します。次に記事のテキストを単語に分解します。その際に地名や時間名に関する語彙を抽出して、地名についてはデジタル地名辞書に当てはめて緯度・経度に変換します。単語単位に分割されたテキストにトピックモデルという手法を使って、内容を 200 種類のトピック言い換えれば主題に機械的に分類します。最後に、記事を地図上に可視化します。これによって、地域を 200 の観点から俯瞰することができるようになります。字句解析とトピックモデルの処理は機械学習の応用です。



これまでの地域研究のように、少数の文献やデータを緻密に読み解くのではなく、ざっくりであるが全体的に眺めるというアプローチです。最近 Distant Reading という言葉が流行っていますが、それに類した方法と考えています。必ずしも正確とは言えない大量のデータを含めて分析することで、地域をより全体的に俯瞰できるようになり、新しい研究のヒントが見えてくる可能性があれば、素晴らしいかなと考えています。

最後に、研究データの長期保存はいま喫緊の問題です。検討を始めましたという段階なので、笑い話としてお聞き下さい。

メディア変換、いまはハードディスクやクラウドを使っているので、ほとんど問題ないです。ですが、フロッピーディスクなどの昔の媒体に入っているデータの変換は大変です。何といっても、装置がなくなりつつありますから。

ワープロ専用機なんて若い方は殆ど知らないでしょうね。このころに作成された文書データは、今となってはほとんど使えないと思いますが、プレインテキストでよければ、何とかなるかもしれません。

ゲーム、これは実は簡単だと思っていたのですが、専門家にお聞きしたら、非常に難しいということが分かりました。プログラムはエミュレータなどを使えば何とかな

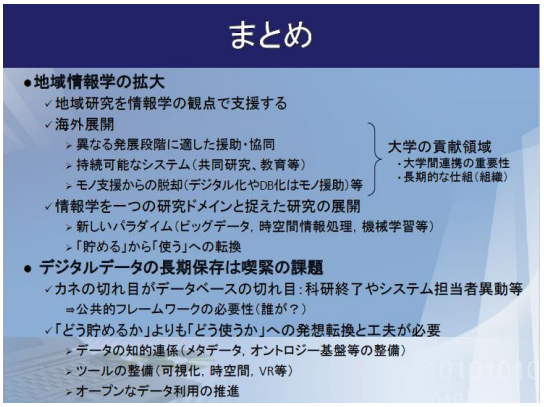
りますが、問題はジョイスティックなどのハードウエアです。設計図を保存しておいて、必要に応じて作るしかないようなので、これは意外に難しいです。

肝心のデータベースです。もしデータベースソフトがなくなってデータしか残っていなかったら、上の図のようなバイナリデータを分析するしかありません。かつて、この経験をしたことがありますが、本当にたいへんな作業です。

次の図は、データだけが読めた場合です。何のことか分かりません。

次の図ですが、これだと血液検査データかなと、何となく分かる気がしますが、実際には使い物になりません。まともな研究者だったら絶対に使わないでしょう。なぜかというと、例えば、最初の項目は GOT で単位は U/L と分かります。ちなみに GOT は肝臓の機能を調べるものです。しかし、どんな計測法を使ったのか、どんな人が検査の対象だったのか、どのような状態で血液を採取したのかなどが明らかでなければ、比較研究に使うことができません。

先ほど言ったように、研究データの長期保存が喫緊の問題となっていて、各大学では図書館やメディアセンターなどが中心となって検討しています。ですが、話を聞いていると、ビットデータを安全に蓄積することが話題の中心になっていて、データを利用する際に必要となる情報についての議論はあまりされていないし、その重要性を述べても、なかなか理解されていないという印象を受けています。適切な標準メタデータが定義されれば良いのでしょうが、当面は、機器の説明書、実験ノート、論文、あるいは手書きのメモなど関連する文書情報をデータ本体に混載させるしかないのかもしれません。



メタデータの話からだいぶズレてしまいましたが、まとめです。

ここでも最後にちょっと書いたのですが、研究データの長期保存の問題は、どう保存するかよりも、どうやって使い続けるかを考えたほうが良いのではないかと考えています。そのためには、先ほど述べたデータのオープン化とデータの関連付けを続けて、色々な経路から古いデータに辿り着ける、あるいは発見できるようにすることが重要になると思われます。そのためのオープン環境あるいは情報インフラの整備を進めていく必要があり、そうなってくると、メタデータのあり方がとても重要で、ことによったら、もう一度再考しなければなら

いのかもしれません。

まとまりのない話になってしまいましたが、以上です。ご静聴、どうもありがとうございました。

原 正一郎（はら しょういちろう）

1957 年、千葉県生まれ。京都大学東南アジア地域研究研究所教授。専攻は情報学。共著に『歴史 GIS の地平』（勉誠出版）、『ナチュラルコンピューテーション 1』（パーソナルメディア）、翻訳書にM・ジェームズ『人工知能 BASIC』（啓学出版）など。

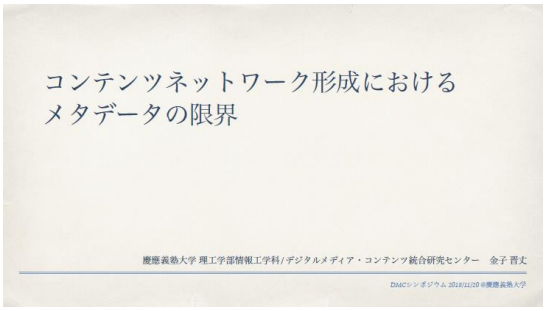
講演

「コンテンツネットワーク形成における メタデータの限界」

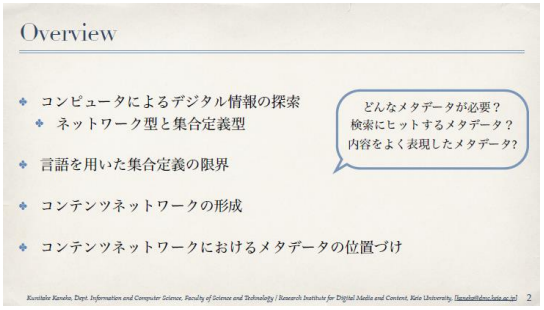
金子晋丈

（慶應義塾大学理工学部専任講師

・DMC 研究センター研究員）

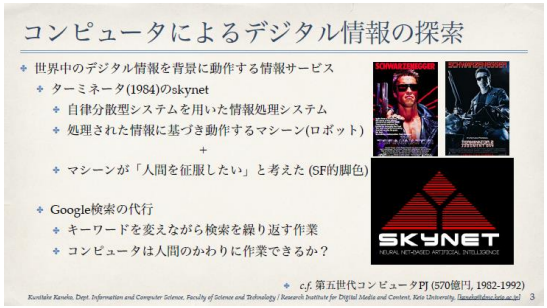


ご紹介ありがとうございました。理工学部情報工学科の金子と申します。DMC 研究センターの研究員も務めております。きょうはコンテンツネットワークとメタデータに関してお話ししたいと思います。ここにいらっしゃる方はコンテンツネットワークとはなんだと思われていると思うのですが、そのあたりも含めてお話しできればと思っております。きょうの内容ですが、コンピューターによるデジタル情報の探索はどういうプロセスを踏むのだろう、人間でもそうですが、情報の探索のフェーズみたいなところを、最初お話しできればと思っています。



現実にいまどういう形で探索をやっていて、本当はあるべき姿とはどういうものなのかなというところから始めて、現在のキーワード検索に代表される言語を用いたやり方の限界が伝わればいいかなと思っております。一方で、「ネットワーク型と集合定義型」と書いていますが、ネットワーク型のほうがコンテンツネットワークというものなのですが、それはどういうふうにしたらできていくのだろうとか、最終的にはメタデータはどういうものなのかというものをお話しできればと考えております。

まず、コンピューターによるデジタル情報の探索ですが、こちらにある映画のタイトルを持ってきました。『ターミネーター』、ご存じない方はいらっしゃらないと思いますが、『ターミネーター』の最初のものが左側、2 個目が右側、その後続編も出ていますが、スペースがないので二つだけです。



最初のものが出されたのが 1984 年です。僕はつい最近までこの 1 は見たことがなかったのですが、『ターミネーター』のストーリーは皆さんご存じですか。漠然と覚えていらっしゃる方もいると思いますが、マシンというロボットとコンピューターが一体化したようなものが人間と戦っていくという話です。その悪の権化みたいなやつが skynet というもので、ニューラルネットベースの artificial intelligence が動いています。これはべつに僕が付け加えたわけではなくて、これが当時から言われていたトピックです。ですから、皆さんが最近 AI と言っているのは、1984 年の話をしているということになります。

1984 年はどういう時代かと考えると、たとえば国内では第五世代コンピュータプロジェクトというのがありました。知っている方は知っていると思いますが、調べましたが、莫大な 570 億円を 10 年間かけて使い、人工知能ベースのコンピューターをつくるという新しいプロジェクトです。ですから、時期がドンピシャであるというのを、まず一つ分かっていただけると思います。

もう一つが、インターネットです。インターネットというのは、自律分散型のシステムで動くと。その二つを組み合わせて、『ターミネーター』というのは『アバター』をつくったジェームズ・キャメロンという監督がつくっているのですが、ジェームズ・キャメロンはたぶんここの二つを組み合わせて、このストーリーを組んでいるのではないかと、僕は現段階では思っています。

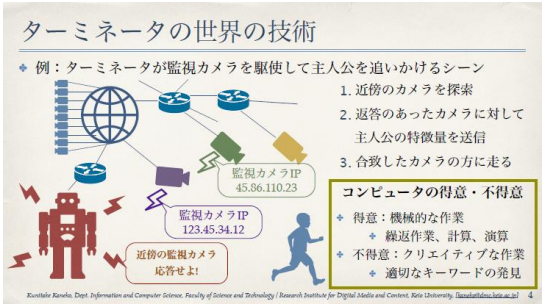
整理しますと、自律分散型インターネットにそれぞれが独自のルールに基づいて動きながら、全体として協調的な動作をしていく。自律分散型システムを用いた情報処理システムと、処理された情報に基づき動作するマシン、ロボット、それが技術的な背景にあり、SF 的な味付けとして、マシンが人間を征服したいと考え始めたということで、この『ターミネーター』はストーリーとしては動いていることになります。

もちろん誤解なく言っておきますと、これは SF 的脚色ですので、べつにコンピューターが頑張ったところで、マシンが人間

を征服したいと考え始めることはないので、これを議論すればするほど世の中が破滅に向かうのかとか、金子の研究は軍事研究だとか、全然そういうわけではないのでご安心ください。

こういった『ターミネーター』の世界が、1984 年にある種 SF として出てきたにもかかわらず、一向にコンピューターが独自に情報を解析して、それを使っていくという形にまだまだなってきていないのはなぜかというと、技術的には考える必要があるかなと思います。

これが 1984 年の例ですが、いまのわれわれの生活でいくと、Google 検索の代行をできるコンピューターがあるかというのと、ほぼほぼ似ている考え方になると思います。皆さんが Google 検索をやるときは、キーワードを入れて結果が返ってくる。少し不満のある、所望の結果が返ってこなかった場合は、キーワードを変えてもう一度入れて結果が返ってくる、それを何回か繰り返すという作業が Google の検索作業になると思うのですが、これはコンピューターに代行させることができません。なぜかというと、コンピューターは次に浮かぶキーワードが浮かんでこないからです。これがそもそもの上の話（ターミネータの skynet）と下の話（Google 検索の代行）で実は共通しているバックグラウンドということになっています。



さて、技術者ですので『ターミネーター』の世界を実現しようと思ったら、どんな IT システムがあればできるのかを考えるわけです。SF を見ながら、「これはどういう技術があったらできるかな、これは無理だね」ということを考えるのは楽しいのですが、『ターミネーター』のリアルなやつを持ってくるとあれかなと思いまして、ロボットにして、これは逃げている主人公みたいな絵です。

すごく典型的なシーンとして、スライド 4 に、監視カメラが街中にあって、世界中に監視カメラがインターネットにつながって存在しているみたいな絵を描いています。では、『ターミネーター』のシーンでいくと、マシンが逃げている人を探し出すためにはなにをしなくてはいけないかと。よくイメ

ージ的には監視カメラを捉えて、あそこに映っているというのが分かって、あちらにいるんだと追いかけるシーンになっていると思いますが、それをやるためにはものすごく大変なことが必要なんです。

なぜかというと、ロボットはこの人が映っているカメラがどこにあるかを知らない。そもそも自分の周りにどんなネットワークカメラがあって、それが本当に自分の近くにいるかが分からないので、仮にたとえば世界中に 1 億台の監視カメラがあったとして、1 万台でもいいですが、その中から自分の近辺もしくは逃げている主人公の近辺のカメラを、10 台でも 100 台でも限定してチョイスすることができないので難しくなってきます。

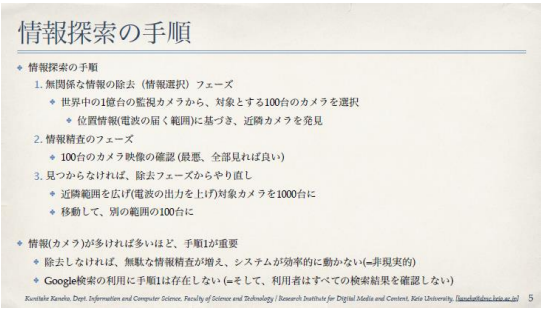
ネットワーク的には違う位置にそれぞれいます。この監視カメラはこんな IP アドレス（45.86.110.23）、このカメラはこんな IP アドレス（123.45.34.12）、違う場所にあって、違う場所にあるから、ただ単に IP アドレスの一番下の桁を 1 増やせばいいとか、1 減らせばいいとか、そういう話ではなくて、なにかしらのメカニズムでこのカメラを見つけてこないといけないというところに、最初の技術ポイントがあります。

ところが、映画ではこんな話はなくて、自動的に見つかったみたいな、「やばい、やばい」と思わされるわけですが、見つけてくることになります。一番僕が安直に思い付いた技術的な解決策は、ロボットが自分から電波を発して、近くの監視カメラがそれに対して応答するわけです。電波が届く範囲にいる監視カメラが「僕の IP アドレスなんだよ」というのを返してあげる。そうすると、このロボットは自分の近くにある監視カメラの一覧を入手できるので、それに基づいてその監視カメラたちに、「こういう特徴の顔をしたやつがいたら教えろよ」みたいなことを言うと、『ターミネーター』の世界が出来上がってくるかなと思うわけです。なんとなく伝わりましたか。

コンピューターには得意なことと不得意なことがあるわけです。ここを理解しないとうまくシステムがつくれないわけですが、コンピューターが得意なのは完全機械的な作業です。繰返作業やこういう計算をしなさい、こういう演算をしなさいというのがコンピューターの得意なこと。不得意なことは、なにか新しい発想を求めること、これはコンピューターには現段階ではできないものになっています。

これを考慮してこんなシステムをつくっ

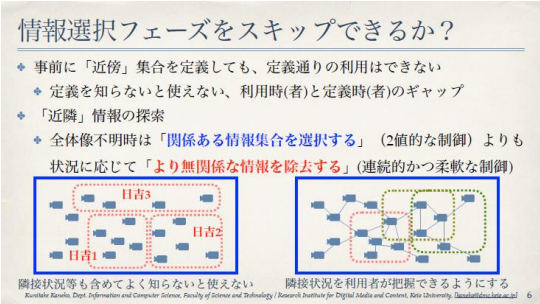
てみたわけですが、これをもう少し整理してみます。これが情報探索の手順の一つのサンプルと思って整理しているわけですが、最初はその 1 億台なり 1 万台なりの監視カメラから、追いかける相手が映っているであろう、映っているかもしれないエリアの 100 台を選んでくるフェーズ、逆に言えば 1 億台から 100 台以外を消すフェーズが最初にあります。その次に出てくるのが、その 100 台のカメラ映像から顔が一致するものを見つけてくる、情報を精査するフェーズになる。100 台で見つからなかったら、もう一度 100 台を 1000 台に変えて探し直しましょうというのがだいたいの流れになります。



こう考えたときに、いまの検索技術はなんなんだろうとか、いまわれわれがやっている情報の探索はどういうことなんだろうというので、たとえば 1 番の情報選択フェーズをカットするとなにが起こるのだろうか。1 億台世の中にあるので、仮に 1 億台のリストがあったとしたら、1 億台にすべて問い合わせて、その顔が映っていますかというのを返す。それだけ聞くと、それでもいいのではないかと思うかもしれないですが、そういうことをしたいユーザーが 1000 人いたら、1 万人いたら、10 万人いたら、そのシステムは動くのだろうか。そうすると、それは動かないわけです。リクエストが集中して処理ができないので、動かないということになります。

では、皆さんがやられている Google 検索は、この例でいうと、実は最初から情報の精査のフェーズに入ってきます。キーワードを入れてこれに合致するものだけを返してくださいということです。たとえば、Keio University で Google 検索すると何件ヒットするか分からないですが、軽く 1 万を超える数字がヒットしましたみたいなものが出てくると思います。皆さんが見るのは最初の 10 個ぐらいです。それは精査していないことになりますね。1 万件ヒットしたうちの 100 件だけ、最初の 10 件だけを見ると、残りの 99 万 9990 件は無視されるということで、実は適切な情報探索のフェーズを経ていない、情報探索ができていないというのが、実際のいまの IT における情報探索の問題と考えています。

最初の情報選択、ここには絶対あいつはいないというのを排除するフェーズをスキップできるかという問題は、いろいろ考えさせられる問題です。それを二つのやり方で対比的に書いております。



たとえば、『ターミネーター』の世界で、頭の中はアメリカのハリウッドのような感じになっているところで、日吉というすごくローカルな地名を出してあれなのですが、たとえば日吉に監視カメラがあって、この監視カメラ群は日吉１丁目にあります、この監視カメラ群は日吉 2 丁目にあります、この監視カメラは日吉３丁目にありますとあらかじめ定義してあげて、それを使って逃げ続ける主人公をうまく見つけられるかという話をここでは取り上げたいと思います。

たとえば、逃げる相手がここだけでぐるぐる回っていて、逃げているのが日吉１丁目だというのが分かれば、なんとなくオーケーそうです。でも、逃げているときに、日吉３丁目のほうに行ったり２丁目になったり、実際日吉１丁目と日吉３丁目は隣接していないのですが、架空の話だと思っていただいて。ぐるぐるやっていると、いまその主人公は何丁目にいるのだろうか、あの通りを隔てると日吉何丁目に行くんだろうか、そういったことを全部把握した上で、いまどこの集合に対して検索をかけないといけないのかということをやらないといけないのが、実はこちらのモデル（スライド 6 の左の図）になっています。

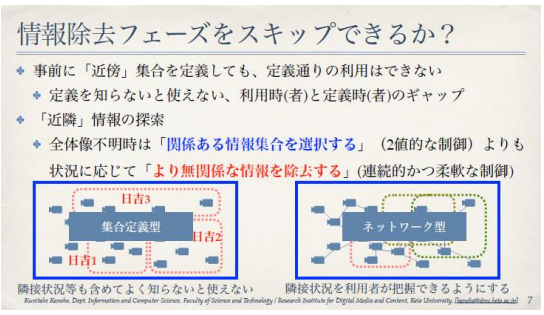
こちらに別のモデル（同、右の図）を書いてみました。こういう集合を捉えるのではなく、それぞれのカメラが相互に近いカメラのことを知っているだけの世界です。この世界だと１カ所場所が決まって、いまここに映ったよとなると、ユーザーはそこから近い位置にいるので、その近傍のカメラを自在に設定することができます。仮に 100 台のカメラでいなかったら、1000 台に広げてあげましょうとか、ユーザーが移動したらユーザーが移動した方向でまた円を描いてあげましょうという形で、変幻自在にその集合を変えてあげることができるのが、こちらのネットワーク型のモデル（同、右の図）になっています。

すなわち、ここで言いたかったのは、こういう形態（同、左の図）を取るときは、関係ある情報の集合が相互にどういうつながりになっていて、どのエンティティがどこに属しているのかを完全に把握している

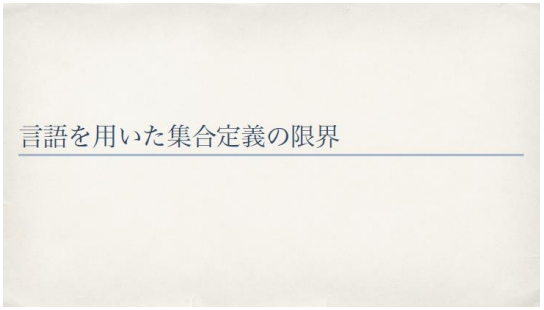
場合には、非常に速やかに場所を特定できて、情報を選択することができるのですが、よく分からないときは実はこれはうまく使えないことになります。いまそもそも自分は日吉 1 丁目にいるか 2 丁目にいるかすら分からない場合には、左のモデルはなかなか使えないということになるかと思います。

言いたかったのは、たとえば Google 検索をするときに、皆さんは世界中のボキャブラリーを知っていて、それぞれがどういう位置で集合が管理されていて、それが分かっているのであれば、キーワードを入れて適切に情報を持ってくることができるけれども、それを知らないのであれば、うまく使えないということをこの例は表しています。

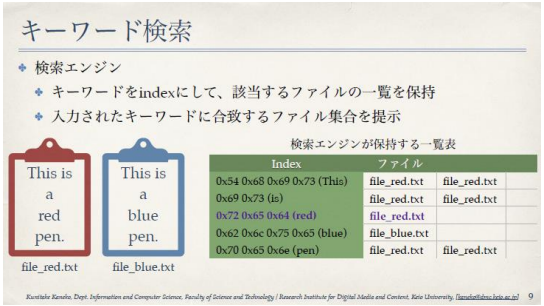
どちらかというと、世界中の情報を知らないのであれば、こちらのモデル（同、右の図）のほうがより適していて、自分が少なくとも知っているところから徐々にたどって、本当に知りたいところに近づいていく、そういうことができるのがこの右のモデルになろうかと思います。



以下では、この左のほうを集合定義型、右側をネットワーク型という名前で呼びたいと思います。



次に、集合定義型で言語を用いて集合を定義した場合に、さらにどんな問題が加わってくるのかについてお話ししたいと思います。ここは先ほどの原先生のお話と微妙にオーバーラップする部分ですが、たぶん原先生のほうが非常に具体的な事例を挙げてご説明いただいていて、僕は少し抽象的な話になっているかもしれないので、うまく頭の中で merge させながら聞いていただきたいと思います。

