

Title	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
Sub Title	
Author	原, 正一郎(Hara, Shōichirō) 金子, 晋丈(Kaneko, Kunitake) 安形, 麻理(Agata, Mari) 久保, 仁志(Kubo, Hitoshi) 石川, 尋代(Ishikawa, Hiroyo) 安藤, 広道(Andō, Hiromichi)
Publisher	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
Publication year	2019
Jtitle	慶應義塾大学DMC紀要 (DMC review Keio University). Vol.6, No.1 (2019. 3) ,p.63- 81
JaLC DOI	
Abstract	
Notes	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 開催日時：2018年11月20日(火) 14:00～17:30 開催場所：慶應義塾大学日吉キャンパス西別館1
Genre	Departmental Bulletin Paper
URL	<a href="https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO32002001-00000006-0063">https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO32002001-00000006-0063</a>

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the KeiO Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

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。時間が無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るのですが、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、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ブという言葉が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ものもありますし、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と。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**、**Semantic 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、

金子先生のコンテキ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」ということもあります(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、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ですか、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かけは、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步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ブ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、パースペクティブを束ねる客観的な値だと思われるメタパースペクティブというんですかね。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有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ます。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考えています。

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。「あっ、これ読んでみたいな」と思って開いたら、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ます。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。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く、「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っていくって、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こと、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かれないとか、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もの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っすぐ置いてみて。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、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と言われてたら、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、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ぬような 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。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のは、言葉でなかったとしても、必ずメタデータの観点があるのかあるのではないかと感じてしまうのですが、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ブを反映しているという話からすると、そ

こ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な必要じゃなくて、ただ単に言われたとおりにリンク張りしました的な計算をまずは想定しているんですけど。

**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：**そうなのでしょうね。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ども。最初におっしゃられたように、どちらもメタデータではないかとおっしゃったように、私の場合はどちら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ど。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、そのきっかけ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、それが分かかってないと探せない。使えない。そ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、例えば場所や時間を追加

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つ張ったから何が見えるかということ、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ただ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、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を上げるとのことなんですけど、レイヤーを上を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。よろしくお願ひします。先ほどの話を普段と離れたところから今年はこちらで聞いていて、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ど思ったんですけど、やはり情報組織化とか、それとはまったく違う話ですよ。

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、ただ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と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っただけですね。

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と。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。



**谷口：**文学部の谷口と申します。私などが話している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捉え方でしょうか、使い方というものがあるって、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が。

せっかくですので、高山先生、よろしくお願いたします。



**高山：**いまはもう所属はございません。元文学部にい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行っていくかということは、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す。それで一つご質問させていただきませう。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。私たち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よ。

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、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う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こと、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いくて、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と。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

それはそれであり得るけれども、ベースとなっていて動くのは、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す。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す。



**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が、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、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します話だと思います。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、そして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と。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考えております。

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。