

人文科学分野におけるデジタル・アーカイブについてのノート

金子 貴昭 (立命館大学COE推進機構COE研究員)
E-MAIL tkaneko@fc.ritsumei.ac.jp

立命館大学アート・リサーチセンター（以下、ARC）では、人文科学分野において「資料」として扱われる、様々な文化・芸術コンテンツのデジタル・アーカイブ活動が行われており、その一連の工程は全て、ARC内部で行われている。筆者は人文科学研究分野の出身であるが、過去数年間に渡って、ARCのデジタル・アーカイブ活動に深く関わってきた。特に静止画については、資料のデジタル化、画像処理、膨大なデジタル・コンテンツの管理、メタ・データの付与、データベース構築から公開まで、ほぼ全ての工程に数多く従事してきたと自負している。つまり筆者はいわば現場作業員であり、デジタル・アーカイブに関する研究を行ってきたつもりは全くない。従ってデジタル・アーカイブの理論や新手法を提示することはできないが、従来、研究者によって事例報告や技術紹介が行われたことはあっても、現場のアーカイブ担当者からデジタル・アーカイブに対する見解が述べられたことはほとんどなかったのではないだろうか。本稿では、現場サイドからデジタル・アーカイブ活動についての気付きを述べ、人文科学分野におけるデジタル・アーカイブの問題点を把握することにした。

技術上の留意点

一般には、「デジタル・アーカイブ」という語から、情報工学技術を駆使した非常に高度なものが連想されるかもしれない。しかし少なくともARCで行われている、閲覧に供することを目的とした静止画のデジタル・アーカイブについては、それほど難解な情報技術や画像処理技術を扱っているわけではない。例えば全くの初心者でも、

デジタルカメラで撮影された画像の余白を切り取るトリミング作業などは、マウスとキーボードのごく簡単な操作さえ可能であれば、誰でも初日から作業を行うことができる。

しかし「デジタル」という言葉から、「後でも何とでも処理できるだろう」「やり直しがきくだろう」という安直な発想を生む場合があるように思われる。せっかく捉えたデジタル化の機会が台無しになることもあるため、このような誤解を解いた上で上で作業に臨まねばならない。

例えばデジタルカメラで撮影した画像について、傾いた被写体を水平にするため、トリミング処理の工程で画像を若干回転させることがあるが、この処理は画質を劣化させる。静止画像はもと、ピクセル（正方形のドット）の集合体であり、各ピクセルを整数の座標上に配置することで表されるデータである（図1）。従って90、180、270度以外の角度を任意指定し、画像を回転させた場合、計算結果に小数が発生する。静止画像は整数の座標であるから、小数は表現できない。結果、例えば白黒2値画像の場合、白か黒しか存在しないため、小数以下は繰り上げたり、繰り下げられることによって、もとの形は崩れてしまう（図2）。近年はデジタルカメラやスキャナの普及



図1

図2 図1を白黒2値画像として、時計回りに5度回転させたもの。

によって、フルカラーでデジタル化されることが多い。フルカラー画像の場合には、小数に当たる部分は、もとの色とは別の中間色で補間される。結果として、輪郭がとぼけて全体的にぼんやりとした画像になってしまう(図3)。



図3 図1をフルカラー画像として、時計回りに5度回転させたもの。

図1～3は、分かりやすく例示するために、ピクセル数の小さい画像を拡大し、かつ回転角度を大きくしている。しかし一般に最も信頼され、利用されていると思われる画像処理ソフトでも、同様の現象は起きている。画像を一度保存してしまうと、可逆性はない。ことに活字資料や、筆者らが頻繁に扱う、和紙の上に墨で細かな文字が書かれたり摺られたりしている資料については、劣化の影響を受けやすいため、注意が必要である。極力、傾きのない状態で撮影を行い、デジタル化を行うのが好ましく、回転処理の繰り返しなどはもってのほかである。また、保存されるデータがすでに劣化したデータであれば、オリジナル・コンテンツをデジタル・データの形で恒久的に保存するという、デジタル・アーカイブの根本的な目的に反することにもなる。