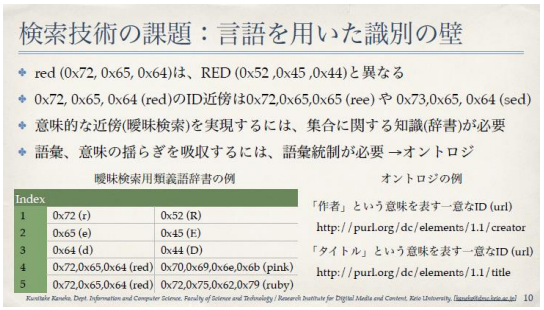


非常にシンプルなキーワード検索の実現方法を書いております。キーワード検索をやるときには、たとえば file_red.txt と file_blue.txt があったときに、それぞれの要素である単語がそれぞれ表に管理されることになります。「This」というのは This

という意味ではなくて、実際コードになって落ちていきますので、0x54、0x68、0x69、0x73、これが This です。ですから、This のコンピューター的な識別は「This」ではなくて「0x54、0x68、0x69、0x73」、この数字の列によって一意に決まってくるということになります。「This is a red pen.」「This is a blue pen.」、これを両方全部分解していくとこういう形になって、それぞれの数字の列が含まれているファイルが一覧表になると、キーワード検索のベースが出来上がることになります。

すなわち誰かが「This」という検索キーワードを入れると、入力された文字列が 0x54、0x68、0x69、0x73 だったら、これだねと言ってこの二つが検索結果に返るし、「red」の 0x72、0x65、0x64 というのが与えられると、この file_red.txt が返ってくるというのが検索の基本的な動きです。

あえてここでこの数字の列を、分かりにくいにもかかわらず出したのには理由があります。それは、たとえば小文字全部の「red」と大文字全部の「RED」は数字としては違うからです。「0x72、0x65、0x64」と「0x52、0x45、0x44」というのはどう見ても似つかない。red と RED だから一緒だと、それは意味の世界であって、コンピューターの世界では完全に数字の列で判断しますので、それは全く違うものということになります。



よりコンピューターの立場に立って考えると、0x72、0x65、0x64 に近いものはなにか、すごくコンピューター的に簡単に近いものを出すと、0x72、0x65、0x65、1 ビット違いですが「ree」です。全然 red とは違います。もしくは、red の r のところだけが変わって 0x73、0x65、0x64 で「set」。こういうふうに ID 近傍、機械的にすぐプログラムができて近くを見つけるというのは、意味的にはまったく異なるところを指していることになります。

そうすると、コンピューター的な数字の列でなにかしようというと難しいので、意味的な近傍もしくは曖昧検索ともいいますが、それを実現するには集合に関する知識（辞書）が必要になってくるということになります。一番端的な例ですと、先ほどの red と RED は同じだよということをやりたければ、72 が入力されたら 52 にも変換して検索しなさいとか、65 が入力されたら 45 にも変換して出しなさいというテーブルを持つわけです。そうすると、大文字、小文字を区別なく検索することができるようになります。

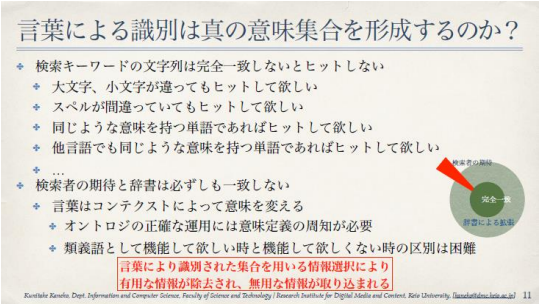
さらには、もっと高度な曖昧検索をしたくて、赤みがかった色を言うときに、赤と言うのか、ピンクと言うのか、朱色と言うのか、いろいろあると思うのですが、そういったものの広がりを検索で実現しようと思うと、このテーブルに全部書かないといけないことになります。0x72、0x65、0x64 は 0x70、0x69、0x6e、0x6b、これは pink ですが、変換してそれも探しなさいとか、同じように red が入ったら ruby となっていますが、0x72、0x75、0x62、0x79 で ruby の色を探しなさいということをやるわけです。

こういうテーブルを持ち始めると、なにが起こるかというと、いまは赤色を探したいから red と pink でいいかもしれないですが、red と pink を区別して調べたいときには、これは邪魔にしかならないわけです。もしくは、Red と ruby を区別して用いたいときには邪魔にしかならない。それが全部一緒くたになってテーブルに入ってくるので、こういったものはなかなか難しいということになってきます。大文字、小文字ぐらいはまだ、大文字、小文字を区別して検索するのはよくありますよね、これをやっているだけです。それ以上のものが入ってこないというのは、弊害も大きいからということになります。

こういう意味の広がりをどういうふうに対処するか、言葉のばらつきをどういうふうに対処するかというので、一つ出てくるのが語彙を統制する。Red というとこういう意味だということを定義するわけです。先ほどの原先生の話でも出てきましたが、たとえばオントロジーとして「作者」という意味があります。作者という意味が一意な ID として url がこれで指定されるのですが、先ほど「作者はいろいろな定義がありますから、熟読してくださいね」とさらっと流されましたが、まさに熟読して、それを理解して、それとたがわぬ使い方をしない限り、実はこういうものはうまく動かないということになります。

言葉による識別というのは、本当に真の意味集合を構築できるのかというのが、そもそも疑問に挙がってきます。利用者からすると、検索キーワードの文字列は完全に一致しなくても見つけてほしいという気持ちになると思います。僕だってなります。スペルミスをしたら直してよとか、大文字、小文字はもちろんのこと、同じような意味を持っているのだったら、それも探してくれたらいいじゃないかと思います。もしくは、日本語で出てこない情報だったら、英訳してくれてそれも探してくれたら幸せだよねと思うのですが、ベースはキーワードの文字列は完全一致でないと駄目ですよと。文字とその文字コードというところで全然違ってきますので、ちゃんと ID 化されるとそれが区別されてしまうという世界で、

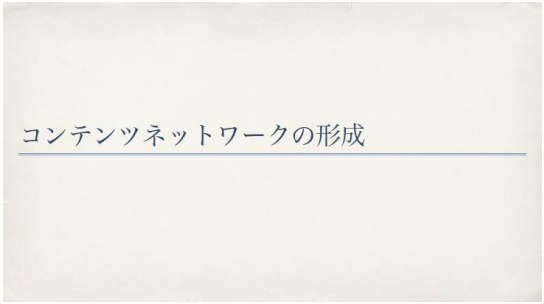
こういったミックスはなかなか難しいことになります。



スライド 11 に絵にしてみましたが、検索者の期待というものがあって、完全一致で出てくる部分、辞書によって先ほどのマッピングのテーブルで出てくる部分、たぶんユーザーが期待している部分。イメージ、すごく期待している部分のほうが辞書による拡張よりももっと先に行っていて、辞書によって、拡張によって得られたのは少しというところが現実的なところなのかなと思います。

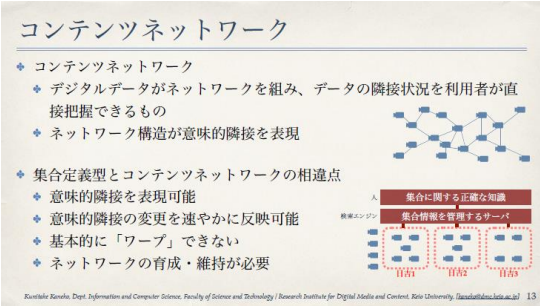
それは、そもそも言葉はコンテクストによって意味が変わってきますし、オントロジーの正確な運用には意味の定義の周知が必要であり、類義語として定義するとしても、それを機能させたいとき、させたくないときのコントロールが非常に難しいです。コンピューターは人間の状況を把握できないので、それをつかむことは非常に困難になるかと思います。結果として、言葉で集合を定義するとなにが起こるかというと、有用な情報が除去されて、無用な情報が取り込まれるというのが言葉による識別の難しいところなのかなと考えています。

ここまでのところで二つお話をしました。情報を選択してくるというときに、集合を定義するとフレキシビリティがないという話と、その集合が言葉によって定義された場合、そこに曖昧性という言葉のコンテクストへの依存性があって、うまく機能しないということになります。



さて一方で、もう一つ対抗馬として出したネットワーク型についてお話ししたいと思います。コンテンツネットワークの形成と書いていますが、ネットワーク型の話です。

いろいろ出てきますが、コンテンツネットワークとはなにかというと、デジタルデータがネットワークを組むと。データの隣接状況を利用者が直接把握できると書いていますが、言ってみれば先ほどの『ターミネーター』のカメラとカメラが相互につながっているイメージ。カメラが情報に置き換わっていくと思っていただけるといいかなと思います。



ある情報がある情報とつながっている、この情報はより自分に近いような位置にいる、もしくは遠い位置にいる、そういったものがこのグラフ情報によって分かってくると。ホップが離れていくと遠い、ホップが近いということになります。すなわち、ネットワーク構造というのはグラフ構造と思っていただいてもいいですが、意味的隣接を表現可能だというのが一つ大きな特徴になっています。

このネットワーク型を集合定義型の先ほどの例と比較していくと、いくつか違いがあります。先ほど言った意味的隣接を表現可能だということが一つと、意味的隣接が少し変わったら、その部分だけ変えてあげれば、それを使う人はそのときになって反映された状況を使えますということ。

あとは、一方でワープができないです。ワープができるというのはどういうことかというと、たとえば日吉１丁目でグループ化される、集合として定義されている、このカメラはクリックで切り替えられるような集合だと思うのですが、それはこのネットワーク的にはここととここに対応しているかもしれない、ここととここになっているかもしれない。でも、これはネットワークがつながっていないので、直接にはそこには行けないということになります。

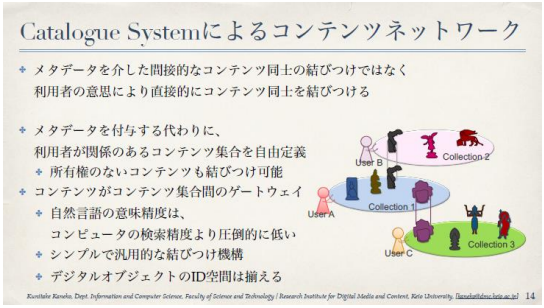
ネットワーク構造を取った場合は、そのネットワークの上でしか通っていくことができないので、離れたところにジャンプはできないことになります。でも、頑張って誰かたどった人がいて、これとこれが実際近いことが分かったら、ここに線を引いてあげるだけで、そのあとの人はワープできるというのがネットワーク型の特徴になります。

こうやって話をしていると、ネットワーク型はイケてるよねと思っていただけたら僕はうれしいのですが、一番難しい点があります。コンセプトは非常にシンプルで、近いものをグラフで表していくというすごくシンプルな構造なのですが、大変なのがスライド 13 の下に書いた、ネットワークの育成と維持が必要になります。なぜ難しいかというと、ネットワークにはメカトーフの法則というものがあります。これはネットワークの価値が n の 2 乗に比例するというものです。

一番シンプルな例でいきますと、いまどきは携帯電話だと思いますが、携帯でも携帯ではなくても電話があるとしたときに、世界中に 1 台しか電話がなくて、それを所有していることに意味はまったくないわけです。誰にもかからないし、意味のないものです。2 人だけが持っていたら、その 2 人だけがコミュニケーションできるので、その 2 人の関係が親密であれば非常に有効に使えます。でも、それだけです。ホットラインみたいなものです。3 人になってくると、3 人がそれぞれ話せるようになってくるので、2 人だけしか持っていないときよりも便利だよねと。4 人、5 人、6 人、7 人とどんどん増えていけばいくほど、その価値は上がっていくということになります。

ですから､ネットワークサービス全般に、たとえば Facebook もそうですが、あれは 5 人でやっていても全然つまらないと思うんです。でも、あれだけのスケールで動いているから新しい情報がやってきておもしろいよね、みんなやっているなら僕もやろうとなるわけです。これがネットワーク技術です。

インターネットも同じです。3 台のパソコンがインターネットにつながっています。3 台だけでインターネットが構成されている世界。「べつにそれは要らないのでは」とみんな思うと思うんです。でも、世界中のネットワーク、世界中のコンピューターが同じネットワークにつながっているからこそ、インターネットをつないでおかなきゃとなるわけです。これが先ほどのメカトーフの法則という考え方で、そこのしきい値を超えるまでが、実はネットワーク型システムの一番難しいところになります。



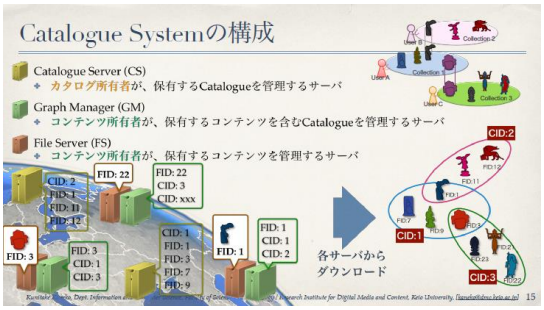
スライド 14、うちの DMC でやっている研究の紹介をさせてください。うちのネットワーク型のシステムです。うちはネットワーク型でシステムをつくるというチャレンジをしています。一般的なメタデータで物事を見つけてくる検索のようなやり方、それはメタデータを介した先ほどの言葉で表されたつながりです。同じ ID を持っている、共有しているという間接的なコンテンツ同士の結び付けではなくて、利用者の意思により直接的にコンテンツ同士を結び

付けると。

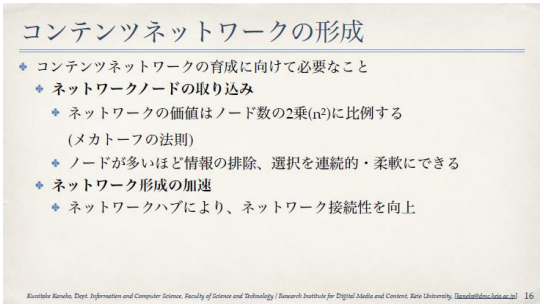
たとえばAさんがこのコンテンツとこのコンテンツは関係あるよねと思ったら、ここで1個決めてあげる、結び付ける、線を引いていると思っていただいても構いません。メタデータを付与する代わりに利用者が自由に定義してあげましょうと。言語化せずにただ単に集合を与えるだけ、もしくはリンクをつないであげるというのが基本的なシステムの考え方です。

これはあとで出てきますが、所有権、自分が持っていなくてもこのリンクをつくれるというので、実はネットワーク育成を促進させようとしています。コンテンツが、たとえばこの黒いオブジェクトを見ていただくと、User B も User A も Collection にそれぞれ入れている。このコンテンツ同士が自動的につながるメカニズムを Catalogue System は持っています。単独で Collection だけをつくっていたらネットワークは分離した世界なので、まったく使い勝手が悪いのですが、同じコンテンツを持つと、そこが自動的につながるメカニズムを導入することで、ネットワークが構成されるようになっているというのが Catalogue System です。シンプルで簡単に結び付けられるというのをキーにした技術になっています。

具体的な方法を話しても、頭がこんがらがるといけないので簡単に説明しますが、先ほどの上の User A、B、C がそれぞれ自分で関係をつけたグループをつくった、もしくはネットワークをつくった。それぞれが LAN をつくると思っていただいても構いません。



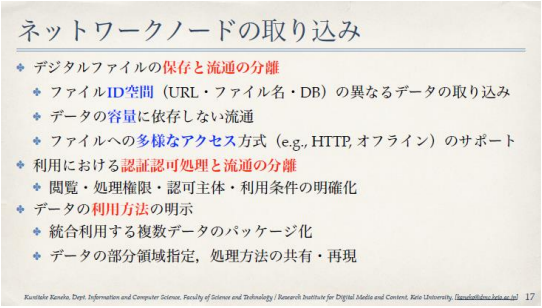
それが実際のデータとしては、左のような形で入って、CID:1、CID:2、CID:3 のような形でそれぞれネットワークが組まれることになっているのが Catalogue System の構成になっています。詳しい話は技術展示のところで聞いていただければ、説明してもらえると思います。



さて、コンテンツネットワークをつくっていこうとしたときに、難しいことが二つあります。一つはノードの数をいかに増やしていくかという話と、その増やしたノードの間にネットワークをどう組ませるかというところが、コンテンツネットワーク形

成においては非常に重要なポイントになってきます。

ネットワークノードが増えるほうがいいというのは、先ほど話したことです。プラス、ノードが多くなってくると、情報の粒度、どのぐらいまでの情報が欲しいかというバリエーションがたくさん出てきますので、ユーザーに応じて適切な情報を選択することができるという意味でも、もちろんノードが増えれば増えるほどいいかなということになります。ネットワーク形成の加速ですが、ハブをあちこちに置こうね、そうするとネットワークにつなぎやすくなるでしょうというのが、簡単な理解になります。



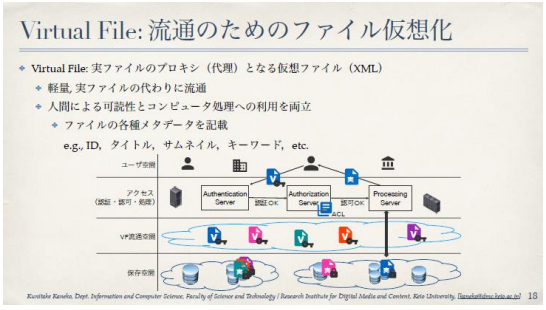
まず、ネットワークノードの取り込みですが、どうやったらこの関係のあるネットワークにノードを持ってこられるか、どんどんファイルを増やしてこられるかというのをいろいろ考えました。それが実際の設計としてこういうふうに落ちてくるのですが、まず保存されているファイルをそのまま流通させることはやめようと。保存する話と流通する話、そういったファイルがあるという話は完全に切り離しましょうというのが一つ目の話です。

なぜかというと、あるファイルは URL が HTTP でアクセスできるところにあるけれども、あるファイルはオンラインにないかもしれない。オンラインに上げておくのは非常に管理の手間がかかるので、そういうのはないかもしれない。もしくは、それがデータベースに入っているかもしれない。それを透過にアクセスしないといけないとなると、全部オンラインに上げてくださいとか、ものすごくシステムの負荷が上がってきます。こういった別々の場所に保存されているものを、きちんとハンドリングしたいので、まず切り分けましょうというのが一つ目。

二つ目が、最近だとデジカメやスマホで撮っても軽く 5MB とかいくような時代に、5MB、10MB のファイルをスパンスパンとやりとりして幸せかというと、スパンスパン行けばいいですが、行かないので不快なわけです。そうすると、データの容量に依存しないような形でやるためにも保存と流通は分離しましょうと。あとは、先ほども言いましたが、HTTP、オフラインのサポート等いろいろあるのを、どんな形でもできるようにするためにも、流通にはそういうエンティティは要るけれども、実際のファイルはどこかに保存されていてもいいみたいなことを言っています。

あとは、ファイルを使うときに課金したいとか、誰だったら使っていいみたいな話が出てきます。そういったところもファイルがあるよとか、このファイルは便利だよという話とは別のところで管理しましょうと。お金を取る、誰が見られないようなのと、実際にそういうのがあって便利だよという話は別だということです。

あとは、最後にデータをどう利用できるのか、先ほども少し出てきましたが、こういうフォーマットで書かれていて、どういう情報が含まれているか分からないと使えないので、そういう利用方法を明示したり、ある誰かがつくったコンテンツの一部だけを流通させたかったら、そういう部分指定もできるようにしたりするというのが、ネットワークノードの取り込みの話です。



いまのコンセプトを全部突っ込んでつくっているのが、スライド 18 の Virtual File という話で、あとでここで技術展示のポスターを学生さんがやっていますので見てください。保存の話、Virtual File といいますが、その流通の話、認証の話、実際使う話、先ほどの Catalogue System を使う話はもっと上にいるのですが、それを切り離して動かすようにしましょうと。

流通したときに、これは一種のチラシみたいなものです。実際のサービスではなくてチラシみたいなもので、チラシにどんな情報を載せておけばアトラクティブなのかみたいなことは考えないといけないので、ここにメタデータがいろいろ載ってくるのかなと考えています。たとえば、タイトルがなにとか、サムネイルも載っていてとか、最低限の流通情報が Virtual File には記載されるのかなと考えています。

もう一つ、ネットワーク形成の加速ですが、ただノードがネットワーク上に存在するだけでは、ネットワークの価値は上がってこなくて、ネットワーク化される、要するにどんどんいろいろなところとつながってくることによって、あちこちに行けるようになるわけです。

まず、システムとして一番肝というか重要視しているのが、デジタルコンテンツを所有していなくてもネットワークの線を引けるということです。インターネットの世界で極端な話、たとえば僕は CNN をよく見ますと。CNN を見るためには日米間の回線が太くないといけないといった場合、日米間の回線を太くするみたいなことを自分が思っていないにもかかわらずやるということです。それができることで、ネットワークがどんどん出来上がってくるという

ことです。日米間のと言っても、日米間の既存のラインの横に引くのではなくて、ショートカットで行けるようなものをつくってしまうということです。

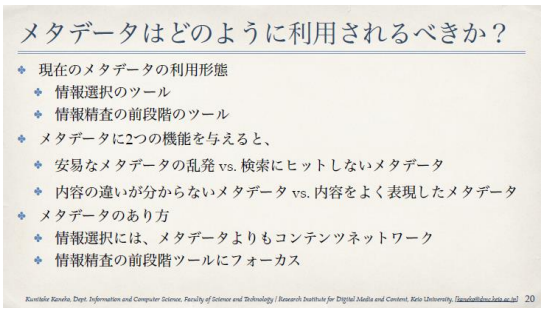
もう一つ、別のアイデアでネットワーク形成を加速しようとしているのですが、それがハブを用意するという考え方です。三つのやり方をスライド 19 に書いています。モノハブ、位置ハブ、イベントハブです。

モノハブはその名のとおり、同じモノに対して、たとえばこれですと、「伊右衛門 500ml ペットボトル」というモノを、違うデジタルファイルに、いま四つになっています。同じモノを三つのデジタルファイルが共有していることになります。あるときはこれを A さんがつくり、あるときは B さんがつくり、それは全然別の場所で行われるかもしれないけど、工業品なのでなかなか難しいですが、違う場所でつくられて、違う人によってつくられたかもしれないけれども、こういったモノがハブ機能になるというのは、ネットワークを形成するのに便利かなというのが一つ目です。

二つ目が、位置を軸にいろいろなものがつながってくると。たとえば、慶應だと日吉の記念館をいま建て替え工事中です。下は古いやつ、上は新しいやつ、これは想像ですが、それがこの場所でリンクしてくると、別の人が同じ場所で体験したものに飛んでいけることになります。

最後のイベントハブというのが時間軸です。たとえば、去年の DMC シンポジウムでこんなのがあったよと、Twitter のハッシュタグみたいなものですが、違う人がそれぞれ撮った写真であっても、それが同じイベントでそこをハブにいろいろ広がっていけるというような工夫をすることで、ネットワーク形成が加速されるかなと考えております。

さて、いままで集合定義型の話、集合定義型が言語、言葉で表現されたときの課題、一方でネットワーク型、ネットワーク型の課題とその解決の仕方をお話ししてまいりました。ここで改めてメタデータはどのように利用されるべきではないかというところに立ち返りたいと思います。



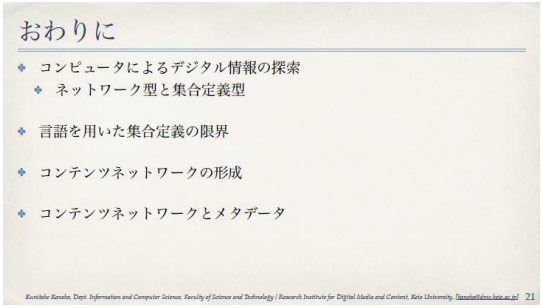
言うまでもなく、集合定義型、言語による集合定義というのがいまのキーワード検

索であり、ほぼほぼいまのメタデータ検索のやり方になっていると思います。現在のメタデータの利用形態は、先ほどの情報探索の手順の 1 番と 2 番を両方やっているのではないかと思います。すなわち、情報選択のツールであり、たくさんの情報リソースの中からそれを選んでくる、1 億個から 100 個を選んでくるための道具であり、そしてその 100 個の中から、いったいなにが書かれているのか中を見極める、でも本文までは読みたくないみたいなときに使われる情報精査の前段階のツール、この二つが混在しているのがいまのメタデータの利用のされ方かなと思っています。

しかしながら、この二つの機能をメタデータに同時に与えてしまうと、なにが起こるかというと、情報選択のツールだと思うと安易なメタデータを乱発し始めると。引っ掛かってくれなかったら困るといって、なんでもかんでも、富士山が小さくでも写っていたら富士山、富士山と入れるということです。でも一方で、情報精査の前段階のツールだと思って入れたキーワードは、すごく精緻なキーワードであるかもしれないけど、決してほかの人が思い付かないみたいな話、これは相反する、対立するような軸かと思います。

もう一つが、メタデータは与えたけど、こちらの写真に与えられたメタデータとこちらの写真に与えられたメタデータは全く同じだよね、それは本当に中身を表しているのかということです。一方で、きちんとその違いが分かるようなメタデータ、そういったところがこの二つの混在によってミックスしてくることになります。

僕が思うに、情報選択というのはメタデータを使うのではなくて、コンテンツネットワークを使うべき、そこで情報の選択をして、情報の選択をしたあとにより精査するときに、そのプロセスを簡略化する、もしくは簡便にするためにメタデータがあればいいのかなと思っているという状況です。



最後に「おわりに」ですが、コンピューターによるデジタル情報の探索はどんなシステムでできるのかな、どんな手順になるのかなというところからスタートして、メタデータ、既存のキーワード検索の限界というものを伝えたつもりです。一方で、ネットワーク型にすることで、柔軟、フレキシブルでいろいろなユーザーの状況に応じた探索をすることができるというのがネットワーク型で、ネットワークの形成の仕方、そして最後にメタデータが現状とどのような関係であるべきかというところをお話し

いたしました。

金子 晋丈（かねこ くにたけ）

慶應義塾大学理工学部専任講師・DMC 研究センター研究員。専門はアプリケーション指向ネットワーキング。特に、デジタルデータの利活用を促すデジタルデータのネットワーク化について研究を行っている。2001 年東京大学卒業。2006 年同大学院情報理工学系研究科博士課程終了、博士（情報理?学)。同大学院新領域創成科学研究科での特任助教を経て、2006 年 9 月より慶應義塾大学デジタルメディア・コンテンツ総合研究機構、特別研究助教。2007 年、同機構特別研究講師。2012 年 4 月より現職、デジタルメディア・コンテンツ統合研究センター研究員を兼任。

話題提供

「書誌学者の立場から見たボーダレスなデータ利活用のための情報組織化」

安形 麻理

（慶應義塾大学文学部准教授）

ご紹介いただきました安形と申します。メタデータということでしたら、うちの専攻のもっとふさわしい先生も会場にいらしているのですが、私は本日は書誌学者としての立場からということで、先ほど冒頭にありました原先生のお話とだいぶ重なるところもありますが、研究に関わってきた中でのメタデータについて思うところを話題提供という形でお話をさせていただきます。

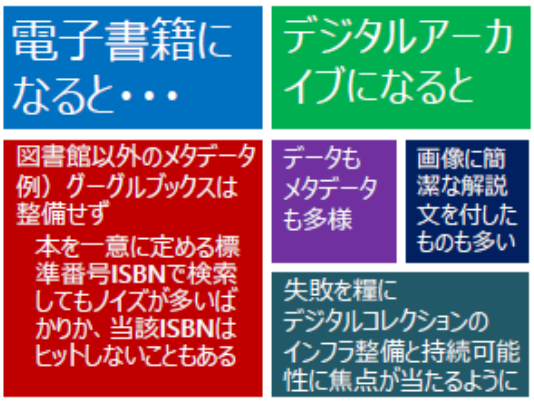
最初に図書館におけるメタデータを考えみると、人類の知的な文化財を扱う機関としては、原先生のお話にもありましたように、非常に早くから標準化が進んできた分野であるといえます。基本的には近現代の刊行物、複製品がたくさんある本を扱ってきたため、目録の標準化が非常にしやすかったということもあって進んできたわけです。

最近はインターネット情報資源など、図書館が物として所有していない情報資源についても記述しなくてはいけないという状況に対応するために、また一方では博物館や文書館とも、デジタルなデータであれば、お互いのメタデータの共有が進むだろうという期待から、ダブリン・コアといった共通の枠組が出てきているわけです。

ただし、一般のいわゆる複製物としての印刷本でしたら何がメタデータかは明確ですが、私が研究対象にしているような 15 世紀の印刷物（incunabula、インキュナブラ）、あるいは会場後方にも 15 世紀の活版印刷物やその前に作られた手書きの写本の

リーフが展示されていますけれど、ああいう資料についてのメタデータをとる際には、メタデータなのかデータなのかという境界はかなり曖昧だといえます。つまり、その物が何であるかということを同定して記述すること自体が研究であるという意味で、メタデータとデータの境界が必ずしもはっきりしないところがあるかと思います。ある場合にはメタデータ、ある場合にはデータともなりうるわけです。

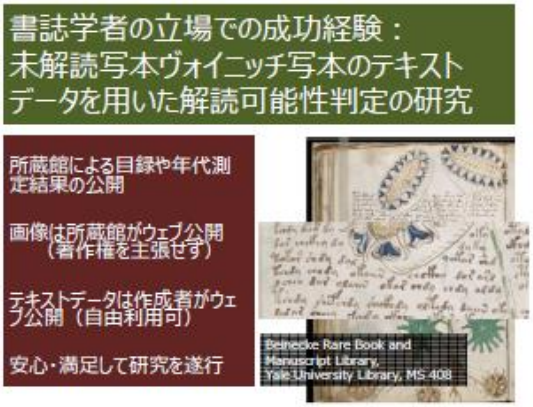
欧米の国立図書館などでは、書誌データを Linked Open Data（LOD）の形で提供するということが行われていて、私も昨年から共同研究者と一緒にインキュナブラの LOD モデルを考えてみたらどうだろうということで取り組んでいます。そうする中でやはりメタデータをどうとるかというのが非常に重要な課題であると認識しました。そちらはまだ検討している途中ですので、今日はちょっと違ったお話をしたいと思います。



目録は非常に標準化が進んできたと申し上げたのですが、本が電子書籍化されたり、デジタル化された場合になると、図書館以外のアクターもいろいろ関わってきます。たとえばグーグルブックスなどでもたくさんの本がデジタル化された形で手に入るわけですが、実はグーグルブックスでは、紙の本を一意に定めるはずの ISBN がそのデジタル版ときちんと対応付けられておりません。検索で一つの ISBN を入れても非常にたくさんのリストが出てきます。なおかつ、その中には該当する ISBN が 1 件も含まれていないということもあります。グーグルブックスではそういった形でのメタデータを整備しようという意識はないということが見てとれます。これは先月行われました第 66 回日本図書館情報学会研究大会の「米国における電子書籍化の現状」というポスター発表で知ったことなのですが、そういった状況です。

また、デジタルアーカイブも様々な機関が様々な方針の下で、それぞれのメタデータを付けているわけです。データそのものも、そこについているメタデータも多様で、目録と違い標準化は進んでいません。メタ

データとして書誌・所蔵情報やフォリオ(ページ）番号、撮影の詳細などを付したものもあれば、本一冊分の画像に簡潔な解説文を付してメタデータにあたるものとして提供されているものも多く見られます。



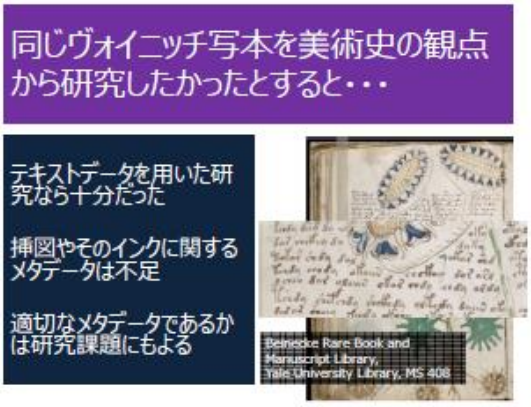
ここで、メタデータと研究の関わりの例を紹介します。スライド 4 に挙げましたのは、ヴォイニッチ写本という、発見から 100 年程経っているのにまだ誰も読むことができていない、オカルト界隈でも大人気の謎の写本です。読めそうなのに読めない文字で書かれているのです。私と共同研究者がこの写本を真面目に研究してみようと取り組んだ 10 年程前の時点では、16 世紀の写本もしくは 20 世紀の偽物というようなメタデータがついておりましたが、その後年代測定が行われて 15 世紀ということは間違いないということになり、メタデータも変わったりしています。

我々はテキストデータをページごとにクラスタリングした結果と、ヴォイニッチ写本はほとんどのページに挿絵がついているので、その挿絵の種類が一致するかどうかを見ることで、文書が一貫した構造をもち内容に意味がありそうかどうかを判定しようという研究を行いました。

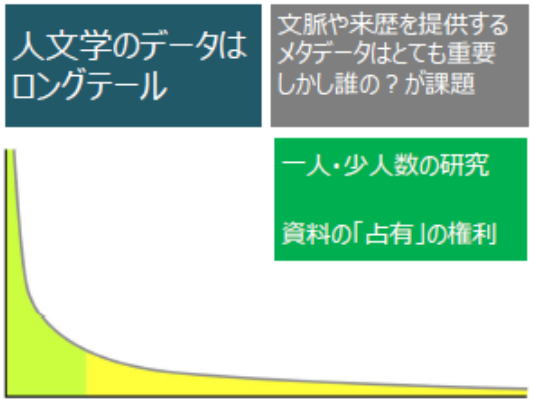
この研究では、比較的詳しい目録データが所蔵館であるイェール大学のバイネッケ図書館の OPAC から入手でき、画像は所蔵館のウェブサイトから公開されていて、なおかつそこにはもう著作権保護期間が満了している資料なので、イェール大学は何の権利も主張しないし、許諾も拒否もしないということが明示されています。権利情報に関するメタデータもきちんと示されています。なおかつ翻刻されたテキストデータも、個人的に翻刻された研究者がウェブで自由利用可として公開されていました。

この写本については、権利関係も非常にクリアということで、安心してスムーズに研究を行うことができました。やはり研究者が知財の法律の専門家である必要があってはならないと思うわけですが、そういった意味でのメタデータがわかりやすく整備されていました。

しかし、この同じヴォイニッチ写本をたとえば美術の観点から研究したかったとしますと、ヴォイニッチ写本の各ページに示されている挿図やそこに使われている顔料、インクといったものに関するメタデータはありません。

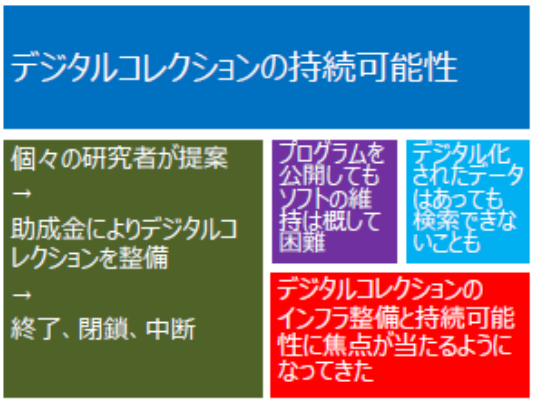


本日の会場の展示コーナーにはたくさんの顔料が並んでいるわけですが、ああいった情報は付与されていません。同じデータとメタデータのセットであっても、研究者が持っている課題によっては適切な十分なメタデータであったり、そうでなかったりするといえるでしょう。



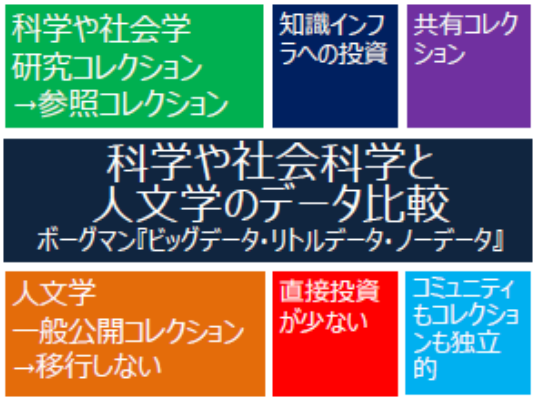
そういう意味でいいますと、人文学のデータ自体がいわゆるロングテールにあたるかと思います。宇宙科学やタンパク質の構造といったビッグデータと関連が深い研究領域もあるわけですが、人文学の研究というのは、しばしば１人、あるいは少人数で行われたり、あるいはその資料を持っている、アクセスできるということ自体が価値であったりすることが珍しくなかったわけです。

それぞれの研究者が様々な課題をもって研究を遂行していくため、そこで必要とされるメタデータが非常に独立的で多様であるというところが一つの特徴であるといえます。



また先ほど原先生が最後の方でおっしゃっていらしたように、いろいろなデジタルコレクション、特に人文学のデジタルコレクションを公開する際には、どちらかというと研究者の側でそのデジタルコレクションの構築がいかに大事かという提案書を書き、助成金を獲得して実現させる形が主流です。待っていてもお金は下りてきません。助成金でデジタルコレクションが整備されても、それがやがて中断、閉鎖、あるいは終了してしまうと、大規模なプロジェクトであっても、データだけは残っているがもはや検索はできないなど、どう使ってよいのか分からないようなデータというものが残されてしまっているわけです。

あるいはプログラムは公開されているものの、そのプログラムを実際にデータに適用し動かしていくようなところでの維持は非常に困難だということも出てきます。欧米の大規模なプロジェクトでもそうした例が多く出てきていることを受けて、最近では、助成機関側がデジタルコレクションのインフラ整備と持続可能性にも焦点を当てるようになってきたというのは、明るい兆候であるといえるでしょう。



最後に、人文学の特徴として『ビッグデータ・リトルデータ・ノーデータ』(勁草書房, 2017) という本にあった比較をスライドに挙げました。これは全体としてビッグデータにしろ、人文学のようなリトルデータにしろ、きちんとしたメタデータやそれを提供する知識インフラが整備されていないかぎりノーデータになってしまう、つまり、使えない、アクセスできない、誰も知らないデータになってしまうということを述べているものです。

人文学のコレクションは、個別の課題、個別の研究課題というものに密接に対応するために、なかなかそのコミュニティ全体で参照して使うような参照コレクションに移行しにくいという特徴があります。直接投資が少なく、個別の研究課題があるからこそです。そうしたことを考えた場合に、人文学のメタデータのあり方を一口にまとめるのは難しいといえます。本当に話題提供という形で、結論も出ていないのですが、私の話は以上です。どうもありがとうございました。

安形 麻理（あがた まり）

慶應義塾大学文学部教授（書誌学、デジタル書物学、図書館・情報学）。学部生時代から貴重書デジタル化プロジェクトに様々な立場で関わってきた。主著に『デジタル書物学事始め：グーテンベルク聖書とその周辺』(勉誠出版, 2010 年)、「利用者による特殊コレクション資料の撮影の許可：北米の研究図書館における動向」『図書館は市民と本・情報をむすぶ』(勁草書房, 2015, p.

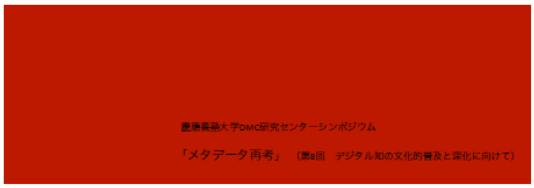
52-60)、「科学は贋作を判定できるか」『書物學』no. 14, 2018, p. 34-41.等がある。

話題提供

「パースペクティヴとメタデータ」

久保 仁志

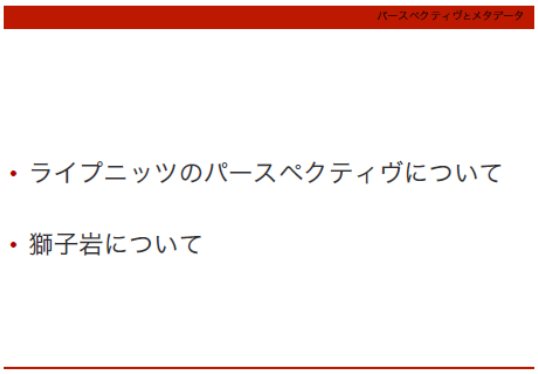
（慶應義塾大学アート・センター所員）



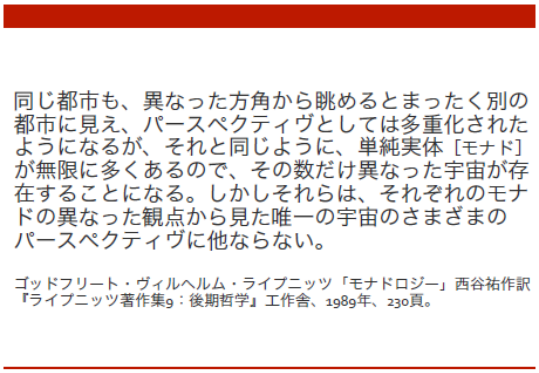
本日は普段自分が関わっている知の体系とはまた違うところに迷い込んでしまった感が実はありまして、いまからする話が場違いなことにならないか若干不安でいます。

最初にお話をいただいた「メタデータ再考」という一番の主題がある。そして「ボーダレスなメタデータの利活用のために」という副題があるというときに、そもそも「メタデータ」とは何かを一つのモデルを通して提示できないかということと、ボーダーというのが、どこにどういうふうに存在するのかということを示せないかと考え、本日の発表をいたします。

これからお話しする内容は、原先生がペットボトルのお茶を使ってご説明されたことをかなりクロースアップして、末端肥大的にご紹介するような形になると思います。



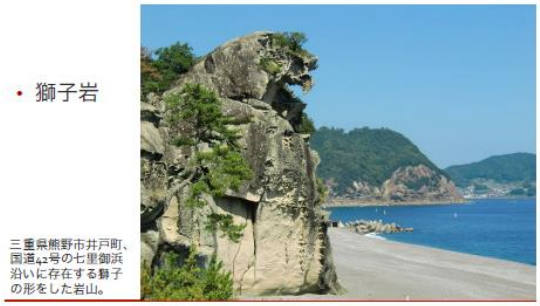
ライプニッツのパースペクティヴという概念を説明したあと、「獅子岩」と呼ばれている岩を通してパースペクティヴがどういうものなのかということをお話しさせていただければと思います。



まず簡単にライプニッツの引用をしてしまいます。「同じ都市も、異なった方角から眺めるとまったく別の都市に見え、パースペクティヴとしては多重化されたようになるが、それと同じように、単純実体［モナド］が無限に多くあるので、その数だけ異なった宇宙が存在することになる。しかしそれらは、それぞれのモナドの異なった観点から見た唯一の宇宙のさまざまのパースペクティヴに他ならない」。

いまからすると「そもそも宇宙一つです

か」という問いがあるでしょうし、「多様体」という考え方があるので、「唯一の宇宙」という言葉に引っかかる方もいるかもしれませんが、ものとしてそこにあるように見える、みんなが一つとして振る舞っているものについて、これは語っているわけです。また、ライプニッツにとっては理念的な「唯一の宇宙」でもあるわけですが。



この「パースペクティヴ」について獅子岩を通して考えてみたいということです。

三重県熊野市井戸町というところに七里御浜という浜があります。スライド4の写真はそこにある岩ですが、皆さん獅子に見えますかね。人によっては獅子ではなくて、鳥のくちばしのようにも見えるという方もいたり、学生に聞いてみたらパンダとかクマとかいう人たちもいたりして、人によっておそらく見え方は違うと思うのですが。

一応、念のために説明しますと、これが（獅子の横顔とすると岩山の右上の小さな凸部を）鼻、かつ口ですね。これ（その下）はおそらくひげですね。または、これ（さらに下の大きな切れ目）を口に見立てて、（前述のひげ部分を）牙というふうに見ることもできます。

ただ、これは別にこういうふうに見ることが厳密な獅子岩ですと定義されているわけではないので、たぶんいまお話した二通りの見方で見たとしても、おそらくどちらも正解だと考えられます。

そして実はこの獅子岩は狛犬を兼ねているんです。

獅子岩がある場所というのは、井戸町という、スライド5の地図の中央部あたりなのですが、そこから結構離れた場所に大馬神社と呼ばれる社がありまして、実はこの大馬神社に狛犬は置かれていないらしいのです。神社の狛犬の代わりに、この獅子岩が用いられ、信仰の対象になっているという

ことなんです。

　では実際、この獅子岩がどこから見ても獅子岩に見えるのかを検証してみましょう。

スライド６の写真は獅子岩周辺の写真をグーグルストリートビューからクリッピングしてきたものです。例えば、この地点（右上段写真）に立つと「獅子岩」には見えませんね。これをお城に見立てることもできるかもしれないですが、ただの崖に見えます。この見えの位置が「獅子岩」に対してこの位置（地図該当箇所）ですね。