こういった危険性は画像の回転に限ったことではない。色調補正や体系的なファイル管理など、デジタル・アーカイブ活動の各工程について同様のことが言える。個々の問題について、その場で適切な処理が行われなければ、膨大な作業の中で問題点が埋もれてしまい、後にその問題点にたどり着くのが困難になる。もしくは、時すでに遅し、という状況に陥るのである。

デジタル・アーカイブ研究の問題

さて近年、文化・芸術コンテンツのデジタル・アーカイブに関する研究が盛んに行われ、数多くの技術紹介や事例報告が発表されている。人文科

学研究分野であれ、情報科学研究分野であれ、デジタル・アーカイブ活動に関する成果となると、多くの場合、デジタル・アーカイブ化された何かではなく、「論文」という形になるようである。それらは研究の世界で行われていることであるから、論文という形になるのは当然のことである。しかし事例報告を見て利用してみようとする、思いのほか実際に利用できるデジタル・アーカイブが少ないのは、いささか残念に思われてならない。

また、筆者らが実際にアーカイブ活動に従事するに当たって、何らかの知見を得るために論文を参照しようとすることはある。ただし正直を言えば、現場の活動にはあまり役に立っていない。現場が常に必要とするのはノウハウである。筆者らのグループが細かなノウハウを学ぼうと論文を参照しても、ノウハウは各研究グループの「企業秘密」でもあるためか、論文の表面にはほとんど理論しか表れてこない。ノウハウが記されていないのであるから、先に記したような画像の扱い方に関する基礎などは、載っているはずもない。従って論文を参照するよりも、懇意の同業者と情報交換を行った方がよほど早く欲しい情報に行き着くということになる。

メタ・データの構築についても同様のことが言える。あまり意識したことはなかったが、筆者らが実際にメタ・データを構築しようとする際、数多くのメタ・データ構築事例はほとんど参照していない。改めて考えれば、アーカイブ対象となる資料に関する論文ばかりあさっていることに気づくのである。

例えば情報科学分野でしばしば言及される、15の属性要素から成るDublin Core Element Set[1]という仕様が存在する。Dublin Coreは、多様な分野に渡って共通して用いることができるように設計されている。従って様々な事例において、Dublin Coreはその汎用性が注目されているように思う。今試しに、筆者らが構築してきたメタ・データの属性項目数を確認すると、67項目あるものと、54項目あるものがあった。もちろんこの

数には作業用の項目も含まれているため、いくらか項目を減らすことは可能である。しかしこれらを最初から15個の属性要素に絞り込むことはどう考えても不可能なように思われる。不可能というよりも、15個の属性要素ではアーカイブ対象となる資料の性格を十分に表現することができないのである。

それではDublin Coreやその他の統一規格をもとに、新たな属性要素を追加すればよいのではないか、と思われるかもしれない。しかし筆者は、少なくとも新たにメタ・データを構築しようとする際に、統一規格を採用したり、意識したりしようとは思わない。メタ・データの構築という工程は、個々のデジタル・アーカイブ活動の思想を決定づける、特に重要な要素であり、デジタル・アーカイブ化された資料を効率的に活用することができるか否かを左右する、まさに人文科学分野の視点が入らなければならない工程といえよう。従ってその設計は、デジタル・アーカイブ構築の目的や個々のアーカイブ対象の特徴に即した形で設計されなければならない。「統一規格」や「汎用性」は情報技術の分野で用いられる概念であり、おそらく人文科学分野にはなじみにくいものであろう。その枠組みに「人文科学」を当てはめてしまうことに無理はないだろうか。

メタ・データに汎用性を求めた場合でも、アーカイブ対象に対して的確なメタ・データさえ付与されていれば、汎用的に必要なとされるコアな情報は自ずと内包されているはずである。従って各種データベースの仕様の相違が問題となった場合においても、十分対応可能と思われる。ただし、そういったシステムはレファレンスサービスには好都合かも知れないが、人文科学研究において効率的に活用できるかという点については、少々疑問を感じる。