　初めに簡単に説明してしまいますが、この同じ方位というのかな。もう少し遠ざかって見る（右中段写真）と、実は獅子岩っぽく見えなくもないんですね。おそらく人間が実際にこの浜に降り立ってこの岩を見たときに見える姿としては、たぶんこれが一番獅子岩に見える位置だと思うのですが、先ほどスライド４でお見せした写真よりも獅子性の度合いが下がっていますよね。

　これを見ると、やはりクマとかパンダというのはあながちうそではないというように感じられます。たぶんいまご覧いただいている皆さんそれぞれによっても見え方はかなり違うと思います。しかし、おそらく獅子を見いだそうとする中で、なんとなく獅子らしさみたいなものが見えてくる。

　もうこれ以上近づいてしまって見上げる（スライド６の右上段写真）と、その見えは、パースペクティヴは崩れてしまう。

　ほかにもいくつかあるのですが、獅子岩のように見える場所の反対側にちょうど回ってみたとき。ちなみに日本のスフィンクスなどと言われることもあるらしいのですが、これ（左上段写真）はおそらく足を折り曲げてお尻の部分を見せている獅子岩というふうに見えないこともない（笑）。だけど､そういうふうに見ることができるのは、この岩の塊を一度「獅子岩」として見た後ですよね。

　そして、これ（左中段写真）はいわゆる顔の正面ですね。横顔が見えるなら正面から見ても獅子岩かというと､そうではない。のっぺりして、全然獅子岩に見えない。ただの岩の壁、岩肌しか見えない。

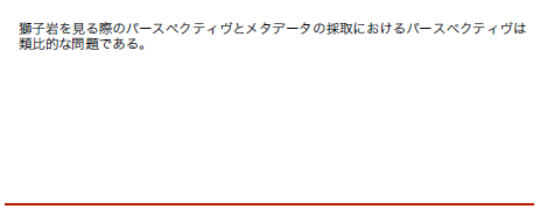
　では、真裏に回ってみるとどうなるかと

いうと、ただの盛り上がった藪です。たぶん獅子岩と知らずにここ（左下段写真のあたり）を車で通り過ぎてしまう人は相当多いと思うんですよね。この見えしかないので。

ただ奇跡的に車から獅子岩に見えるポイントというのが、実はありまして。ここなんですね（右下段写真。地図、道路上の南地点）。道路を北から南に走って来た人にとっては、知らなければ絶対に獅子岩と気づかないのですが、逆に南から北に走って来た人にとっては獅子岩と知らなくても、「ひょっとして、あれ、なんか動物っぽくない？」というような、気づきがある可能性があるパースペクティヴです。これは最初にお見せした写真の位置で、獅子岩と言われて一番納得できる位置です。

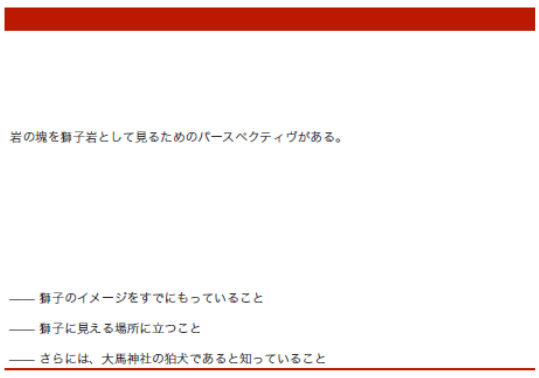
いまお話ししたのは獅子岩をどこから見るのかというアングル、空間的なパースペクティヴの問題なのですが、実は時間的なパースペクティヴにおいても、このギャップというものがありまして、スライド 7 を見て下さい。

左の写真のほうはなんとなく獅子に見えるのですが、右の写真を見ると獅子が消えてしまう。一度獅子岩に見えてしまうと不可逆的にこの写真でも獅子岩を見出してしまうかもしれませんが、この光景をはじめに見て獅子岩と見ることは難しいと思います。ここから分かるのは獅子に見えるためにはシルエットだけが獅子に見えることが条件ではないということですね。灰色だったり、光が当たっていたり、暗かったりとか、その凹凸のイメージも含めて実は獅子性というのが立ち現れている。右は夕暮れどきですが、肌理が消え、獅子らしさがなくなっているといえると思うのです。だから時間的なパースペクティヴにおいても、この獅子岩が現れるポイントがあることになる。



「獅子岩を見る際のパースペクティヴとメタデータの採取におけるパースペクティヴというのは類比的な問題」があるのではないかと思っています。別の言い方をすると、メタデータの項目というのは、まさにこのようなパースペクティヴの設定そのものなのではないかなと思います。また、メタデータの項目というのは、どこからそれを見るのか、いつそれを見るのかということ。もう少し具体的な言い方をすると、名称は何なのか、大きさはどれくらいなのか。そういう問いがメタデータの項目となりますが、その設定によってかなり現れてくるものが違ってくる。つまり、パースペクティヴやメタデータの項目とは、どのように対象にアプローチするのかという問いのことです。

もう一度、ただの岩の塊を獅子岩として見るためのパースペクティヴの条件が何であるのかと考えると、まずは獅子というものを知っていること、獅子のイメージをすでにもっていること。



これがある対象を獅子と同定するための前提条件です。次に獅子に見える場所に、見える時間に立つこと。さらには、それが単に獅子岩に見えるというだけではなく、ある種の信仰の対象としても見えるということです。(大馬神社には狛犬がなくて、この獅子岩は狛犬としての機能も持っている)。

では、この条件のうちの一つを客観的に規定できるのかを考えてみましょう。例えば、空間的座標を定めたときに、この獅子岩というものが明らかに獅子岩として立ち現れてくるのか。緯度経度を指定すれば少なくとも立つべき位置は分かる。しかし、それも実は測地系によって変わってしまいます。

震災のあと国土地理院は 2011 年に測地系を更新したようで、それに従った座標を発表しているのですが、ほかのウェブで検索できる地図は実は古い測地系に従っています（2018 年時点)。日本が今まで用いいてきたのはおよそ 3 種類ありまして、2011 年以降の「日本測地系 2011」、2002 年に採用された「世界測地系」、それ以前の「日本測地系」です。それで全部数値が変わってしまうんですね。だから、緯度経度が分かっていれば獅子岩に見えるような位置に立てるか、そのパースペクティヴに立てるかというと、実はそうではない。ある測地系に則った緯度経度が指定され、高度も、見る角度も、時間も分かってないといけないということになる。さらには、それらの誤

差や、各人の個体差の修正などのプロセスを経てようやく獅子岩のパースペクティヴが現れてくることになるはずです。

また、ボーダーの話ですがぼくは、パースペクティヴの設定の間にそれが存在すると考えています。

結論は別に特にないんですが、こういうパースペクティヴの問題というのが、メタデータを考えるときに一つ大きなきっかけになるのではないかなと思い、本日は発表させていただきました。

久保 仁志

慶應義塾大学アート・センター所員、同大学非常勤講師。アーキヴィストとして 2002 年より現在に至るまで様々な資料体の構築に携わる一方、美術理論・アーカイヴ理論・映画の研究を行い、映画作品《Cargo 1　なにかいってくれ いまさがす | 半影のモンタージュ》(2010 年) 等を制作。近著に『〈半影〉のモンタージュ：アーカイヴの一つのモチーフについて』(「JSPS 科研費 26580029」レポート、2017 年)、「ある書斎の事件記録??瀧口修造と実験室について」『NACT Review 国立新美術館研究紀要』5 号（国立新美術館、2018 年）等がある。現在、慶應義塾大学アート・センターにてアーカイヴを考えるための企画『プリーツ・マシーン』を行っている。

話題提供

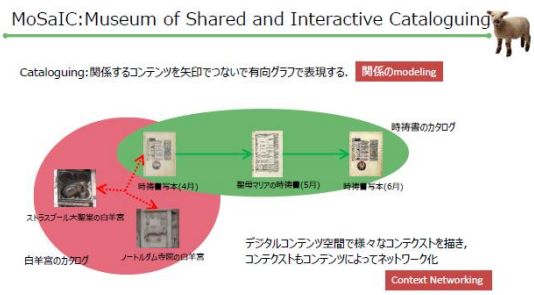
「「MoSaIC」におけるデジタルコンテンツの情報組織化」

石川 尋代

（慶應義塾大学 DMC 研究センター特任講師）

MoSaIC というシステムに関しての話をさせていただきたいと思います。今回のテーマがメタデータということだったんですが、私は文化財を持っているわけでもなく、それをデジタル化してデータベースをつくるわけでもなく、メタデータをつくったこともなければ、それをつけたこともないエンジニアリングのほうにずっといた者なので、なにをどういっていいのか分からなかったのですが、ここで MoSaIC では情報の組織化ということをやってきたと思うので、きょうはそれについてお話をしたいと思います。

この MoSaIC（Museum of Shared and Interactive Cataloguing）というのは略語でありまして、最後に Cataloguing ということが出てきますが、この Cataloguing が先ほど金子先生の話に出てきた Catalogue System にのっかるデータのことになります。これを MoSaIC のほうでは関連するコンテンツ同士を矢印でつないだり、まとめたり、有向グラフで表現して、これを「関係の modeling」と呼んでいます。

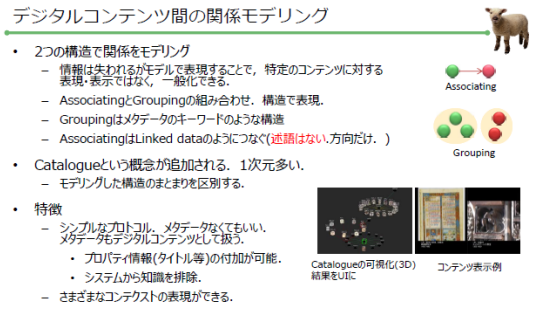


スライド 2 の中央に簡単な例がありますが、緑色の矢印でつながっているのは時祷書ですね。これは松田先生の写真を拝借してしまいましたが、4 月、5 月、6 月、これを並べて矢印でつないでみたカタログになります。赤いほうは 4 月ですから、おひつじ座ということで、おひつじ座に関係する

ものを集めたものです。

ざっくりと言いますと、このようなことを MoSaIC では Cataloguing と呼んでいて、このデジタルコンテンツがたくさんある中で、この矢印を引いてつなげていくことで、コンテクストを描くということに値するのかなと思っております。

そして、このコンテクストはあそこの共通する、共有するコンテンツによってさらにネットワークしていく。これを「Context Networking」と呼んでおります。



ここで（スライド 3）でもう少し関係モデリングについて説明します。MoSaIC では二つの構造でモデリングを行います。右側に書いてある Associating、これは二つのコンテンツの関連性を示すもので、それを矢印でつなげるもの。もう一つは Grouping と読んでいまして、なにか関係するものをなにかの観点で集めたものを表しています。この二つの構造で、デジタルコンテンツ間の関係を記述できるということを提案しております。

Grouping というのはなにかの観点があるので、これはメタデータのキーワードのような構造ともいえるかもしれません。Associating のほうは Linked data のようにつなぐということができますが、この特徴としては述語の意味はもたせていないということがあります。これは方向だけがあって、これからこれにつなぐよということだけが記述されるものです。

そして重要なのは、この構造で表した関係のまとまりを Catalogue ということで識別する、区別をすることが可能となっておりまして、これによって 1 次元増えるわけですが、いろいろな表現ができるようになります。特徴としては先ほどもありましたが、とにかくシンプルです。メタデータはなくてもいいんです。画像があって、なにか Catalogue をつくるぞって線を引けば、それで一つのデータになります。Catalogue になります。

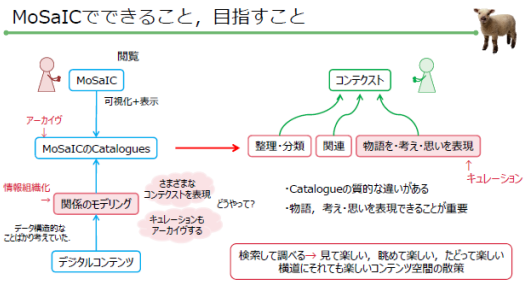
最初のころに議論があったのですが、メタデータはなくてもいい。最初はメタデータいらないと言い張っていたのですが、最近はちょっとメタデータいるよねということになりまして（笑）。メタデータはなくて

もいいよということを、少し弱気になっていますが、言っています。これは関連性をたどっていくことで、必ずしも検索、メタデータ的な検索ではなくても、なにか探していけるだろうな、探していけるといいなということで言っておりましたが、先ほどの金子先生の発表を聞いていますと、メタデータがメタデータちっくになって、メタデータとして入っていくなと思って、ちょっと裏切られた感があるんですけど（笑）。

メタデータというのはもうデジタルコンテンツの一つとして扱おうと思って、この 7 年間やってきました。メタデータもその概念ではなく言葉として、もう言葉が存在するんだ。存在論的な感じで、言葉が存在するから、それも電子データに言葉で文字データだったり、なにかすれば、それをくっつけてやることで人間が見れば分かる。もしくはコンピュータもある程度がんばれば分かるから、それでいけるのではないかと言っていたのが 5 年ぐらい前ですかね。もう最近はやはり一般的なメタデータのほうに寄っていったかなと思います。

このようにコンテクストの表現ができると言っておりますが、ここで一番の特徴が、その Catalogue をつくったあとに、Catalogue を可視化することですね。つながりをすべて 3 次元空間中にこうノードをおいてデジタルコンテンツを置いて、それを触ると中身が表示できるというようなシステムになっています（スライド 3 下図参照）。

これは、環境をモデリングするということは、すべてのコンテンツを一様に扱えるということになりまして、なにか特別なコンテンツを表示するものではなく、とにかくなにかつなげていけば、このように一様な空間に表示できるので、それが特徴となります。



そして、これまでずっと私はエンジニアリングのほうで関係のモデリングですね。その情報組織化、そのあたりのことばかり考えてやってきまして、では、この MoSaIC の Cataloguing でなにが表現できるのか。そういうことをずっとやらなければいけないと思いながらも後回しにして、Catalogue は誰かものをもっている、デー

タをもっている人たちがつくってくれると思ってやっていたのですが、なかなかどうもうまくいかないので、ようやく自分できちんとCatalogueをつくってみようという段階にきております。

よくよく考えてみたら、このCatalogueというのが従来であるデータベース、整理・分類してデータベースにきちんと登録するという側面もありますし、Linked data のように関連をどんどんつなげていくという側面もあります。

一つこの、ほかにそれにない新しい特殊な考え方としては、このCatalogueでは物語をつくったり、考え・思いとかをつなげていって、その順番であったりとか、その構造というか関係とかを表現できるのではなかろうか。これはキュレーションということにつながっていくのかなと思って、いまこれを、ここの部分がどうやったら。モデルケースですね。こうやったら、こういうのができるみたいな、そういったことを考えようとしはじめています。

そしてMoSaICとして今後、目指していきたいところが、検索して調べるところから、コンテンツを見て楽しい、そしてコンテンツにつながっている Graph の様子を見ても楽しい。それで一つずつたどっていって順番に見ていっても楽しいし、横道にそれて見ていっても新しい発見があって楽しいというようなコンテンツ空間の散策ができるようなアプリケーションの開発をしていきたいと思っております。以上で終わらせていただきます。

石川 尋代（いしかわ ひろよ）

慶應義塾大学デジタルメディア・コンテンツ統合研究センター特任講師。博士(工学)。2010 年慶應義塾大学大学院理工学研究科開放環境科学専攻博士課程単位取得退学後，同年博士（工学）取得。その後、慶應義塾学理工学研究科特任助教を経て、2012 年より現職。専門はビジュアライゼーション、3 次元視知覚、アプリケーション開発。2012 年より DMC にて Catalogue の記述法、可視化の研究、インタラクティブ展示 MoSaIC の開発に従事している。

パネルディスカッション
「ボーダレスなデータ利活用のための情報組織化とは」

（パネリスト）

原 正一郎（京都大学東南アジア地域研究研究所教授）

金子 晋丈（慶應義塾大学理工学部専任講師／DMC 研究センター研究員）

安形 麻理（慶應義塾大学文学部准教授）

久保 仁志（慶應義塾大学アート・センター所員）

石川 尋代（慶應義塾大学 DMC 研究センター特任講師）

（モデレーター）

安藤 広道（慶應義塾大学文学部教授／DMC 研究センター副所長）

安藤：それでは、これからディスカッションに入っていきたいと思います。時間がおしていることもありまして、話題が限られてしまうかもしれませんが、精一杯盛りだくさんの議論ができるようにしていきたいと思います。

先生方の講演とポジショントークを受けて、すぐに議論に入っていくほうがいいのかもしれませんが、私も一言だけお話しさせていただきたいと思っています。時間がないのでどうしようかと思っていたのですが、やはり少しだけお話をさせていただきます。

私自身はメタデータということに関してはまったくの素人と言っていいのですが、本日、いろいろとお話を伺っていて、データベースに蓄積した情報を利用するためにやはりメタデータは不可欠なんだなと、改めて思っているところです。ただ、私自身は世代的にポストモダンという世界の中で研究を進めてきた人文学者ですので、そういう目でメタデータを見ていると、いろいろと疑問とか心配になってしまうことがどうしても出てきてしまうんです。

どういうことかというと、久保先生をここにお呼びしたのは大正解だったなと思っているのですが、メタデータは蓄積された情報を利用するための組織化に必要なものである。ということは、あらかじめそれらの情報を利用する主体にとって有用な、ある特定の利用のあり方を前提に構築されているということですね。つまり、ある一群の情報の特定の利用のあり方を共有するためのものである。

久保先生はパースペクティヴという言葉を使われていらっしゃいましたが、逆に言

うとそうした利用方法を共有する利用主体のまとまりが、あらかじめ想定されているということにもなります。となると、その利用のあり方を共有する利用主体のまとまり、たとえば使用言語のまとまりというのもありますし、学問分野や関心領域というまとまりもあるわけですが、当然のことながらそのまとまりごとにメタデータの構成は異なってくる、つまり多様性を持つことになる。そこにそれぞれのまとまりの領域の横断を難しくするボーダーが張り巡らされていくのではないかということです。

問題はそのボーダーの存在だけではありません。利用のあり方をあらかじめ前提として情報を組織化するということは、ある意味で、その利用のあり方を共有する利用主体のまとまり内に対しても、利用を通じて人々の思考の方向性を限定してしまう、つまり思考の自由な拡がりを制限してしまう力が働いてしまうのではないかと、私などひねくれた人間は考えてしまうわけです。

たとえば同一分野や関心領域の中にも、本来メタデータを巡っては、久保先生のお話にもあったように、さまざまなパースペクティブが存在するのが当たり前でありまして、そうした中で、無理矢理統一したメタデータによって情報の利活用を行う。仕方がないかな、これでいいかなと合わせて利活用し続けていると、知らず知らずのうちに、本来多様な視点があり得るものであることが忘れられてしまうのではないかということです。

メタデータのことを考えていくと、どうしても私はそういうところに引っかかってしまうんです。つまりメタデータのボーダーによって、われわれの思考が制限されてしまっているのではないか。そんな疑問と心配がどうしてもぬぐい去れないということです。ほかにもいろいろあるのですが、これ以上話していると時間がなくなってしまうので、まずはそうしたボーダーというものを乗り越えていくことに関して議論をしていきたいと思います。

本日、二人の先生のご講演では、そうしたメタデータのボーダー。特にシンクロニック（共時的）な位相におけるボーダーを乗り越えていくための現在進行形の研究についてのご紹介がありました。原先生のご研究はある意味、私の感覚ですが、正道と言っていいアプローチのように聞こえました。つまり、メタデータマッピングというところから始めて、まずは資源の共有システムをつくっていく。加えて、現在どんどん研究が進んでいる Linked Open Data、Sematic Web などを応用することによって、メタデータによって張り巡らされていくボーダーを乗り越えていこうという考え方、方向性だと思いました。

一方、先生方のお話をきちっと理解できているかどうかは不安ではあるのですが、

金子先生のコンテクストネットワークのご研究、そして、展示してある石川先生のMoSaIC の開発に関わる研究は、これまでの情報の組織化の研究としては、これも私の勝手なイメージなのですが、変化球という感じがするんですね。

今ちょっと内紛が起こっているみたいなので心配ですけど（笑）、メタデータに依存をしないあり方､最初はそうだったらしい。つまり、メタデータに依存しない意味的隣接の Graph 情報というのでしょうか｡そうしたものに基づいたデータ間の関係を重視して、メタデータ界のボーダーを乗り越えていこうというご研究だったと思います。

こうした研究事例だけでなく､安形先生、久保先生のポジショントークでも面白い話をたくさん伺うことができましたので、それらも踏まえてメタデータのボーダーというものの存在をどう受け止めるのか。そして、そのボーダーを乗り越えていくためには、今後どのような見通しをもった研究が必要になってくるのか。そのあたりのことを私的にはぜひ議論をしていただきたいと思っています。

ただ「そのような議論していたらいつまで経っても終わらないよ」ということもあると思います（笑)。ですので、ここからの展開は先生方の議論の成り行きに委ねるとして、まずはそれぞれの先生方のお話に対する、たとえば質問とかご意見とか、そのようなところからでもいいので、どなたか口火を切っていただけるとうれしいなと思います。あまりしゃべると怒られてしまいますので、私からはこれぐらいにしたいと思います。いかがでしょうか。

原：私の研究所におけるミッションは、研究資源をデータベース化して研究コミュニティに供するようにすることというか、それが採用の際に示された用務でした。データベースといっても、人文系の場合は目録的なものがほとんどでしたから、その意味ではメタデータありきでした。

先ほど申しましたが、メタデータは、ある対象のプロキシ、資料へアクセスするための代用物に過ぎませんから、必要ではあるけれども、真剣に考える必要のあるものとは最初のころは考えていませんでした。むしろ、データベースシステムを効率的に管理・運用する上で、多様なメタデータは邪魔で、統合するのは当たり前と考えていました。

だから、みんなで共通のメタデータを使

いましょうと言ったのですが、驚いたことに、標準メタデータを使っているはずの図書館ですら、賛成してくれませんでした。対象とする史資料を標準のメタデータでは記述しにくいということと、MARC 以前から使っていたメタデータを今更変えたくないということのようでした。研究者の場合、これも先ほどの発表でも述べましたが、たとえ同じ資料のメタデータであっても、研究の目的が異なればメタデータは異なるので、共通化は言語道断であり、耳も貸してくれませんでした。ですが、研究者の側に立ってバラバラなデータベースを幾つも公開するメリットと、利用者の側に立って一回の検索で資料を検索できるメリットのどちらを優先すべきか、またシステムの効率的な管理も勘案すると、共通メタデータの方が良いと判断して、数年ほどは粘り強く説得を続けたというか抗ったのですが、最終的には諦めました。説得しているうちに、研究者がオリジナルなメタデータを重視するのも宜なるかなと考えるようになってきました。