人文科学分野の問題

筆者はデジタル・アーカイブ活動に関わりながら、筆者らが構築したデジタル・アーカイブが何

かを導き出すとは考えていない。あるいは、デジタル・アーカイブを活用することによって大量の情報が扱えるようになり、研究上の新視点が獲得できるかもしれないという期待は持っているが、やはりそれ以上のものには思えない。つまり、何か調査や研究に際して、便利に使用したり、使用されれば良い、というのが本音である。つまり筆者は、研究におけるデジタル・アーカイブを、単なるツールと見なしており、そのツールを使って行う研究は、従来の研究と特に変わりはないと考えている。しかし人文科学分野においては、この段階から議論が遅々として進まなくなる。

例えば最近、人文科学分野から見たデジタル・アーカイブ批判として、「欲しい資料がデジタル化されていない、底本の選定が甘い」[2]という点が紹介されている。前者は身につまされる部分もある。しかし後者については、デジタル・アーカイブ活動においてあらかじめ準備されることではないように思われる。なぜなら、本来「底本の選定」は、可能な限り全ての諸本を確認した上で研究者が行うべきことであり、「底本の選定」自体が人文科学の研究として十分に成り立つからである。また、諸本の確認に関わる地理的な制約や時間のコストを削減し、資料との接触を容易にするのがデジタル・アーカイブの大きな役割の一つでもある。どの程度多くの人文科学研究者がこうした批判を持っているのか、筆者には分からない。しかしこの批判は、デジタル・アーカイブを研究利用するより先に、その結果を求めたものであり、本末転倒と言える。

こうした批判は、人文科学研究者が情報技術に対して抱いている、ある種の夢に起因するのであろう。人文科学研究者の中には、情報技術が人文科学研究における何かしらの結論を導き出してくれると考える研究者が、少なからず存在するようである。この考え方の存在は人文科学系の研究分野にコンピュータがもたらされた頃の議論において、すでに指摘されている[3]。後にコンピュータは、便利なツールとして盛んに利用されるようになった。むろん、当時のスタンド・アロンのコ

ンピュータと、ネットワークを介した広域な展開が可能な現在のデジタル・アーカイブを一概に比較することはできない。しかし現在、デジタル・アーカイブを話題にするにあたって、情報技術はツールとして有効であるが、情報技術だけでは何もできない、という当たり前の議論をまた繰り返さなければならないのであろうか。

一方では、人文科学分野でもデジタル・アーカイブを活用した研究が見られるようになってきている。最近の研究では、武内恵美子氏による上方役割番付の統計分析[4]が、データベース化のみならず、それをもとにした考察まで行われており、大変興味深い。しかしその土台となるデータベースの構築法については、引っかかるところがある。武内氏の研究は「江戸時代の京都と大坂の劇場で上演されたものの番付約18000点中、重複分を省いた約4500点を調査しデータベース化」されたものという。当該分野のデータベース構築法から考えれば、順序が逆であると言わざるを得ない。氏は所蔵機関ごとの既刊の目録に即して調査されたものと思われるが、上方役割番付に限らず、この種の資料には、細部に入れ木による部分訂正が施されているものが多く、諸機関が発行する目録の情報から重複を判断するのは難しい。重複を厭わず18000点を調査し、後に重複を省くのが本来のあり方ではなかったか。膨大な調査数の中から重複を省くという処理こそ、データベースが最も効力を発揮するところである。重複と見なされた中に考察結果に対する例外は含まれていなかっただろうか、といった勝手な不安を覚えるのである。

確かに18000点を1人で調査するのは現実的とは思えないが、4500点ずつ4人で調査を分担するとなれば、いささか現実味を帯びるだろうか。武内氏の手法は合理的という点で十分肯定できるものであり、筆者は氏の手法を批判したいのではない。むしろ人材の不足という問題が垣間見えてくる点を指摘したいのである。