そこで、バラバラなメタデータを取り込むことができて、しかも管理や運用も容易な、言ってみれば研究者にとってもシステム管理者にとっても「最小不幸？」となるデータベースシステムとして工夫したのが My データベースです。My データベースは、基本的には、どのようなメタデータであってもデータベース化できます。

しかし、My データベースで作った個別のデータベースを繋ぐ段になって、当然のことながら、メタデータの違いは壁になりました。言い忘れましたが、今話しているのは、前の職場で扱っていた国文学データベースです。ですが、同じことは、今の職場である地域研究にも当てはまります。国文学では、ジャンルや時代が違えば、使う語彙が違ってくるので、壁が出てきてしまうんです。これを乗り越える方法としては、語彙の関連あるいはマッピングを整理したオントロジーといいますか、シソーラスが必要です。ところが、厄介なことにシソーラスの整備されている学問分野はほとんどありません。少なくとも国文学にはありませんでした。さらに困ったことに、誰もオントロジーを作ってくれませんでした。これも国文研究者の立場にたてば、自分の研究に直接関係ないことにエフォートを割り振りたくない気持ちは理解できなくもありませんが、ドメイン全体の将来を考えると問題となるかもしれません。

ということで、今のところ、語彙間のマッピングにはダブリンコアを使っています。このマッピングはあまりにも粗雑というかラフなので、幾つかのオントロジーの組合せを試みています。これは、書籍データベースであれば図書館用標準メタデータである MARC にマッピングし、文書データベ

ースであればアーカイブ用標準メタデータである EAD にマッピングし、書籍データベースと文書データベースはダブリンコアを使うといった、マッピングの対象と粒度に注目した使い分けです。この方法は RDF を使った次世代情報基盤で実装中です。

この次世代情報基盤では Linked Open Data の実現を目指しています。情報アクセスのきっけは、相変わらずキーワードですが、それ以降は地名辞書や Wiki などの知識ベースを介したリンクを辿ることになるので、ドメインを超えたデータ共有、あるいは知の共有に一歩近づくことができるのではと期待しています。

最後の Web ビッグデータはキーワードによらない情報アクセスと分析の試みです。たとえば新聞記事を検索しようとしても、量が多すぎてキーワードを付けることができません。文字列検索も可能ですが、かなりのゴミも拾ってしまいます。そこで、機械に勝手に分類させてしまおうと言うのが、ここでの発想です。具体的にはトピックモデルを使っています。文書中に現れる語彙の出現確率に基づいた分類を行う方法なので、キーワードではなく内容の類似性によりデータを分類します。新聞の場合、似たような内容の記事を抽出します。トピックを色分けして地図に表示したりしています。機械が統計的に分類するので、「何で？」というパターンがある反面、これまで気がつかなかった面白そうなパターンも散見されるので、新しい研究の視点あるいはヒントを得ることができるのではないかと期待しています。

久保：つながっていないのですが、よろしいですか (笑)。まず言い忘れたことを先に申し上げたいのですが、メタデータのボーダーがどこにあるのかということで、皆さんもお気づきだと思うのですが、私の発表ではパースペクティヴの差にボーダーがまずあるだろうと思っています。そして、もう一つは「利用主体のまとまり」と安藤先生が仰っていたことですが、パースペクティヴを束ねる客観的な値だと思われているメタパースペクティヴというんですかね。そういうものにもボーダーがあるだろうと思います。

おそらくこれはいまのデータベース系の問題にそのまま交換可能な問題設定だと思うのですが、メタデータをなしに、そもそもコンテンツネットワークを形成するというのが、なぜ無理だったのか。メタデータ

を利用するようになってきたというのが、それはなぜなのか。そもそもメタデータのないネットワーク、データベースというものがあり得るのかというのが、ふと素朴な疑問としてありまして。そのあたりをちょっと教えていただければと思うのですが。

金子：それは最初に言うべきだったかもしれないんですけど、私はいまでもメタデータはいらない。なくても動くシステムだと思っています。そういう意味で、まず石川先生に誤解しないでねという（笑）。まず、それが一つです。内紛ではないです（笑）。いやいや、よくある話です。

そもそもメタデータというのは意味理解を必要としますという話と、ボーダーが存在するという、それは理解とセットですという話について私はそのとおりだと思っていて、究極な世界を考えたときには、メタデータなんてなくして、ただ単に渡されたほうが、それがなにかを考えるきっかけになるのではないかと思っています。

それをさぼりたいと思う、人間の甘い考え方がメタデータという存在を生んでいるのではないか。ちょっと全部読みたくないけど、あらすじだけ知りたい、といったような。某航空会社のラウンジのアプリだと、面白いんですよ。本を並べているかのように見せて、それは忙しいビジネスマンのためのあらすじ集なんですね。「あっ、これ読んでみたいな」と思って開けたら、あらすじしか書いてないんですよ。

それって本当にそれを分かったことになるのかなということなんですけれど、イコール、メタデータで分かった気になって勝手にその情報として除外しているようなところがあると思っていて。そんなのだったらいっそのことそんなメタデータなんてやめて、それこそ獅子岩の現場に行って、神社まで歩いてみて、あそこの周りは360度、365日ずっと周遊し続けて、獅子岩とはなんたるかを考えるのが、本来獅子岩を分かるということに非常に近い。もしくは、その獅子岩の周りで生活している人に密着して、その獅子岩の位置づけを肌で感じるというのが本当の理解だと思うんですが、たぶんみんなそれをサボりたいからメタデータを使いたいんじゃないか。そういう真に理解する人の数と、その甘っちょろくちょっとだけ分かった気になりたい人の人口比を考えたときに、後者の割合が多いんじゃないか。「まあ、メタデータ入れとく？」というふうになってきていると理解していた

だいたほうが正確かなと思っています。

久保：すみません。本当に素人に説明していただくつもりで答えていただきたいのですが、私からすると、メタデータなしのデータベースってよく分からないんですね。つまり、そもそもあらゆるデジタル・コンテンツはメタデータなんじゃないのと。どういうものをメタデータではなくて、どういうものをデータというふうに区別しているのかというのもご説明いただいていいですか。

金子：Graph というものが実は情報そのものになっているんですね。ですから、別に地図がなくても町歩きはできるじゃないですか。その道がここで、交差点があります。そこをすべての道に入っていって、「ああ、ここを行くと、ここに出るんだ」みたいなことを。たとえばイタリアとかの旧市街を歩くと、小さな町だけだけど、たぶん 3 時間ぐらい歩くと、すべての小道も全部入れていって、頭の中に地図が再現できるみたいことはできると思うんですが、それに近いものだと思うんですよ。

そうすると、一種その地図情報というのをメタデータとして頭の中に入ってくる。では、メタデータがなくて町歩きができないかというと、そうではなくて、ガイドブックもなにもなく街を歩くことは可能だし、そこから学び取っていくことは可能だと思うんですね。だからメタデータがないとデータベースがつくれないとか、情報が構築できないというのは、それはいわゆる既存のデータベース構造にすごく寄った考え方なのかなと思いますね。

Graph という面白い情報を保存し、共有するメカニズムがあれば、それも一種のデータ。それもデータベースなんですけどね。

久保：獅子岩でいうと、たとえばどういうふうに考えられますか。先ほどおっしゃった理想的な意味の到達地点みたいなものが一つあるとして、そこに至るまでに、獅子岩をコンテンツネットワークの形成みたいなモデルとしてなにか具体的な方法論を考えるとしたら、どんなものになるでしょうか。

金子：たぶん私だと、先ほど見せていただいたのをどういうふうに Graph 化するかという話ですが、モノのハブみたいなのもちょっと説明しましたけれど、モノ ID というのを獅子岩という一種の抽象概念を設定して、そこにいっぱい撮っていただいた写真を Graph 上に交差点のなんていうんですかね、凱旋門の伸びる放射状の道路のように、その先に写真をバンバンとまず置いてみて。またその獅子岩からずっと伸ばしたところに、ここに神社を配置してみた

いに。その一個のポイントから複数のところにいざなえるというか。

久保：すみません。私からすると、その写真がメタデータではないのかと思うんですけど、それはメタデータではない？

金子：そういう意味でいうと、Graph 構造がメタデータになっているといえば、なっていると思います。逆に、その Catalogue でポイントにしているのは、その言語は使わないけれども、関係があるかないかを示しているんですね、実は。関係があるというメタデータが Graph に存在しているのではないかと言われたら、そのとおりです。でも、そこにたとえばもっと付随する、補足的な情報を言語として与えているかというと、ベースの考え方としてはそれはなく、少なくともなんらかの関係があるよという情報をメタデータとして提示しているのが Graph だと私は思います。

原：金子先生にお聞きしたかったことがあります。先ほど、自分たちのデータを RDF 化したと言いましたが、これを可視化したらどうなるか試してみたところ、使い物にならない。ものすごい数のノードとリンクが画面中を覆い尽くしますから、最初はスゴイと感動するのですが、「だからどうなっているの」と思い直して眺めると、さっぱり分からない。結局、RDF の主語や述語の語彙を利用してフィルタリングしないといけないんです。ある国で行った別の研究でも似たようなものでした。地域住民の関係図を作ったのですが、最初は A さんから B さんへという単純な有向グラフでした。ところが可視化したらゴチャゴチャで使い物になりませんでした。結局、親子関係や師弟関係などのいろいろな関係をグラフに追加することになり、言ってみれば意味付き有向グラフになってしまい、結局、リンクに付けた語彙を使って、関係をフィルタリングしたり、色付けして区別したりしました。つまり、グラフを使っても、言葉に頼って操作しています。

この経験を写真の例に当てはめると、私が持っている 100 枚や 200 枚程度の写真データならば問題ないと思いますが、組織全体が持っている、たとえば 100 万枚の写真の場合、先生のやり方がどうなるのか気になります。

金子：数の問題ですけど、まず私が立って

いるところは、すべてのデータが精緻にそういう Graph 化されていくかということにおいてはノーだと思っています。やはりじっくり研究されたものに関しては、いっぱい線が引かれるであろうし、もしくは著名なコンテンツであればいっぱいエッジが出てくると思いますが、では、日々スマホでカシャカシャ撮っている一個一個にそんな RDF を書くような細かい単位の情報をエッジとして与えていくのかというと、最初はやるかもしれないですが、たぶんそれはほぼ不毛な作業だと気づくと思うんですね。そうすると私は人間のレイジーさというか、そこもちょっと含めて考えているんですが、そうすると一個の写真を撮る、もしくはあるイベントで複数個の写真を撮ったときに、ほかのコンテンツと結びつくのはたかだか数枚の内の 1 枚だけが 2 本ぐらいとつながって、残りは同じイベントですよという、1 本ぐらいしかつながらないような Graph の世界を私としてはイメージしている。

いかに埋もれるデジタルコンテンツをなくすかというのをまず考えていて、そのためにはやはりつながっているということが重要で。でもリッチにいっぱいつながりすぎると、もちろん先生がおっしゃったみたいな問題になってくる。だから逆に私が恐れているのは、すごく時間をかけてじっくりといっぱいリンクを一個のコンテンツに張る問題よりも、エッフェル塔とかというすごく著名なものに対してものすごくたくさんのエッジが多くの人によって引かれることのほうがどんな世界がくるんだろうという恐れを感じています。伝わりましたか。

安形：私も金子先生に質問です。Graph でいろいろつないでいくにしても、なんらかの計算はしてつないでいくわけですよね。もしかしたら、理解が間違っているのかもしれないのですが、たとえばそのエッフェル塔にたくさんつながるのか、それともほかのもの同士でつながるのかというところがあるときに、どういう形で計算していくかというところは、言葉でなかったとしても、必ずメタデータ的な観点がなにかあるのではないかと思ってしまうのですが、そのあたりいかがでしょうか。

金子：特に計算ということは考えてなくて、まずは人力です。結局さっきのメタデータ、その関係があるのは人間のパースペクティヴを反映しているという話からすると、そ

れは個人に依存しているので、個人が線を引きたいと思ったら引けばいいんじゃないかというのが、いまの立ち位置です。ですからコンピュータ的な計算が必要かというと必要じゃなくて、ただ単に言われたとおりリンク張りました的な計算をまずは想定しているんですけれど。

安形：そうすると、一つのものを見て、私がつなぐものと金子先生がつなぐものが違ってくるわけですよね。そうした場合に、なかなかほかの人のものを活用したり、統合したり、あるいは横断的に見ることはどちらかというともう諦める方向なのでしょうか。それとも、その先にやはりなにか共有の方法というのは、どんな形であり得るのでしょうか。

金子：私は、そのコンテンツ空間が全部あって、それを本当に左上から右下まで横断的に探すというのは完全に諦めています。でも諦めていないのは、自分が興味を持っているところから、たとえば Graph 的に何ホップかいった範囲内。そのある中心から同心円状に広がりがあったとして、そのある範囲内はきちんと見られるようにしたいねというのがいまの立ち位置です。ですから、その範囲に本当にほしいコンテンツがなかったら、残念みたいな感じですね。

安藤：そうなんでしょうね。ぜひ石川先生にもこの問題についてコメントいただきたいと思います。

石川：本当にメタデータを分かっていないんだなということを、よく分かったんですけれども。最初におっしゃられたように、どっちもメタデータではないかとおっしゃったように、私の場合はどっちもデジタルコンテンツでいいじゃないかと思ったほうで。そのへんの違いもよく分からずというか、それでいいんじゃないかと私も思ったんですよ。

もう一つは、メタデータは使うだけだったんですが、その使い方というのもあまりこれまで絶対に有効ではなかったですね。いろいろなデータベースのサイトに行きます。それでなにか調べようと思っても、専門外なのでなにを入れていいか分からないんですね。キーワードが分からない。メタデータに何が入っているか分からない。専門用語はさらに分からない。一番分かるのは年代かなといって年代を入れてみると、

いろいろなものが出てきてもうなにがなんだかという。

そんな状態で、なにも知らない素人のユーザー側からはメタデータがあまり見えてきていないのではないかなと。なので、ユーザー側から見たら、どこにどうやってそのメタデータが有効に使われているのかというのが分からない。データベースも全部別々ですし、一つのデータベースで探して、次のデータベースで探してと。結局、なにか面白いものも発見できずに終了するということが多々ありまして。

なにか調べるというトリガーがないんですね。エンジニアリングで古文書なんて絶対、そんなに調べないんですよ。でも、それがなにか面白いかもしれないと思って、そのデータベースに行ってみてみるトリガーが必要なのだと思います。

久保：関心はあったとしても、そのとっかかりとなるものが必要ですね。

石川：そうですね。なので、最近は全面的になにかトピックが出てきて、そこからクリックして入っていくというのが出てきたので、まだ少しぐらいはいいかもしれないですけど。本当に一般の人たちから見たらメタデータの議論はしているんだけど、果たしてどこにどうやって活躍しているのかがよく分からないところが現状です。

金子：補足で、われわれがつくっているのは、先ほど原先生は研究者向けだとおっしゃったんですが、うちらはいかに文化財データとか、あらゆるデジタルデータを一般に開放して使わせるかというところに結構違いがあるかなと思っています。

そうすると、さっきのメタデータのキーワードとか項目とか要素とか、それが分かってないと探せない。使えない。そこの敷居を下げるために、やはりそこから外れないといけないよねというのは考えているポイントとしてありますね。

安形：デジタルアーカイブでは、たとえば赤い絵とか、そういうもので手軽に見ることができるとか、テーマ別にいろいろ見られるというものは確かに結構あります。特に美術館・博物館のアーカイブなどだとそういう面白い見せ方があると思うのですが、一方でそういう方に注力しているアーカイブは、メタデータとしてはそんなにがんばっていなくて検索しづらい傾向があります。

テーマとして出てきているところでは見ることができるけれど、この絵を実際に見たいときにどうやったら検索して出てくるだろうといったときの、メタデータの整備の状況はかなりいろいろです。でも、なぜそういうものが増えてきているかというのも今日はよく分かって、よかったと（笑）。ただ、検索を考えた場合にメタデータというのは非常に重要になってくると思います。

原：データベースを作るときには目的があるので、それは重視する必要があると思います。ですから、研究者が作るデータベースは、どんなに良い物であっても、恐らく小学生や中学生が使うことはなかなか難しい。誰でも使えるデータベースであることに越したことはないし、工夫してできないことはないと思うけれど、それは本来の研究とは全く違う別の研究テーマなんですね。

研究者のデータベースというのは、目的に応じて、自分が整理したいようにデータを組織化するんですよね。たとえば、写真を対象としても、色に注目したデータベースがあれば色についての情報は豊富であっても、内容についての情報は乏しいかもしれない。だからと言って、研究データベースを統制することはできない。そのような研究データベースを少しでも広く使えるようにするには､私たちがやっているように、言語ではない情報、例え場所や時間を追加するとか、オントロジーあるいはシソーラスを使って、語彙を拡張して繋いでいくしかない。

でも知らない言葉って実はいっぱいあるんですね。最初、国文学の世界に触れたとき、音を漢字に変換できない語彙がたくさんありました。たとえば「ほんこく」が「翻刻」なんて分かりませんでした。あるいは、国文学者はテキストデータベースを「本文データベース」と書いていたのですが、「ほんぶん」と読んだら笑われてしまいました。つまり、オントロジーあるいはシソーラスを如何にうまく作るかが鍵になっている。

だから金子先生がおっしゃるメタデータを使わないという趣旨はよく分かるんですが、だからといって線で引っ張ったから何かが見えるかというと、やはり違うのではないのかと。そこにはなにか語彙的な枠組みが必要なのかなと、どうしてもまだ思ってしまう。

金子：私はネットワーク屋なのでアナロジーでちょっといろいろ考えるんですが、そ

の電話網とインターネットって二つの通信サービス、通信ネットワークがあるんですよ。端的にいうと、電話網というのは電話サービスオンリーで、それに最適化したシステムなんですね。インターネットは、最初はもちろんボイスという音声コミュニケーションをサポートできず、そのコンピュータ間の通信だけが世界中でできるよというものでしたが、そのうちグングーッと性能が上がっていき、いまや普通に国際電話なんてせずにスカイプで国際電話しますという世界にどんどん変わってきています。

いま Graph 化しようというのは、いろいろなもちろんアプリケーションごとに違った特性のあるメタデータで検索できる。そうだろう、それは私にとっては電話のイメージです。でも、それがもしプラットフォームとしてインターネットのように共通化できて、それを複数のサービスに使っていけるようになるのであれば、それはもっと新しいネットサービスを生むだろう。逆に言えば、もっと新しいデジタルデータのサービスを生むでしょうと。

なにが違いかというと、先ほどフィルターの話もありましたが、結局どう使っていくかの話をそのプラットフォームを共通化した上で、ネットワークの用語で一個ユーザーのレイヤーを上に上げるということなんですが、レイヤーを上に上げたところで処理することによっていろいろなサービスを共存させる。同じ情報を使いながら複数のサービスを共存させる方向に、いま私は進みたいと思っているというのが全体の考え方ですね。

安藤：いかがでしょうか。会場から松田先生が手を挙げていらっしゃいます。一応、所属とお名前を。

松田：文学部の松田と申します。よろしくお願いします。先ほどの話を普段と離れたところから今年はここで聞いていて、なんていうか、金子先生が目指しているのは私も本日あらためて真のデジアナ融合だなと思ったんですけど、やはり情報組織化とか、それとはまったく違う話ですよね。

金子先生は、つまり一人の人間がどうやってデジタルと関わっていくか。それがアナログの人間として関わっていくかというところをやろうとしていて、それはわれわれもそうだし、MoSaIC はそれをやろうとしていると思うんですね。一方で、だけど情報を情報としてちゃんと組織化していく必要がもう一つあると。

それと MoSaIC とはまったく別なものであって、それぞれあればいいと思うわけです。それはいま私が関わっている、理事もおっしゃっていた慶應ミュージアム・コモンズでもやはり金子先生がいうようなものをミュージアムがもつデジタルの側面と

して出していこうと考えているところがあるので、そこをちょっと、こういうことやりたいんだけど MoSaIC ではどうやるんですかとか、そういう話ではないのではないかというのが一つです。

もう一つは、メタデータという使われ方については私が考えていたものよりアバウトになってきているように思います。私が考えていたメタデータは本来、触る必要がないもの。最初にそれが出てきたのは、たとえば安形先生などと貴重書のデジタル化をやっていたときに、これはいつ誰がどういうふうにして撮影したのか、JPEG なのか TIFF なのかということであって、それは普段は触る必要がなくて、なにかあらためてもう一回写真を撮りましょうというときに、初めて見ればいいものであって、リーダーズ・ダイジェスト的なものとか本のタイトルとか、そういうのは全部データだと思うんですね。

あえていうならパラデータであって、メタデータではないのではないか。だから、ここで聞いている話は全部データの話ではないかと、極端なことをいうと私には聞こえてしまうんですけど、どうでしょうか。

安藤：なかなか難しいお話だと思うんですが、やはりある情報に対して付属している情報はメタと言っていいと思うんですね。ある情報本体に対して付属して、それに対してなんらかの利用の方向性、検索のあり方、そういうものを規定していくものはすべてメタデータで私はいいと思いますけれど。そのあたりはもしかすると、谷口先生がちょっと苦笑していらっしゃいますので、もっと高いところから、このあたりのところを解説してくださるのではないかと思いますが。

谷口：文学部の谷口と申します。私などが話していいのかという気がしますが、おっしゃる皆さんの立場というんですか。とらえ方はそれぞれ妥当な。つまりもともとメタデータというのも、また先ほどの話で定義を誰がどう決めて安定して使われているかという話になりますので、そこはとらえ方によってデータに対してのメタデータという、メタレベルの。おっしゃっていた管理情報みたいなのは、あるいは測定に関わるもろもろの証拠情報というんですかね。そのあたりのものをメタデータとおっしゃる場合と、それから世の中にある事物に対してなんらかの形でデータとして表現した

段階で。先ほどの写本に対して、それがタイトルが誰、あるいはどこで見つかった。なんでも構わないのですが、ある実在物に対してデータとして表現した段階で、もうそれがメタデータだというとらえ方でしょうか、使い方というものがあって、両方の使い方が混在しているというのがややこしい部分ではありますし、どちらも妥当な使い方だと私は思っております。

安藤：どちらかというと、本日はいま谷口先生がおっしゃった後者的なところに比重が置かれていたと思いますね。とはいえ、どっちがメタデータなのかとか、二者択一的なものでもないと思いますし、どちらであってもそれぞれ問題点があれば、それをカバーしていくなんらかの改善点を見つけていくことがやはり大事なのかなとは思いますけれども。

せっかくですので、高山先生、よろしくお願いいたします。

高山：いまはもう所属はございません。元文学部におりました高山と申します。いまちょっとお話があった続きでいいますと、やはりメタデータの本来の意味は、対象をコントロールするところにあると思います。メタデータがなんであるか。デジタルはどうであるか。あるいはデジタルデータがどうであるか。それから DMC としてなにをどういうふうにやっていかれるかということは、それぞれの先生方がそれぞれのお立場からお考えを深めていただければいいわけであって、それを本日ここで交流させていただいて、一つにまとめる必要は毛頭ないと思うんですね。

ぜひ、こういう場で先生方のご意見を外延的に広げていただきたいと思っておりまして、それで一つご質問させていただきます。メタデータが先ほど来お話がありますように、一つは利用のためにという側面から論じられていた。もう一つは、これはコントロールをするため、管理をするためという側面もあるんだというお話は、先ほどの谷口先生のお話も含めてあると思うんですが。

そのコントロールをするために、これから特に慶應義塾の場合に新しいミュージアムを考えておられると考えたときに、これは特に安形先生にお尋ねしたいのですが、図書館の世界では書誌コントロールという概念があるわけですね。そういう場合に、そのメタデータのコントロールをどういう形で、どういう機関が中心になってやるか。

仮に慶應義塾の新しくできる博物館が、少なくとも自分が所有権を主張するところのデータをコントロールするとなったときに、どういう形でコントロールをしたらいいのか。それを裏付けるための法律の問題はどうなるんだろうか。あるいは、いや、その法的な問題まで発展させる気はないということであれば、ビジネスベースでコントロールの実際を実現するためにはどういうふうにしたらいいのだろうとか。そのあたりのお考えがあるかどうかをお尋ねしたいと思います。安形先生と言いましたけれども、ほかの先生方もご意見がありましたら、ぜひご教授いただきたいと思います。

安形：書誌コントロールを谷口先生の前で私がいうと、大変（笑）。それはぜひ谷口先生から簡単にご説明をいただけますと幸いです。

谷口：私にはあまり取り上げたくない用語ですね。非常に説明が難しいと思います。私らの領域というんでしょうか、一つの先ほどのボーダーの中においても、たぶんいくつかの定義の仕方、説明の仕方があってしかるべき用語。私なんかはどちらかというと戦略的な用語なのかなという感じが、要するにもともと中身があるわけではなく、そういう言い方にして、それをやりたい。あるいは、それによってお金なり人なりを集めて、なにかを動かしていきたいという人たちが、その時点において使ってきた言葉なのかなというのが正直なところです。

一般的に、書誌コントロールというのは、書誌と言っているのは、原先生の前でこう言うのもなんですが、われわれが図書館なりなんなりの領域が扱っている資料群、情報資源、それを社会的にコントロールする。なにがどういうものがあったのか。あるいは、どこにあるのかということが社会的にコントロールされる。つまり、それにわれわれとしてアクセスできる。あるいは、そういう情報を知ることができるということを実現する。それが、単に一つの組織ができる部分も当然ありますし、それが寄り添ってというんでしょうか。連携してもう一枚上のレベルでそういうコントロールをしていくというような意味合いもあって、全体、そういうものの総称だと私は理解しておりますが、高山先生、このような回答でよろしいでしょうか。（笑）。

金子：どこまで立ち戻って話をするかにも

よるんですが、デジタルデータを、先ほどの書誌コントロールの話でいきますと、どこまで保存し、管理してアクセスできるようにするかと考えたときに、その保存のコストというものを絶対に忘れることができないと思っています。何回かこのシンポジウムを通してしゃべっていることなんですが、デジタルデータを保存するという行為は、実は媒体を変えながらずっと永続的にやろうと思えばできることだと思っています。したがって、イコールそれは経済的な裏打ちがあれば、金持ちであればデジタルデータを永続させることは技術的に可能だし、論理的にも可能だと思っています。

しかし、すべての人、すべての組織がそういった経済的資源を持ち合わせていない場合において、データが失われる危険性が出てくる。いろいろな考え方があると思いますが、「すべてのデジタルデータは貴重な人類の資源であるから、すべて保存しておくべきだ」的な考え方から、「いや、使われないのは消しちゃえばいいんじゃない。だって金もったいないし」という考え方まであると思うんですけど、私はどちらかというと後者に近い考え方です。かといって、完全に使われないものを捨てるべきなのか。それって 10 年後も使うかもしれないよねと。

なぜ Graph のような考え方を持っているかというと、Graph にしておけば、少なくともつながりがあるので、一個手前、二個手前までアクセスした人がそこにリーチできる可能性があり、そこにそれの延命のチャンスが与えられるのではないかなと思うんですよ。

もし既存のデータベース構造でなにかにマッチするものだけが出てくると、いかにそれが情報的に隣接していたとしても、それはマッチしないので永遠に日の目を浴びないということになると思うんですね。その情報をやはり保存する側ってそれがイコールお金になってくるし、利用する側はそこのお金の部分が見えていないので、保存する側というのは苦しみを持っていると思うんですね。その苦しみに対してなんらかのインプットを与えてあげないと、その苦しみを耐えられないと思っていて、そういったときに近くまで欲しい人が来てますよという情報があるだけで、実は救われる部分が多いのかなと思っています。

ですから、法制度によってそれを保証しようとか、たとえば国でアーカイブ組織をつくってそこで保証しようという考えはどちらかというと私は持っていなくて、基本的にはより多くの人類がより多くのデジタルリソースを利活用することによって、その健全なデジタルのエコシステムというか、デジタルデータのエコシステムというのが形成されるのではないか。それに対して国がなにかしらの援助をするのであれば、そ

れはそれであり得るけれども、ベースとなって動くのは、国が社会を動かしているわけではなくて、人々が国をつくって社会を動かしているわけですから、やはりそこで動くデジタルデータというのも人々によって動かされ維持されるのが基本ではないかと考えています。お答えになっていますでしょうか。

高山：ありがとうございました。私もそう思うのですが、先生がおっしゃるようにネットワークの中で利用者と権利者とが話し合えばそれでいいんだというところへ持っていきたいのですが、それで果たしてどこまで可能かというところに疑問を感じたものですから。でも、それをともかくやってみるよりしょうがないと、こういうことですよね。

金子：そうですね。

高山：はい。ありがとうございました。

安藤：議論が尽きないところ恐縮です。メタデータという話でここまで多様なパースペクティヴが出てくるとは思わなかったというか、もちろん多少覚悟はしていたんですが、やはり広がってしまったなというところですね。

実は予定の時間を十数分過ぎていますので、そろそろ終わりにしないといけないと思うのですが、せっかく 1 名の方からご質問がありましたので、金子先生、簡単にお答えいただけますでしょうか。

金子：「1 億台の監視カメラを全部検索するのは大変というお話がありましたが、全部検索するスピードを上げようというアプローチについてはどう思いますか。将来的な可能性はあるのか」。1 億台の監視カメラを全部検索するスピードを上げようというアプローチもあってもいいと思いますが、それはエンジニアリング的にはあまり効率的ではないと思います。

それはどういうことかというと、もっと身近な例でいくと、皆さん、おうちありますよね。最寄りの駅からいろいろなところに通学、通勤されていると思うんですが、そのときに、いや、私、その目的地まで早く行きたいから、全部直通の電車をつくります的な話だと思います。それってちょっと効率悪いよねと、エンジニアとしては思うんですね。もちろんスピードを上げるための CPU リソースをどんどん向上させて

いくという考え方もありますけど、そしたらそれはもっと別のところに使ったほうが便利なのではないかなと思うので、1 億台の監視カメラに、絶対映っていないところに対してジョブを投げるのはあまり効率的ではないかなというのは、エンジニアリング的な観点のコメントです。

安藤：どうもありがとうございました。まだまだお話をしたいことがたくさんあるのではないかと思います。今日は私のモデレーターとしての能力不足から、議論がどんどん広がっていくことになってしまったのですが、一方で、メタデータをめぐっては本当に多様な論点があり得るということ、いろいろな課題があり得るというところは、おぼろげながらではありますが、この議論を通じて見えてきたのではないか。そうしたことを皆さんにも了解していただけたのではないかと思っております。

　まだお話をしたいという方は、このあと交流会がございますので、ぜひそこで納得のいくまで議論をしていただければと思います。今日は、お二人の先生方、そしてコメントをしてくださいました 3 人の先生方、本当にどうもありがとうございました。それから会場の皆さま、最後までお付き合いくださいましたことに感謝申し上げます。

研究ノート
「インターネット前提時代におけるパブリック・ヒューマニティーズへの挑戦」

宮北 剛己
（慶應義塾大学 DMC 研究センター研究員）

1. はじめに

インターネット前提時代[1]とも云われる昨今、デジタル技術を活用した教育・研究活動は大きな分野に成長している。そのひとつにデジタル・ヒューマニティーズ（デジタル人文学）と呼ばれる分野がある。その定義や解釈は多岐にわたるものの[2,3]「Big Tent」[4]あるいは「umbrella term」[5]とも隠喩されるように、デジタル・ヒューマニティーズとは、文学や歴史学、美学などの人文学研究を一つの枠組みに集約し、「デジタル技術を応用することで人文学を深化・発展させようとする一連の営み」[6]のことを指す。

筆者は、このデジタル・ヒューマニティーズに係る営みのなかでも、研究活動の裾野を広げていく試みであるパブリック・ヒューマニティーズの分野で研究に取り組んでいる。パブリック・ヒューマニティーズに関する定義や解釈もまた広範囲にわたるが[7]、学術研究を特定の研究者や専門家だけでなく、専門的な知識・研究経験を伴わない公衆/市民−パブリック−にも開いていく研究として位置づけられる。その他の多くのデジタル・ヒューマニティーズ研究では、デジタル技術を駆使した新たな発見・解読（の遂行）に重きが置かれる一方、パブリック・ヒューマニティーズでは、いかに、そしてどのように人文学の知をパブリックに公開・共有し、公衆/市民とエンゲージしていけるかが重要とされる[8]。

本稿では、まず、現在パブリック・ヒューマニティーズを取り巻く状況や技術的動向について述べた後、筆者の（研究）アプローチを紹介する。次に、筆者がこれまでに取り組んだ研究活動のうち、このパブリック・ヒューマニティーズの文脈に即したかたちで行われた研究事例を紹介し、最後に、今後の取り組み、展望について述べる。

2. パブリック・ヒューマニティーズを取り巻く状況・技術的動向

前述したように、パブリック・ヒューマニティーズとは、学術研究を広く社会に開いていく試みのことを指す。日本ではまだあまり馴染みのない言葉かもしれないが、欧米では、学際的な特徴をもつ、ひとつの学問領域として認識され、複数の大学機関では、パブリック・ヒューマニティーズに関連したプログラムやコースが提供されている[9]。また、日本でも昨今、パブリック・ヒストリ

ー（公共歴史学）と呼ばれ、人文学のなかでも特にヒストリー（歴史学）研究に重点を置いた分野・研究アプローチ（方法論）に対する関心が高まっている[10,11,12]。

このパブリック・ヒューマニティーズと呼応しながら推進されているのが、学術研究のオープンデータ化およびオープンアクセス化である。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」[13]のことを指し、オープンアクセスは、はじめは「インターネット上で論文全文を公開し、無料で自由にアクセスできる」[14]ことを目的に提唱され、今では論文に限らず、様々な学術情報（コンテンツ）を世界中の人々が広くオンラインで享受できるようにすることを指す。

そして、こうした枠組みが普及することで、今日では書物や美術品、文化財といった人文学（に係る）データのオープンな利活用も実現しはじめている。以下では、その技術的動向について、代表的なものを一部紹介していく[15]。

IIIF (International Image Interoperability Framework)

画像（データ）をはじめデジタルコンテンツを相互運用するための国際的な規格として、IIIF-トリプル・アイ・エフ-と呼ばれる（共通）規格がある。IIIFは、「Community Focused」「Defined APIs」「Plug ‘n’ Play Software」の3つを柱に展開され[16]、世界各地で制作・公開されているデジタルアーカイブで導入されている。IIIFに準拠して公開されているデジタルコンテンツは、標準APIを活用することでアーカイブを横断した閲覧、共有、再利用が可能となり、また、利用者自身が（複数開発・提供されているオープンソースのソフトウェアの中から）ビューワー等を選択して、（一定のルールのもと）自由にコンテンツを利活用できる。国内外での普及活動に携わる一般財団法人人文情報学研究所の永崎研宣氏によれば、IIIFは「サイロからコンテンツを解き放つ」[17]ことで、デジタルコンテンツのあり方に幅広い可能性をもたらす。

LOD (Linked Open Data)

LOD-リンクト・オープン・データ-は、オープンデータ（に付随する各種属性情報、所謂メタデータ）をウェブ上で統一的に記述することで、データを相互に接続して「つなぐ」仕組み・技術のことをいう。LODは、ウェブ上で記述・公開されるコンテンツの機械可読性を高めるセマンティック（ウェブ）技術に基づくことで、ただデータを公開するだけでなく、分散されている情報を統合的に扱うことを可能にする[18]。このLODは、ウェブ（World Wide Web）の考案者であるティム・バーナーズ＝リーが提唱する「5つ

星オープンデータ計画」[19]の中でも 5 つ星レベルとされ、研究機関のみならず、行政機関や企業でも導入が進んでおり、分野を超えたオープンデータの公開・利活用が促進されている。

Japan Search（ジャパンサーチ）

上記で紹介した 2 つとは異なり、技術ではなくウェブツール（ポータルサイト）の紹介となるが、機関横断型デジタルアーカイブとして、欧州の Europeana[20]・米国の DPLA (Digital Public Library of America)[21]に次ぐかたちで、2019 年 2 月末より、Japan Search（試験版）の公開が始まった[22]。日本ではこれまでにも、各分野/機関ごとに様々な体裁でデジタルコンテンツが構築・提供されていたが、これにより、今後はこのポータルサイトから国内の多様なコンテンツを備えるデータベース/アーカイブを横断検索できるようになる。また、各種文化資源（情報）のオープン化に加えて、メタデータの整備・集約等も行うことで、データの標準化も促進されていくという[23]。

このように、パブリック・ヒューマニティーズの広まりと並行して、オープンかつパブリックにデータ・学術情報（コンテンツ）を公開していく仕組みが構築されている。学術と公衆/市民をつなぎ、シームレスなアクセスを可能にすることで、知を共有していく取り組みが加速しているのである。

3. 「デザイン」に根ざしたアプローチ

話題は変わるが、筆者の知る限りでは、デジタル・ヒューマニティーズ界隈を俯瞰したとき、主流を占めているのは、次のいずれかの分野の研究者・専門家となる。1）人文学を専攻している/してきた者、もしくは 2）情報科学（コンピューターサイエンス）を専攻している/してきた者である。勿論、両者の知識や経験なくしてはデジタル・ヒューマニティーズの深化・発展は望めないが、ことパブリック・ヒューマニティーズにおいては、人文学を「開いていく」対象は公衆/市民（社会一般の人々）となる。そこで重要になるのが、その公衆/市民に対して「どのような経験・価値を提供したいのか」という考え方に立脚することである。つまり、ただ研究成果を（例えばオンライン上で）公開するのではなく、公衆/市民がどういう人たちで（構成され）、何を知りたいのか/求めているのかを考察しながら研究を実践していく–広義的な意味合いでの「デザイン」[24]に根ざした考え方が有用となる。

筆者はメディアデザインという分野を専攻し[25]、人文学あるいは情報科学（コンピューターサイエンス）を専門としている/してきたわけではない。しかし、だからこそ、そ

れらの分野を専門とする研究(者)と協働しながら、デザイン(研究)の立場から、パブリック・ヒューマニティーズの実践・発展に寄与できるものと考えている。公衆/市民といった「ユーザー」にどのような体験を提供できるのか、ユーザーエクスペリエンス(UX)の観点を中心としたデザインベースドの研究アプローチの有用性を模索・探求している。

4. 研究プロジェクトの紹介（パブリック・ヒューマニティーズの一例として）

ここでは、パブリック・ヒューマニティーズの一例として、筆者がこれまでに取り組んだ研究プロジェクト(概要)を紹介していく。紹介するプロジェクトは、筆者が博士課程で在籍していたメディアデザイン研究科（KMD)、ポスドク研究員として半年間在籍していたフィンランドのアールト大学、そして、2019年4月現在、研究員として在籍しているデジタルメディア・コンテンツ統合研究センター(DMC)で行われた/行われているものとなる。

Keio FutureLearn プロジェクト

2012 年に設立された FutureLearn （フューチャー・ラーン)[26]は、イギリス・ロンドンに本部を置くオンライン教育配信事業体となる。慶應義塾大学は 2015 年に配信協定を結び、これまでに合計 6 種類[27]のオンラインコースを配信しており[28]、筆者はプロジェクト発足時より、コースのデザイン・コンテンツ開発チームに従事している。なかでも、筆者の博士論文の主題材となったのが、日本の書物文化・歴史的典籍を紹介していくコース「Japanese Culture Through Rare Book/古書から読み解く日本の文化：和本の世界」である。

本コースは、世界各地の公衆/市民が−日本語や日本の歴史的典籍に関する知識の有無にかかわらず−体系的かつ網羅的に日本の書物文化・歴史的典籍について学ぶことができるよう、全体プログラムが設計されたが、筆者はそのうち、コース本体[29]と連動して展開された「歴史的典籍のデジタル展示モデル」を中心にデザインした。このデジタル展示モデル（「Narrative Book Collection」と名付けられた）では、歴史的典籍の画像や書誌情報といったデジタルデータを、ただ不文脈に羅列する[30]のではなく、各典籍が備える言語・非言語両側面の特性に応じて、コースのトピックと呼応した「ナラティブな文脈」を取り入れることで、世界中のコース受講生がより能動的にコースを楽しめる環境を構築した。

具体的には、慶應義塾図書館ならびに附属研究所斯道文庫が所蔵する日本の歴史的

典籍（8 世紀から 19 世紀にわたって制作された古書や貴重書）合計 146 点を題材に、それらの書物を制作年代や種類、形や大きさ、（表紙の）色合い等に応じて体系的に整理・分析したのち、デジタルデータならではの特性を生かして、視覚的に分かりやすく、且つ直感的なインタラクションを通じて書物（群）に関する理解を深めることができる Web インターフェースを創出、新しいかたちの（日本の歴史的典籍）鑑賞体験をデザインした[31,32]。（図 1 参照）

Semantic Biographies プロジェクト

筆者が2018年1～6月まで在籍していたフィンランド・アールト大学の Semantic Computing Research Groupでは、LODをはじめとするセマンティック（ウェブ）技術を活用したセマンティック（データ）インフラやオントロジー技術の開発、また、それら技術を用いたウェブアプリケーション・サービ

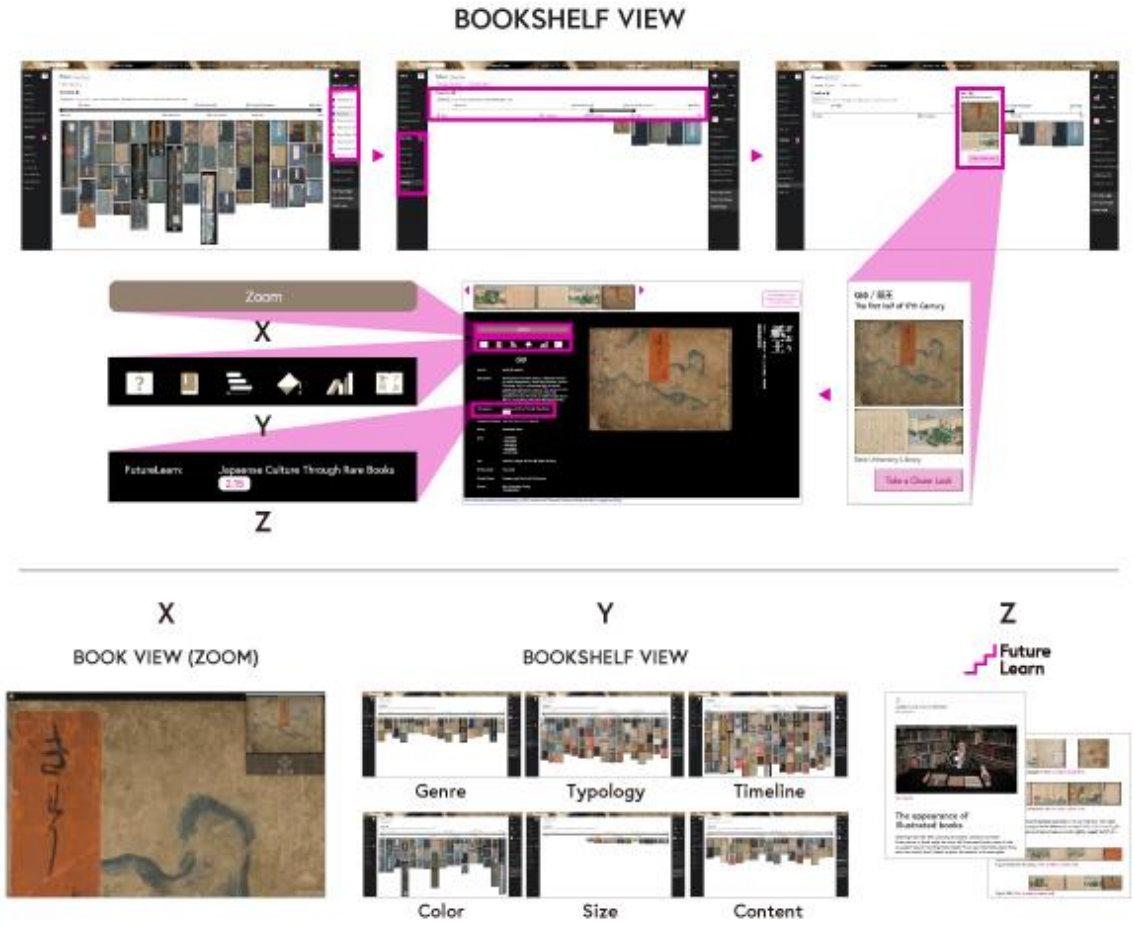


図1：Narrative Book Collection上での鑑賞体験の一例（本棚を模したBOOKSHELF VIEWからフィルター機能を駆使することで、BOOK VIEW［各書物の高精細画像と書誌情報を掲載］やFutureLearnコースを、興味のある文脈に沿って自在に探求できる）

スの開発を行なっている。[33] 研究内容は多岐にわたるが、筆者はそのうち、人物の伝記や人物間の関係（人・物・事との関係性）について、セマンティック技術を用いて研究するSemantic Biographies プロジェクトに携わった。