人材の不足

ARCでは、大学という学生が集まる場の特性をいかし、プロジェクトに参加する学生や学生アルバイトを人材として活用することが多い。学生の活用は、デジタル・アーカイブ作業を業者に委託するよりもコストが圧倒的に安い、もちろん目的はそれだけではない。ARCでは、学生が実際に資料に触れつつ学習したり、作業内容を自身の研究の糧にするなど、研究・教育の場とすることも目的としている。

また学生だからといって、その能力は必ずしも侮れない。筆者はかつて、とあるネガ・フィルムのデジタル化に関わったことがある。一部を業者に委託し、残りは学生チームが引き受けた。フィルムは1本ごとにネガ・シートに収められていたが、1コマずつ切り離され、しかも順不同にケースに収められていた。連続性が必要な資料であるにもかかわらず、その連続性が全く失われていたのである。業者はシートに入ったままを預かり、連続性が失われた状態でデジタル化を行い、連続性が失われたデジタル画像に対して通し番号で画像ファイル名を付けてきた。一方の学生チームは、業者と同様にシート毎に作業を行ったが、各コマの連続性が必要な資料であることを見抜き、ネガの切り口やコマ番号から連続性を回復してきたのである。おそらく、フィルムのデジタル化技術は、業者の方が圧倒的に優れていたに違いないと思われる。しかしデジタル化されたものを研究資料として評価するならば、当然学生チームが行った仕事の方が、明らかに数段レベルが高いと言える。学生チームが資料の特性に理解を示し、数倍安いコストで、数倍の作業量をこなしたことを加味すれば、彼らを活用するメリットは十分認められる。

自己批判を行えば、コストの安さに任せて、学生チームに多くを委ねすぎてきたという嫌いがある。思う存分デジタル・アーカイブ活動に尽力して欲しいという当方の願いに比べて、学生生活というタームはあまりにも短すぎる。しかも個々の

学生には専門分野の研究や就職といった本分があり、デジタル・アーカイブ活動に携わる期間はさらに限定される。「文化財」や「デジタル・アーカイブ」という言葉に飛びついて参加する学生に至っては、目前に積み上げられた膨大で単調な仕事に辟易し、人知れず去ってゆく者が多い。

最も切実な問題として、アーカイブ対象となる資料に対する理解度があげられる。アーカイブ対象に対する理解がなければ、デジタル・アーカイブは成り立たない。特に学部生は専門意識がまだ十分に確立しておらず、相対的にアーカイブ対象に対する理解度が低いということになる。デジタル撮影においては、資料原本の取り扱い方を熟知していなければならない。先に技術的には誰にでもできると記したトリミング処理についても、どこまでを資料とみなし、どこからを余白と見るのが判断できなければ、何もできない。特に対象が複雑な形態の資料（貼交帖など）ともなれば、アーカイブ対象に適切な勘を持たない者はお手上げ状態になってしまう。撮影や画像処理という技術的な側面の強い作業工程においても、アーカイブ対象に対する理解が必要となる。

近世以前の資料を取り扱うにあたっては、くずし字の読解能力も必須となる。極端な例をあげれば、非常に薄い和紙に文字が印刷された一枚物の資料を学生に撮影させた際、表裏を誤って撮影してきたことがあった。彼がコンピュータとカメラの扱いに気を取られていたせいであろうが、被写体に何が書かれているのか、少しでも視界に入ったならば、本来は起こりえないミスである。彼の目には印刷されたくずし字が不可解な記号の羅列としてしか映っていなかったためであろう。

こういった状況であるから、先にその重要性を記したメタ・データの構築など、学生は当然参画できない。結局、実務上できることは非常に限られており、扱いやすいアーカイブ対象の画像処理や撮りこぼしの確認作業といった比較的簡単な工程に徹することになる。そして学生がようやく満身に技術を身につけてきた頃には、彼らがチームを離れる時期になっているのである。