Semantic Biographies プロジェクトでは、LOD形式に変換されたバイオグラフィー（biography）およびプロソポグラフィー（prosopography）関連のデータ群をマッピングやテキスト/ネットワーク分析を通じて定量的・定性的に解析、可視化することで、人物に連なる歴史的資料の新しい価値・見方（解釈方法）を創出することを目的としている。同プロジェクトでは、Sampo[34]シリーズと題して、様々な人物のバイオグラフィー・プロソポグラフィーをLOD化し、デジタル・ヒューマニティーズ研究そのものにも役立てるよう、ファセットナビゲーションを組み込んだウェブアプリケーション（インターフェース）を制作、一般公開している。例えば、第二次世界大戦時におけるフィンランド軍事史を対象とした「WarSampo」[35]や約13,000におよぶフィンランド人物史伝を対象とした「Biografiasampo」[36]がある。

筆者は在籍中、現地の研究者と協働して「US Congress Prosopographer」[37]と名付けられたウェブアプリケーションを開発した。本プロジェクトでは、米国コロラド大学の協力も得て、米連邦議会が開会した第1回議会（1789年）から第115回議会（2018年）まで、約12,000の下院・上院議員のバイオグラ

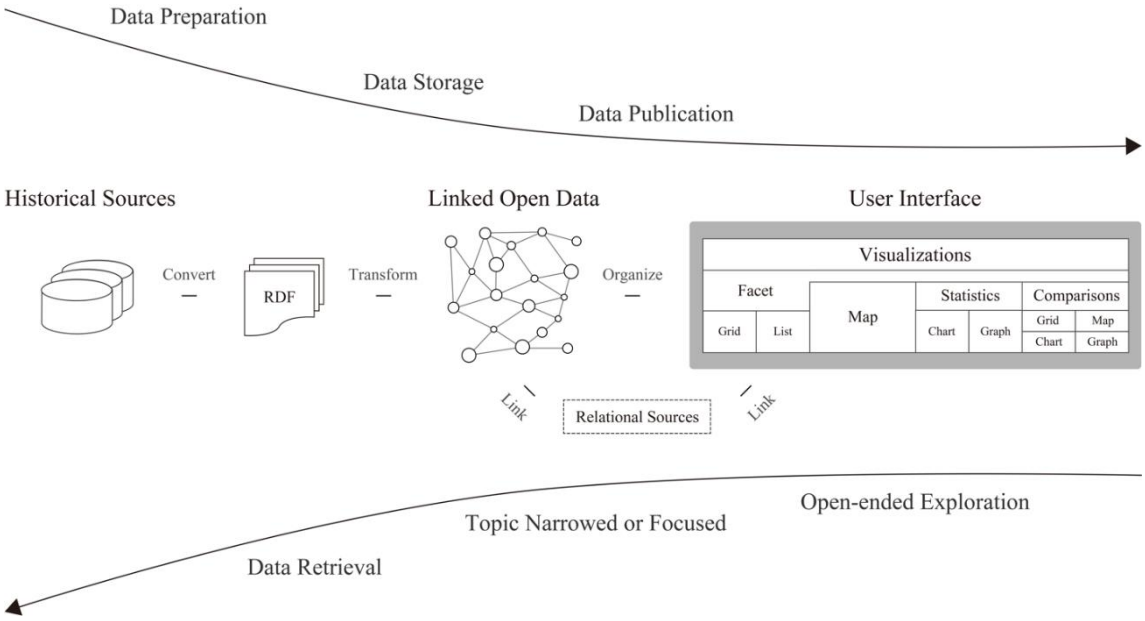


図 2：研究パイプライン（図上の矢印/項目は筆者のフロー、図下の矢印/項目はユーザーのフローを表す）

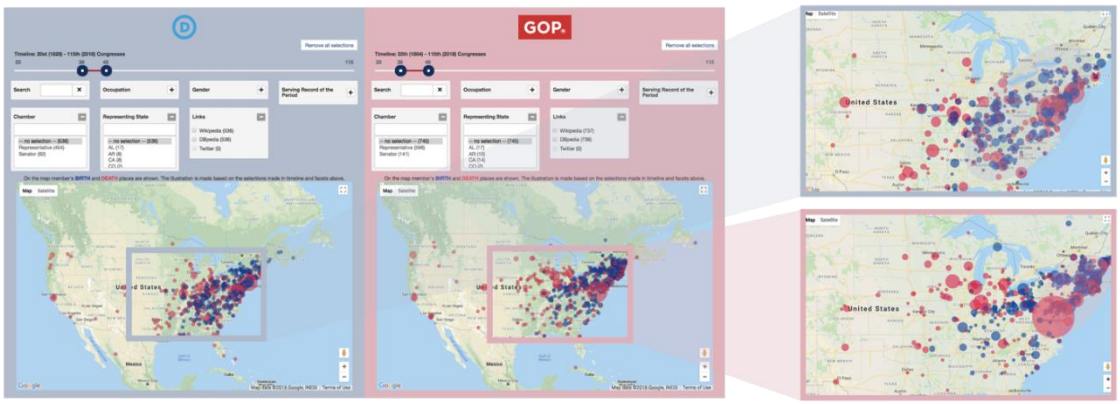


図 3：US Congress Prosopographer 上でのビジュアライゼーションの一例（民主党員[青]・共和党員[赤]の出生/死亡地の比較）

ンド軍事史を対象とした「WarSampo」したのち、各種ビジュアライゼーションを通じて、議員の在任期間や所属政党、出生/死亡地、二大政党（民主党員・共和党員）の類似/相違性について比較検証ができるアプリケーションを制作した[38]。研究期間の都合上、アプリケーション本体は試験版の公開で終わってしまったが、前述したSampoシリーズを含めて公開しているものはすべて、一般のユーザーでも特別な知識なしに利活用できるよう、簡易で操作性の高いインターフェースを実装している[39]。（図2, 3参照）

FutureLearn GLAM プロジェクト

本プロジェクトは、Keio FutureLearn プロジェクトから派生するかたちで発足した。GLAMはGalleries、Libraries、Archives、そしてMuseumsの頭字語から成り、文化・学術系の施設を総称する時に用いられるが、本プロジェクトでは、慶應義塾が有する書物・ツ上だけではなく、そうした実空間でも鑑賞できるように、オンライン空間と実空間とを横断して体験できる「場」づくりを推進している。

試験的な取り組みとなったが、2018 年 11 月 12 日から 12 月 14 日まで慶應義塾大学三田キャンパスで行われたセンチュリー文化財団寄託品展覧会「禅僧の書と書物」[40]では、展示作品の一部–FutureLearn から配信されたコース「Sino-Japanese Interactions Through Rare Books/古書から読み解く日本文化」で取り扱われた書物群–を大型タッチパネルディスプレイで紹介した。ここで来場者は、ガラスケースに陳列された書物を眺めるだけでなく、タッチやスワイプといった自然な操作を通じて（書物の）デジタル画像や（コースで公開されている）ビデオを閲覧することが可能となり、書物のもつ魅力を様々な視点から体験することができた。（図 4 参照）

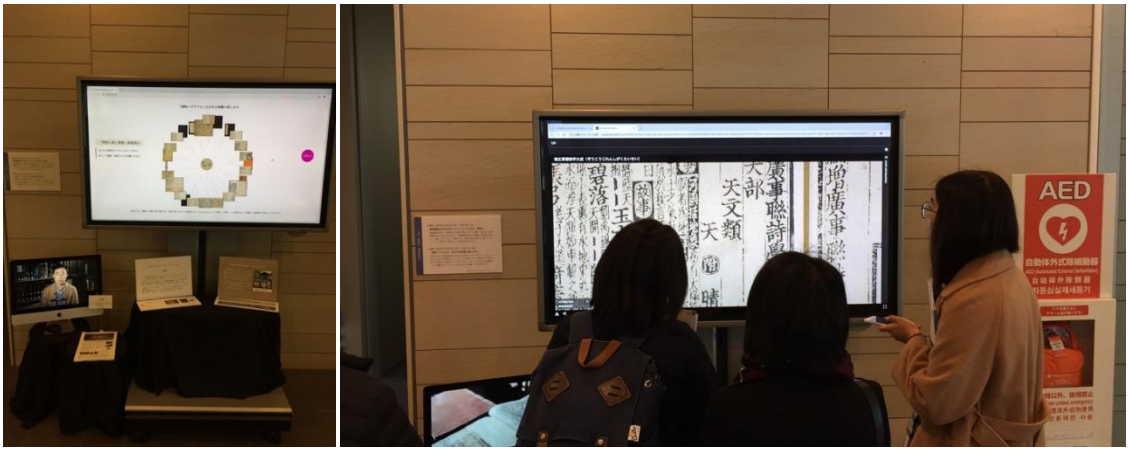


図 4:展覧会の会場前に設置された大型タッチパネルディスプレイ(展示)の模様

5. 今後の取り組み、展望

前述したプロジェクトに加えて、筆者が現在関わっている研究としては、FutureLearn の Book シリーズ[41]の新コースの制作・公開がある。本コースは、大英図書館との共同開発コースとして制作が進んでおり、慶應義塾・大英図書館が所蔵する書物(和書・洋書)を取り上げ、2020 年度中の配信を予定している。また、FutureLearn での取り組みからスピンアウトするかたちで、慶應義塾大学出版会が中心となり、江戸時代の文化・芸術の一翼を担った「名士層」[42]に焦点を当てたデジタルアーカイブの制作やデジタル表現を取り入れた展覧会の開催も今後予定されている[43]。

おわりに

デジタル・ヒューマニティーズを捉える際に「Methodological Commons (方法論の共有地)」という背景理念がある[44]。つまりは、様々な分野にまたがって方法論を共有し、複合的な視野をもつことで、新しい研究・成果物を生み出していけることを意味するが、本稿では、パブリック・ヒューマニティーズの文脈のなかで、デザイン(研究)の立場から、学術研究を公衆/市民にも開いていく取り組みの一端を紹介してきた。詳しい考察は別の機会に譲りたいが、膨大な数のデジタルデータが生産・蓄積されるインターネット前提時代において、オープンデータやオープンアクセスを基盤に、どのように人文学の知をパブリックに公開・共有し、公衆/市民とエンゲージしていけるのか。今後も「パブリック」と「デザイン」をキーワードに研究に取り組み、デザインの

視座から、パブリック・ヒューマニティーズの新しい可能性を創出していきたい。

宮北 剛己（みやきた ごうき）

DMC 研究センター研究員。専門は、デジタル人文学、セマンティック・ウェブ、情報デザイン、ユーザエクスペリエンス・デザイン。DMC では、MOOC コースのデザイン・コンテンツ開発に従事している。

[1] 村井 純. インターネット前提社会からの出発. 電気学会誌, 2017, 137 巻, 1 号, pp. 35-40.

[2] デジタル・ヒューマニティーズに関する論考には、例として、以下がある。
Susan Schreibman, Ray Siemens, John Unsworth ed. A Companion to Digital Humanities. Oxford, Blackwell, 2004. http://www.digitalhumanities.org/companion/, (accessed 2019-04-23.)
Anne Burdick, Johanna Drucker, Todd Presner, and Jeffrey Schnapp. A Short Guide to the Digital Humanities. MIT Press, 2012.
Matthew Kirschenbaum. What Is “Digital Humanities,” and Why Are They Saying Such Terrible Things About It?. differences, 1 May 2014, 25 (1), pp. 46–63.

[3] 国内外の最新の動向は、一般財団法人人文情報学研究所の永崎研宣氏による論考やブログ、発行メールマガジンに詳しい。
digitalnagasakiのブログ. http://digitalnagasaki.hatenablog.com/, (参照 2019-04-24.)
人文情報学月報. http://www.dhii.jp/DHM/, (参照 2019-04-24.)

[4] Matthew Jockers and Glen Worthey. Introduction: Welcome to the Big Tent, n. d., Digital Humanities 2011 Abstracts, http://dh2011abstracts.stanford.edu/xtf/view?docId=tei/ab-005.xml, (accessed 2019-04-24.)

[5] Stanley Fish. The Old Order Changeth. opinionator.blogs.nytimes.com. New York Times, 26 Dec 2011. https://opinionator.blogs.nytimes.com/2011/12/26/the-old-order-changeth, (accessed 2019-04-24.)

[6] 『デジタル・ヒューマニティーズ』発刊に際して. デジタル・ヒューマニティーズ, 2018, 1 巻, p. 1.

[7] パブリック・ヒューマニティーズに関する論考には、例として、以下がある。
Julie Ellison. Guest Column: The New Public Humanists. PMLA, 128, 2013, pp. 289-298.
Seven Rules for Public Humanists. Public Humanities Blog. John Nicholas Brown Center for Public Humanities and Cultural Heritage, 29 Aug 2014. https://www.brown.edu/academics/public-humanities/blog/seven-rules-public-humanists, (accessed 2019-04-24.)
Wendy F. Hsu. Lessons on Public Humanities from the Civic Sphere. Debates in the Digital Humanities 2016, Ed. Mathew K. Gold and Lauren F. Klein, Minneapolis, University of Minnesota Press, 2016, pp. 280-286.

[8] Sheila A. Brennan. Public, First. Debates in the Digital Humanities 2016, Ed. Mathew K. Gold and Lauren F. Klein, Minneapolis, University of Minnesota Press, 2016, pp. 384-389.

[9] 例えば、米国ブラウン大学の John Nicholas Brown Center for Public Humanities and Cultural Heritage やイエール大学 Public Humanities at Yale、英国シェフィールド大学では、専門的に学べるコースが提供されている。

[10] 本稿では詳細に触れないが、パブリック・ヒストリーという言葉/概念自体が、パブリック・ヒューマニティーズと互換的に使われることも多い。

[11] 最近刊行された包括的論文集として 2018 年刊「David M. Dean, ed. A Companion to Public History. Newark, John Wiley & Sons, Incorporated.」がある。また、同 2018 年には International Federation for Public History（IFTP）から、電子ジャーナル International Public History（IPH）の初号が刊行されている。

[12] 日本では、国内複数の研究者が集まり、2019年3月に「パブリックヒストリー研究会」第一回会合が開催されている。
https://public-history9.webnode.jp/, (参照 2019-04-24.)

[13] オープンデータとは何か？. OPEN DATA HANDBOOK. http://opendatahandbook.org/guide/ja/what-is-open-data/, (参照 2019-04-24.)

[14] Read the Budapest Open Access Initiative. Budapest Open Access Initiative. http://www.soros.org/openaccess/read, (accessed 2019-04-24.)

[15] 本稿ではごく一部を紹介している。2019 年 4 月に刊

行された「国立歴史民俗博物館監修 後藤 真・橋本雄太編 『歴史情報学の教科書　歴史のデータが世界をひらく 』 （文学通信）」では、様々な分野の動向や事例が網羅的にまとめられている。

16 International Image Interoperability Framework. https://iiif.io/, (accessed 2019-04-18.)

17 つながる世界のコンテンツ──IIIF が描くアート・アーカイブの未来. デジタルアーカイブスタディ. アートスケープ. 2017 年 10 月 15 日号. https://artscape.jp/study/digital-achive/10139893_1958.html, (参照 2019-04-24.)

18 本稿では詳細な言及は避けるが、技術概要や国内での取り組みについては、国立情報学研究所の武田英明氏による論考に詳しい。例えば「オープンデータの最前線―データの Web を実現する LOD と DOI―. 特集：オープンサイエンス　開かれたデータの可能性, NII Today, 第 71 号, 2016-03.」がある。

19 5 ★ オープンデータ. https://5stardata.info/ja/, (参照 2019-04-17.)

20 Europeana Collections. https://www.europeana.eu/portal/en, (accessed 2019-04-25.)

21 Digital Public Library of America. https://dp.la/, (accessed 2019-04-18.)

22 ジャパンサーチ (BETA). https://jpsearch.go.jp/, (参照 2019-04-24.)

23 非公式ではあるが、試験版の公開にあわせて、セマンティック・ウェブ技術研究者の神崎正英氏による「Japan Search 非公式サポートページ」も開設され、LOD のデータモデル/記述の際の枠組みである RDF（Resource Description Framework）の活用方法等が説明されている。
Japan Search 非公式サポートページ - ジャパンサーチ [利活用スキーマ] unofficial support page. https://www.kanzaki.com/works/ld/jpsearch/, (accessed 2019-04-18.)

24 今日、デザインという言葉が示す範囲は広がりをみせ、表現（形態）のみならず、様々な理論や実践手法が確立されている。またデザインに関する論考も多く存在するが、例えば、ユーザーの立場・見方も記した先駆的な著書として「D. A. ノーマン. 誰のためのデザイン? ― 認知科学者のデザイン原論. 野島久雄訳, 新曜社, 1990.」がある。

25 慶應義塾大学大学院メディアデザイン研究科の修士・博士課程に在籍していた。

26 Free online courses from Top Universities - FutureLearn. https://www.futurelearn.com/, (accessed 2019-04-24.)

27 2019 年 4 月時点。

28 各コースの開発過程やコース内容等の詳細は、以下の論考に詳しい（初出順）。
大川恵子. 慶應から世界へ──グローバルなオンライン公開講座の開講. KEIO Report, 三田評論, No.1207, pp. 118-119, 2017-01.
高信彰徳. FutureLearn プロジェクトにおけるオンラインコースの開発. 慶應義塾大学 DMC 紀要, 4(1), pp. 34-41, 2017-03.
村上篤太郎. FutureLearn 発グローバル展開―義塾のオンライン教育―. 塾監局紀要, 第 33 号, pp. 36-39, 2019-03.

29 各コース（本体）は、ビデオ、アーティクル（テキスト）、ディスカッション、クイズ等の「ステップ」を組み合わせて構成される。

30 学術研究利用を主目的に開発されたアーカイブ・レポジトリの多くは、例えば、テキスト検索結果にも基づいた書誌情報・画像のリスト表示―といったかたちで、表現様式が限定されることが多い。

31 本稿では詳細な言及は控えるが、KMD が提唱している「MAKE、DEPLOY、IMPACT」のプロセスに沿うかたちで研究は進められた。また、研究の詳細に関して興味のある方は、別稿（以下拙稿）を参照されたい。
拙稿. Design of Narrative Book Collection Combining Computational and Visualization Approach― Exploring Japanese Culture Through Keio University's Pre-Modern Japanese Book Collection―. 慶應義塾大学大学院メディアデザイン研究科, Doctoral Dissertation, 2018-03.

32 FutureLearn では、一度配信が完了したコースも（一定期間のインターバルを置きながら）継続して再配信を行なっている。また、再配信にあわせてデジタル展示モデル「Narrative Book Collection」も改良を重ねており、次回は 2019 年後期に再配信/改良を予定している。

33 Semantic Computing Research Group (SeCo). https://seco.cs.aalto.fi/, (accessed 2019-04-22.)

34 Sampo とは、フィンランド神話に登場する宝物のことを指すという。

35 WarSampo. http://www.sotasampo.fi/en/, (accessed 2019-04-22.)

36 Biografiasampo. http://biografiasampo.fi/, (accessed 2019-04-22.)

37 US Congress Prosopographer (試験版). https://semanticcomputing.github.io/congress-legislators/, (accessed 2019-04-22.)

38 Goki Miyakita, Petri Leskinen, and Eero Hyvönen. Using Linked Data for Prosopographical Research of Historical Persons: Case U.S. Congress Legislators. 7th International Conference, EuroMed 2018, Nicosia, Cyprus, October 29 – November 3, 2018, Proceedings, Part II.

39 プロジェクト（内容）とは直接関連しないが、Semantic Computing Research Group を率いる Eero Hyvönen 氏は、ヘルシンキ大学に設置されているデジタル・ヒューマニティーズのリサーチセンター「HELDIG（Helsinki Centre for Digital Humanities）」のディレクターも務めており、フィンランド国内外で精力的にデジタル・ヒューマニティーズ研究・教育の普及活動を行なっている。

Helsinki Centre for Digital Humanities | University of Helsinki. https://www.helsinki.fi/en/helsinki-centre-for-digital-humanities, (accessed 2019-04-25.)

[40] 慶應義塾大学附属研究所斯道文庫、慶應義塾大学アート・センター、慶應義塾図書館の 3 機関が主催。

[41] Book に関連するコースとしては、これまでに「Japanese Culture Through Rare Book/古書から読み解く日本の文化：和本の世界」、「Sino-Japanese Interactions Through Rare Books/古書から読み解く日本文化」、「The Art of Washi Paper in Japanese Rare Books/和本を彩る和紙の世界」、以上 3 コースを制作・開講しており、総じて Book シリーズと呼称している。

[42] 足立郷土博物館学芸員の多田文夫氏によると、名士層とは、江戸時代の「文人文化の担い手」のことを指す。本研究を推進しているメンバーである慶應義塾大学大学院システムデザイン・マネジメント研究科（SDM）研究所の日比谷孟俊氏の祖先にあたる日比谷健次郎氏もその一人とされる。

[43] 慶應義塾大学出版会の安井元規氏、実践女子大学の白戸満喜子氏を中心に活動しており、「足立における郷土の歴史と文化に関するワークショップ」等で成果発表を行っている。

[44] 「Methodological Commons」という用語は、はじめ「Willard McCarty and Harold Short. Methodologies. Pisa, 2002-04.」で用いられ、その後、広く定着したとされる。

記録

会議記録

●DMC ミーティング

4 月 26 日（木）、6 月 1 日（金）、6 月 26 日（火）、7 月 27 日（金）、9 月 4 日（火）、10 月 3 日（水）、10 月 31 日（水）、11 月 27 日（火）、1 月 25 日（金）、2 月 25 日（月）、3 月 19 日（火）

●拡大 DMC ミーティング

4 月 26 日（木）、6 月 1 日（金）、6 月 26 日（火）、7 月 27 日（金）、9 月 4 日（火）、10 月 3 日（水）、10 月 31 日（水）、11 月 27 日（火）、1 月 25 日（金）

●オンライン教育委員会

10 月 26 日（金）

●Future Learn Meeting

4 月 10 日（火）、4 月 27 日（金）、6 月 4 日（月）、6 月 13 日（水）、6 月 29 日（金）、8 月 31 日（金）、9 月 26 日（水）、9 月 28 日（金）、10 月 15 日（月）、10 月 30 日（火）、11 月 9 日（金）、11 月 16 日（金）、1 月 24 日（木）、2 月 20 日（水）

記録

研究・教育活動業績

(2018 年 1 月～12 月)

凡例＝本記録は研究員による研究・教育活動の業績一覧であり、研究員の投稿に、もとづくものである。

1．著書・訳書、2．論文、3．学会発表，4．講演・展覧会・ワークショップ等、5．その他

重野 寛（所長　理工学部教授）

2. 論文

- 小林諒二郎, 篠原涼希, 重野寛: Named Data Networking における Content Poisoning Attack 対策手法, 情報処理学会論文誌 Vol.60 No.2, pp. 440-pp.448, 2019 年 2 月.
- 小林裕樹, 西山潤, 谷遼太郎, 重野寛: 帯域状況に基づく PoI を考慮した被災地情報収集機構, 情報処理学会論文誌 Vol.60 No.2, pp.459-pp.468, 2019 年 2 月.