あるいは、アーカイブ対象に理解が必要ならば、それらは研究者の仕事である、と人は言うかもしれない。しかし昨今の大学事情や知的生産が研究者の本業である点を考えれば、研究者にそれを全て委ねるのも酷であろう。研究者がデジタル・アーカイブ活動を行わないとしたら、誰が行うべきだろうか。筆者はこの問題をクリアするために、継続的に大規模なデジタル・アーカイブを構築する組織においては、研究部門とは別個に、デジタル・アーカイブ活動を専門的に実践する技能集団が形成されるべきであると考えている。研究部門はスーパーバイザーとしての役割を担わなければならないが、デジタル・アーカイブ活動の各工程における煩わしい作業からは解放される。従って技能集団は研究部門と緊密に連繋が取れる位置に設置されるのが好ましいであろう。研究者はそこで構築されたデジタル・アーカイブを活用して、効率的な知的生産を行えば良いのではないか。人材の不足はなお問題として残る。技術的にそれほど難解ではない点、アーカイブ対象に対する理解が必要である点を考慮すると、人文科学分野の人材でも十分に業務を遂行でき、むしろ適任であるといえる。とすれば、今後さらに増えると思われ、国家レベルで救済策が叫ばれている余剰博士や、話題にもものほらない余剰修士に期待を寄せたくなるが、これは卑しい見方であろうか。

おわりに

研究の世界は、この十数年の間に大きく変化してきたと思われる。かつては、広い視野で浅く掘り下げた研究手法と、視野は狭いが緻密に深く掘り下げる研究手法とが存在した。しかし最近では、広く深く緻密な研究成果が要求されるようになってきたのではないだろうか。その要求に対応するためにも、紙のメディアと違って、大量の情報を扱えるデジタル・アーカイブは有効なツールである。しかし現在はまだその土台作りが行われている段階に過ぎない。

近時、セキュリティや権利問題といった、デジ

タル・アーカイブが内包する学術的課題を克服するために、学問としての「デジタルアーカイブ学」という領域が提唱されている[5]。それについては筆者も大いに賛同するし、その領域において人材が育成されることを期待するところである。一方では、すでにデジタル・アーカイブ化され、利用に供されている文化・芸術コンテンツは、ごく一部に過ぎない。従って学術領域の形成と同時に、デジタル・アーカイブ活動を、引き続き精力的に行っていかなければならないのもまた事実である。

以上、筆者の思うままを煩雑に記してきた。やや極端に物事を述べた部分もあり、批判を受けることをおそれるが、過去数年間ライフワークのようにデジタル・アーカイブに携わってきた筆者の実感を述べたつもりである。何にせよ、筆者が関わるデジタル・アーカイブ活動の現場が人材を渴望していることだけは間違いない。そして人文科学分野出身の人材があと少しでもデジタル・アーカイブ活動に振り向けられたなら、人文科学におけるデジタル・アーカイブを取り巻く環境は大きく変わると筆者は考えているのである。

参考文献

- [1] The Dublin Core Metadata Initiative
<http://www.dublincore.org>
- [2] 赤間亮、川村清志、後藤真、野村英登、師茂樹：人文科学にとっての“デジタルアーカイブ”、人文科学とコンピュータシンポジウム論文集「デジタルアーカイブ-デジタル学術情報資源の共有と活用-」、情報処理学会、pp.259-267 (2004)
- [3] 近藤瑞男、阪口弘之、武井協三、原道生、棚町知彌：一九八九年の近世演劇研究、近松研究所紀要、第1号、pp56-87 (1990)
- [4] 武内恵美子：上方歌舞伎役割番付データベースと統計分析による長唄の変遷と組織形態の解明、人文科学とコンピュータシンポジウム論文集「デジタルアーカイブ-デジタル学術情報資源の共有と活用-」、情報処理学会、pp.43-50 (2004)
- [5] [2] に同じ