3. 学会発表

- Ryotaro Tani, Yuki Kobayashi, Hiroshi Shigeno, "Offloading System Based on Estimated Response Time in Multi-tier Environment," IEEE Consumer Communications & Networking Conference (CCNC 2019 Work-in-Progress Paper), 11-14 January 2018.
- Chiaki Doi, Masaji Katagiri, Takashi Araki, Daizo Ikeda, Hiroshi Shigeno, “Is he becoming an excellent customer for us? Customer level prediction method for customer relationship management system”, The 32nd IEEE International Conference on Advanced Information Networking and Applications (AINA-2018), pp.320-326, May 2018.

4. 講演、展示会、ワークショップ

- 谷遼太郎, 小林裕樹, 重野寛, "複数層環境における推定応答時間に基づくオフローディングシステムの検討", 第 26 回 マルチメディア通信と分散処理ワークショップ (DPSWS2018), 2018 年 11 月.
- 小林諒二郎, 篠原涼希, 重野寛: Named Data Networking における Interest のフラグを用いた Content Poisoning Attack 攻撃検知の検討，情報処理学会 マルチメディア, 分散, 協調とモバイル(DICOMO 2018)シンポジウム，pp. 1652-1657，2018 年 7 月

安藤 広道（副所長　研究員　文学部教授）

4．講演、展覧会、ワークショップ等

講演

- 「第五航空艦隊司令部壕の調査成果」鹿屋平和学習ガイド・戦争遺跡調査員主催『あの日を忘れない･･･3･18 鹿屋空襲によせて』鹿屋市中央公民館　2018 年 3 月 18 日
- 「歴史を伝える･･･だけでいいのか」2018 年度極東証券株式会社寄附講座・慶應義塾大学文学部公開講座『橋渡しする文学部』慶應義塾大学三田キャンパス　2018 年 6 月 30 日
- 「近現代考古学の可能性－社会に開かれた歴史をめざして－」大阪経済大学日本経済史研究所『黒正塾』大阪経済大学　7 月 21 日
- 「日吉と鹿屋－鹿屋市第五航空艦隊司令部地下壕の調査成果から－」慶應義塾福澤千九センター主催　学徒出陣 75 年シンポジウム・研究報告『慶應義塾と戦争』慶應義塾大学三田キャンパス　12 月 2 日

展覧会

- 『日吉と矢上の考古学－縄文土器から焼夷弾まで』　開催期間 1 月 12 日～2 月 8 日　慶應義塾大学三田キャンパス図書館（分担）

ワークショップ等

- 「久ヶ原遺跡ツアー」大田区立郷土博物館　3 月 21 日

5. その他

- 展示図録『日吉と矢上の考古学－縄文土器から焼夷弾まで』慶應義塾大学民族学考古学研究室　1 月 12 日（編著）
- パネルディスカッション「ボーダレスなデータ利活用のための情報組織化とは」慶應義塾大学 DMC 研究センターシンポジウム『メタデータ再考』慶應義塾大学日吉キャンパス西別館　11 月 20 日（モデレーター）

大川恵子（副所長　メディアデザイン研究科教授）

5. その他

- "Understanding Quantum Computers" Course Run2, 16 Apr 2018 - 4 Weeks
- "Japanese Culture Through Rare Books" Course Run5, 28 May 2018 - 3 Weeks
- "The Art of Washi Paper in Japanese Rare Books" Course Run1, 16 Jul 2018 - 2 Weeks
- "An Introduction to Japanese Subcultures" Course Run4, 30 Jul 2018 - 4 Weeks
- "Understanding Quantum Computers" Course Run3, 1 Oct 2018 - 4 Weeks
- "Exploring Japanese Avant-garde Art Through Butoh Dance" Course Run1, 24 Sep 2018 - 4 Weks
- "Exploring Japanese Avant-garde Art Through Butoh Dance" Course Run2,7 Jan 2019 - 4 Weks
- "Japanese Culture Through Rare Books" Course Run6, 14 Jan 2019 - 3 Weeks
- "Sino-Japanese Interactions Through Rare Books" Course Run3, 18 Feb 2019 - 4 Weeks

小菅 隼人（研究員　理工学部教授）

1. 著書

- Kosuge, Hayato. “The Expanding Universe of Butoh: the Challenge of Bishop Yamada in Hoppo Butoh-ha and Shiokubi (1975).” in *The Routledge Companion to Butoh Performance*. Routledge, 2018. pp. 214-25.（単著分担）

2. 論文

- 小菅隼人．「祈りとしての舞踏—舞踏家大野慶人に聞く」．『日吉紀要：言語・文化・コミュニケーション』50 号，慶應義塾大学日吉紀要刊行委員会，2018/12/31, PP. 19-51.（単著）
- 小菅隼人．「漆黒の闇から純白の拡がりへ—舞踏家雪雄子に聞く—」．『人文科学第 33 号：慶應義塾大学日吉紀要 H-33』49 号，慶應義塾大学日吉紀要刊行委員会，2018/06/30，pp. 35-77.（単著）
- 小菅隼人．「『カフェ・ミュラー』をアーカイブ映像で見ることの意味について」．『舞台芸術 21：アーカイブを批評する』21 号，角川文化振興財団，21 号，2018/03/25，pp. 120-22.（単著）

3. 国際学会発表

- Kosuge, Hayato. “Curated Panel: The Theatre of Ghosts and the Other under the Threat of Mass Deaths.” IFTR 2018 Migration and Stasis (9-13 July, 2018), U of Belgrade, Serbia. 2018/07/11.（共著）
- Kosuge, Hayato. “Curated Panel: Migrating/Migrated Bodies in Japanese Context.” IFTR 2018 Migration and Stasis (9-13 July, 2018), U of Belgrade, Serbia. 2018/07/10.（共著）
- Kosuge, Hayato. “Theorizing Asian Bodies in Performance.” IFTR International Federation for Theatre Research 2018, Regional conference, UP Diliman, Manila. 2018/02/23.（共著）

金子 晋丈（研究員　理工学部専任講師）

2. 論文

- 金子晋丈, "柔軟広範かつ長期的な利活用のためのデジタル情報のネットワーク化," 信学技報, vol. 118, no. 19, IN2018-1, pp. 1-6, 2018 年 5 月.
- T. Kondo, S. Yoshihara, K. Kaneko, and F. Teraoka, "ZINK: An Efficient Information Centric Networking Utilizing Layered Network Architecture," IEICE TRANSACTIONS on Communications, Vol.E101-B, No.8, pp.1853-1865, 2018 年 8 月.
- R. Ogitani and K. Kaneko, "Activation Server for Bidirectional Limited Usages of Digital Objects," in Proc. of 2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech), Athens, 2018, pp. 1008-1015.
- doi: 10.1109/DASC/PiCom/DataCom/CyberSciTec.2018.00143
- H. Watanabe, T. Kondo, K. Kaneko, F. Teraoka, "Inserting Layer-5 to provide applications with richer functions through common API," IEICE TRANSACTIONS on Communications, Vol.E101-B, No.9, pp.1967-1981, 2018 年 9 月.
- 上村優介, 金子晋丈, "Catalogue System におけるパーソナライズサービスのための部分グラフ決定手法", 研究報告マルチメディア通信と分散処理(DPS), 2019-DPS-177(6),1-8, 2019 年 1 月.
- 荻谷凌, 金子晋丈, "Virtual File：認証認可とファイル取得の独立制御を可能にする広域データ流通フォーマット," 信学技報, vol. 118, no. 466, IN2018-127, pp. 259-264, 2019 年 3 月.

池田 真弓（研究員　理工学部専任講師）

1. 著書

- （共著）池田真弓「初期印刷本の装飾方法––一四五九年マインツ出版『聖務の理論』を例に」『移ろう形象と越境する芸術』八坂書房、2019 年 3 月。

4. 講演

- 講演「インキュナブラの装飾と挿絵」、第３０回慶應義塾図書館貴重書展示会「インキュナブラの時代：慶應義塾の西洋初期印刷コレクションとその広がり」、2018 年 10 月 6 日。

石川 尋代（ＤＭＣ研究センター特任講師）

3. 学会発表

- 石川 尋代, 本間友, 金子晋丈,「MoSaIC による資料間関係の可視化―イタリア・ルネサンス祭壇画研究への利用―」人文科学とコンピュータシンポジウム論文集(じんもんこん 2018)、情報処理学会シンポジウムシリーズ Vol. 2018, No.1, pp.33-38.

4. 講演、展覧会、ワークショップ等

- 石川 尋代, 金子 晋丈, 松田 隆美,「MoSaIC:Connecting digital contents by various contexts：様々なコンテクストによってデジタルコンテンツを繋ぐ」国際シンポジウム「デジタル時代における人文学の学術基盤をめぐって」, 2018 年 7 月 6 日.

5. その他

- 石川 尋代「様々なコンテクスト：エンジニアリングからデジタル人文学へのアプローチ」人文情報学月報第 85 号【後編】巻頭言, Digital Humanities Monthly No. 085-2, 2018 年 8 月 31 日発行.

杉浦 裕太（研究員　理工学部専任講師）

2. 論文

- 小林巧，家永直人，杉浦裕太，斎藤英雄，宮田なつき，多田充徳，距離カメラ付きスマートフォンを用いた簡易的な人体の足の 3 次元形状計測システム，精密工学会誌，Vol.84，No.12，996-1002，2018-12-5.

3. 学会発表

- Yuta Sugiura, Hikaru Ibayashi, Toby Chong, Daisuke Sakamoto, Natsuki Miyata, Mitsunori Tada, Takashi Okuma, Takeshi Kurata, Takashi Shinmura, Masaaki Mochimaru, and Takeo Igarashi, An asymmetric collaborative system for architectural-scale space design, In Proceedings of the 16th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry (VRCAI '18), ACM, Article 21, 6 pages, December 2-3, 2018, Tokyo, Japan.
- Yuta Sugiura, Toby Chong, Wataru Kawai, and Bruce H. Thomas, Public/private interactive wearable projection display, In Proceedings of the 16th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry (VRCAI '18), ACM, Article 10, 6 pages, December 2-3, 2018, Tokyo, Japan.
- Moeko Iwasaki, Suzanne Low, Mitsunori Tada, Yuta Sugiura, Hideo Saito, Maki Sugimoto, 3D Shape Reconstruction of Human Foot using Distance Sensors, In Proceedings of the SICE Annual Conference 2018, 6 pages, September 11-14, 2018, Nara, Japan.
- Kentaro Ino, Naoto Ienaga, Yuta Sugiura, Hideo Saito, Natsuki Miyata, Mitsunori Tada, Grasping Hand Pose Estimation from RGB Images Using Digital Human Model by Convolutional Neural Network,in Proc. of 3DBODY.TECH 2018 -9th Int. Conf. and Exh. on 3D Body Scanning and Processing Technologies, 154-160, October 16-17, 2018, Lugano, Switzerland.
- Takumi Kobayashi, Naoto Ienaga, Yuta Sugiura, Hideo Saito, Natsuki Miyata, Mitsumori Tada, A Simple 3D Scanning System of the Human Foot Using a Smartphone with a Depth Camera, in Proc. of 3DBODY.TECH 2018 - 9th Int. Conf. and Exh. on 3D Body Scanning and Processing Technologies, 161-169, October 16-17, 2018, Lugano, Switzerland.

記録

活動実績

●4 月 11 日（水）
FutureLearn 舞踏収録（三田）

●4 月 12 日（木）
FutureLearn 舞踏収録（三田）

●4 月 13 日（金）
FutureLearn 量子コンピュータ入門収録（SFC）
FutureLearn 舞踏収録（三田）

●4 月 14 日（土）
慶應長寿クラスター　ポリシーダイアログ・ミーティング収録（信濃町）

●4 月 16 日（月）
FutureLearn 舞踏収録（三田）

●4 月 16 日（月）～17 日（火）
大学院政策・メディア研究科紹介ビデオ収録（TTCK）

●4 月 18 日（水）
FutureLearn 舞踏収録（三田、上星川）

●4 月 19 日（木）
FutureLearn 舞踏収録（三田）

●4 月 20 日（金）
FutureLearn 舞踏収録（三田）
KGRI Lecture Series 講演収録 ″生物学応用のための実用的な in vitro 技術プラットフォーム″（矢上）

●4 月 25 日（水）
FutureLearn 舞踏収録（東京大学）

●5 月 9 日（水）
情報の教養学（福井健策氏「ネットのダークマター：止まらない海賊版で、マンガ・アニメは滅びるのか？」）収録（日吉）

●5 月 10 日（木）
医学部広報ビデオ収録（信濃町）

●5 月 12 日（土）
慶應看護 100 年記念式典収録（信濃町）

●5 月 16 日（水）
FutureLearn 舞踏収録（古川橋）

●5 月 21 日（月）
情報の教養学（堀　潤氏「フェイクニュースと民主主義：いま私たちが為すべきこと」）収録（日吉）

●5 月 24 日（木）
BPA（英国パラリンピック委員会）との覚書（MOU）締結式収録（日吉）

●5 月 25 日（金）
早稲田大学大学総合研究センター、9 名来訪

●5 月 28 日（月）
FutureLearn 舞踏収録（三田）
APRU 研究プロジェクト紹介インタビュー（國領常任理事）収録（三田）

●6 月 5 日（火）
APRU 研究プロジェクト紹介インタビュー（中谷教授、ショウ教授）収録（三田）

●6 月 8 日（金）
アーミテージ氏への慶應義塾大学名誉博士称号授与式&講演会収録（三田）
雪雄子舞踏公演収録（日吉）

●6 月 17 日（日）
FutureLearn 舞踏収録（三田）

●7 月 6 日（金）
国際シンポジウム「デジタル時代における人文学の学術基盤をめぐって」ポスター／デモンストレーション「MoSaIC: 様々なコンテクストによってデジタルコンテンツを繋ぐ」参加：特任講師 石川尋代（一橋講堂）

●7 月 12 日（木）
塾長ビデオメッセージ収録（三田）

●7 月 19 日（木）
KGRI Lecture Series 講演収録"超撥水表面の紹介"（矢上）

●8 月 3 日（金）
KGRI Lecture Series 講演収録 （矢上）
#1 "中枢神経系における細胞・薬剤導入のための注入可能な生体高分子"
#2 "ヘルスケアのための体内埋め込み型マイクロデバイス"

●9 月 4 日（火）
FutureLearn 舞踏収録（三田）

●9 月 11 日（火）、9 月 12 日（水）
KGRI Lecture Series 講演収録 （矢上）
「顕微鏡技術とレーザーを用いたイメージング」
「レーザーを用いた細胞・生体組織のマニピュレーションとその応用」

●9 月 21 日（金）
KGRI Lecture Series 講演収録 "前頭側頭型認知症におけるタウの出現"（信濃町）

●9 月 26 日（水）
AV ホール自動収録・配信（文学部・基礎情報処理）

●10 月 3 日（水）
AV ホール自動収録・配信（文学部・基礎情報処理）

●10 月 5 日（金）
第 30 回慶應義塾図書館貴重書展示会ギャラリー・トーク収録
（丸善・丸の内本店４階ギャラリー）

●10 月 9 日（火）
AV ホール自動収録・配信（理工学部・英語スピーキング３）

●10 月 10 日（水）
AV ホール自動収録・配信（文学部・基礎情報処理）

●10 月 16 日（火）
AV ホール自動収録・配信（理工学部・英語スピーキング３）

●10 月 17 日（水）
AV ホール自動収録・配信（文学部・基礎情報処理）

●10 月 31 日（水）
情報の教養学（岡田美智男氏「〈弱いロボット〉的思考のすすめ―他力本願なロボットがひらく、弱いという希望、できないという可能性」）収録（日吉）

●11 月 14 日（水）
KGRI Lecture Series 講演収録
"Volumetric imaging of whole-tumors reveals cancer malignancy"（信濃町）

●11 月 16 日（金）
KGRI Lecture Series 講演収録 （三田）
"22 世紀のグランドデザインを見据えて現在の課題を深く洞察する"

●11 月 20 日（火）
「DMC シンポジウムー第８回デジタル知の文化的普及と深化に向けてーメタデータ再考」開催・収録（DMC 展示室）

●11 月 22 日（木）、11 月 23 日（金）
ORF2018PITCH の収録（東京ミッドタウン）

●11 月 27 日（火）
AV ホール自動収録・配信（理工学部・テクニカル・コミュニケーションⅡ）

●12 月 1 日（土）
じんもんこん 2018 人文科学とコンピュータシンポジウムにて研究発表
（東京大学 地震研究所）参加：特任講師 石川尋代

●12 月 4 日（火）
AV ホール自動収録・配信（理工学部・テクニカル・コミュニケーションⅡ）

●12 月 6 日（木）
KGRI Lecture Series 講演収録（三田）
″2018 年米国中間選挙：米国のアジア政策へのインプリケーション″

●12 月 7 日（金）
KGRI Great Thinker Series 講演収録 ″サイバー文明のプロローグ″（三田）

●12 月 12 日（水）
AV ホール自動収録・配信（文学部・基礎情報処理）

●12 月 19 日（水）
AV ホール自動収録・配信（文学部・基礎情報処理）

●12 月 21 日（金）
KGRI Lecture Series 講演収録（三田）
"Political Transition and News Media: Dynamics of Media-Party Parallelism in South Korea"

●12 月 26 日（水）
AV ホール自動収録・配信（文学部・基礎情報処理）

●12 月 27 日
慶應義塾大学病院１号館広報ビデオ
編集完了（信濃町）

●1 月 9 日（水）
AV ホール自動収録・配信（文学部・基礎情報処理）

●1 月 15 日（火）
AV ホール自動収録・配信（理工学部・テクニカル・コミュニケーションⅡ）

●1 月 16 日（水）
AV ホール自動収録・配信（文学部・基礎情報処理）

●1 月 17 日
SFC 渡辺賢治教授最終講義収録（湘南藤沢）

●1 月 21 日
SFC 渡辺利夫教授最終講義収録（湘南藤沢）

●1 月 22 日
SFC 渡邊頼純教授最終講義収録（湘南藤沢）

●1 月 23 日～2 月 4 日
キャンパス紹介映像制作収録（入学センター・全キャンパス）

●2 月 5 日
塾長・メルケル首相対談収録（三田）

●3 月 8 日
看護医療学部
近藤好枝教授、鎌倉光弘教授最終講義収録（湘南藤沢）

●3 月 30 日
SFC 村林裕教授最終講義収録（三田）

名簿

研究員・職員

2019 年 3 月 31 日現在

所長　重野　寛　慶應義塾大学　理工学部教授
（Ph.D.）
副所長　大川恵子　慶應義塾大学　大学院メディアデザイン研究科教授
（Ph.D.）
安藤広道　慶應義塾大学　文学部教授
研究員　松田隆美　慶應義塾大学　文学部教授
（Ph.D.）
小菅隼人　慶應義塾大学　理工学部教授
金子晋丈　慶應義塾大学　理工学部専任講師
（Ph.D.）
池田真弓　慶應義塾大学　理工学部専任講師
（Ph.D.）
杉浦裕太　慶應義塾大学　理工学部専任講師（9 月 12 日～）
（Ph.D.）
石川尋代　慶應義塾大学　DMC 研究センター特任講師
（Ph.D.）
宮北剛己　慶應義塾大学　DMC 研究センター研究員（7 月 1 日～）
（Ph.D.）

専門員　岡田　豊史
事務長　村上篤太郎（～10 月 31 日）
落合　啓一（11 月 1 日～）
職員　鈴木美佐子（～5 月 31 日）
小川　　武 (6 月 1 日～)
山本　真紀

以　上

編集後記

安藤広道
慶應義塾大学 DMC 研究センター副所長
文学部教授

DMC 紀要第 6 号をお届けします。

「デジタル知の文化的普及と深化に向けて」と題するシンポジウムも第 8 回目となりました。今年度は、私立大学戦略的研究基盤形成支援事業が終了し、これまで駆け足で進めてきた MoSaIC を中心とするプロジェクトを振り返るいい機会になったと思います。そこで今回は、プロジェクトの発想の原点に立ち返り、メタデータの問題を取り上げることにしました。情報の蓄積と利活用において、メタデータに基づく情報の組織化は不可欠ですが、一方でそれは利用者の思考を縛ることにもなるある種の領域を形成してしまいます。その領域のボーダーをどのように乗り越えていくのか、熱い議論をご覧いただきたいと思います。

当センターのプロジェクトのもうひとつの柱である MOOCs（FutureLearn）では、今年度も新たに The Art of Washi Paper in Japanese Rare Books と Exploring Japanese Avant-garde Art Through Butoh Dance の 2 コースを加えることができました。昨年度までの 4 コースの再配信も含め、ラインナップがさらに充実したことで、世界中で着実に受講者が増えています。本号では、今年度から研究員に加わった宮北剛己さんが、パブリック・ヒューマニティーズの観点から、学知を世界に開いていく MOOCs の意義と課題をまとめてくださいました。大学の国際的な情報発信力がますます重要になっている今日、FutureLearn プロジェクトには、塾内はもちろん塾外からも大きな期待が寄せられることになるでしょう。

本年度から、ヒューマンコンピュータインタラクションがご専門の杉浦裕太さんにも、DMC のメンバーに加わっていただきました。すでに各プロジェクトに新たな風を吹き込んでくださっています。一方、長年 MoSaIC プロジェクトにご尽力いただいた石川尋代さんが任期満了となります。この場をお借りしてこれまでのご貢献に感謝申し上げたいと思います。

慶應義塾大学 DMC 紀要　第 6 号（2019.3）
DMC Review Vol.6 No.1 (2019.3)
2019（平成 31）年 3 月 31 日発行

〔編集・発行〕
慶應義塾大学デジタルメディア・コンテンツ統合研究センター
〒223-8523
神奈川県横浜市港北区日吉本町 2-1-1
日吉キャンパス西別館 1
Research Institute for Digital Media and Content, Keio University
Hiyoshi Campus West Annex 1, 2-1-1 Hiyoshihoncho, Kohoku-ku,
Yokohama, Kanagawa 223-8523
TEL 045-548-5807　FAX 045-566-2153