

Title	「プロジェクトをアーカイブする：HUMIプロジェクトのアーカイブの試み」
Sub Title	
Author	池田, 真弓(Ikeda, Mayumi)
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	2015
Jtitle	慶應義塾大学DMC紀要 (DMC Review Keio University). Vol.2, No.1 (2015. 3) ,p.18- 22
JaLC DOI	
Abstract	
Notes	特集：DMC研究センターシンポジウム：第4回 デジタル知の文化的普及と深化に向けて： MoSaICによる多面的アーカイブへの挑戦
Genre	Departmental Bulletin Paper
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO32002001-00000002-0018

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.

「プロジェクトをアーカイブする -HUMI プロジェクトのアーカイブの試み -」

池田 真弓

DMC 研究センター研究員 理工学部専任講師

ただ今金子先生から、我々のプロジェクトのカタログ化について、その概念のお話があったと思いますので、私の方ではそれを実際にやってみるとどうなるかというのを、HUMI プロジェクトのアーカイブというものの実例を示しながら、皆さんと考えていきたいと思えます。

まず HUMI プロジェクトをアーカイブするということですので、その前提として、「プロジェクトをアーカイブする」とはどういうことかというのを、私なりに考えてみました。先ほどの嘉村先生のお話では、藝大のいろいろな財産や資産をアーカイブするということでしたけれども、プロジェクトとなりますと、一体それと何が一緒で何が違うのか、それも考えていければと思います。

「プロジェクトをアーカイブする」とは、まず第一には、プロジェクトが生み出したもの、つまり成果物をアーカイブすることが挙げられるかと思えます。それだけではなくて、プロジェクトのアーカイブの第二の特色として、プロジェクトの活動のプロセスそのものをアーカイブするということが挙げられます。そして3つ目には、これは嘉村先生のお話にもあったかと思えますが、プロジェクトやその成果物に関連する事象、成果物からさらに生み出されたもの、あるいはプロジェクトそのものが生み出したものではないけれども、プロジェクトにリファー（refer）しているもの、あるいはプロジェクトがリファーしている、いわゆるレファレンス（参照）の関係にあるものも、一緒にアーカイブできたら非常に面白いのではないかと考えております。

では今回考えている HUMI プロジェクト、これはそもそも何かと申しますと、ご存知の方もいらっしゃると思いますが、Humanities Media Interface Project というのがフルネームでして、1996 年から 2009 年まで足かけ 13 年活動していたプロジェクトで、その主たる目的は、貴重書のデジタル化でした。HUMI の設立にはきっかけがあります。1996 年の春、私が慶應に入学するよりも前ですが、この時にゲーテンベルク聖書が慶應の所蔵となりました。HUMI はこれを契機に誕生したプロジェクトです。ゲーテンベルク聖書

というのは 15 世紀の半ばにドイツで印刷された本です。何がそんなに重要かといいますと、活版印刷術を用いて印刷された、ヨーロッパでは初めての本ということで、印刷史の歴史では、非常に記念碑的な本となっています。ゲーテンベルク聖書は、アジアでは実は慶應義塾しか所蔵していないんですね。そういう意味でも、非常に重要な本なのですが、せっかく買ったからには何かに使おうということになりまして、デジタル化をしようということになりました。現在は貴重書のデジタル化というのは、どこでもやっていることだと思いますけれども、このときにこれを思いついたというのは非常に先進的で、HUMI プロジェクトというのは、貴重書のデジタル化プロジェクトとしては、パイオニア的な存在です。ゲーテンベルク聖書だけでなく、例えば日本の奈良絵本という、江戸時代に流布したとても美しい、絵本の原形になったような本があるのですが、そういったものも含めて、たくさんの貴重書のデジタル化を進めていきました。さらに HUMI プロジェクトの特徴としては、海外の撮影遠征というものもあります。こちらは先ほど言いましたけれども、ゲーテンベルク聖書というのは、アジアでは一点しかありません。ではもっとゲーテンベルク聖書を撮りたいとなったときには、外に出向いていくしかないということで、随分と海外に出向いて、撮影遠征もしました。それからやはり貴重書のデジタル化プロジェクトとしては、パイオニアですので、デジタル化の手法というのも、自分たちで試行錯誤して、開発していったという特徴があります。

そのような HUMI プロジェクトを DMC でアーカイブしていかうとなりました。アーカイブするにあたっては、まず第一に、状況の把握が必要です。HUMI のデータの特徴、あるいは問題をごく簡単に挙げてみますと、いろいろなメディアの集合体であるということ。もちろん、撮影の完成画像が HUMI のデータのメインとなりますけれども、それを撮影するに至るまでの、さまざまなデータが存在しているということです。それからやはり、これはこのような時限付きのプロジェクトでありがちなのですが、プロジェクトが走っている最中は、なかなかシステム

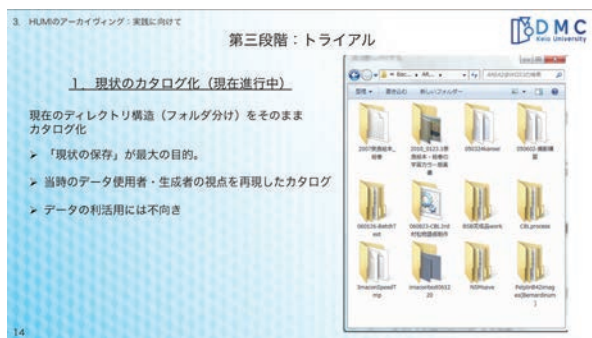


図1 現状のカタログ化

ディックな整理がされず、とりあえず撮ったら置いておこう、撮ったら置いておこうを繰り返していった結果、何がどこにあるのか分からないような状況になってしまっていました。それから先ほども申しましたけれども、海外をはじめとした、他の機関の本も撮影していますので、画像の使用権の問題も、今後公開ということを考えていくと出てきます。

そのように状況を把握した後に、どうやってアーカイブしていこうか、方針と手段を検討しました。どのようなアーカイブにしたいかを考えてみますと、まずはなにしろアクセシビリティを高めたいと考えました。HUMI のデータは、今は全く活用されておらず死蔵してしまっていますが、非常に重要な資産です。研究者や興味のある方に、どんどん見ていただきたい、活用していただきたいと考えております。それから、プロジェクトの活動全体が見えるようなアーカイブができれば、それは「プロジェクトのアーカイブ」としても、意義があるのではないかなと考えております。さらに、プロジェクトに関連する事象も一緒にアーカイブすれば、いわゆる論文のインパクトファクターと同様、そのプロジェクトの意義ですとか、インパクトを視覚化することができるのではないかと考えております。このようなアーカイブを目指すこととし、次に、アーカイブの手段をどうしようかなと思ったところ、ああそうか MoSaIC でやろうと。最初は MoSaIC でやるというのは、思いつかなかったのですが、HUMI のアーカイブを DMC でやるということで、MoSaIC を用いたアーカイブを進めていくことにしました。

先ほど金子先生のお話にもありましたが、MoSaIC のカタログ化というものの特徴としては、多面的なカタログ化ができるということです。技術展示でご覧になった方は分かると思いますが、いろいろなデータのつながりを矢印で表現することが可能である、ということです。また、可視化という点でいえば、アーカイブが目に見える形になることで、直観的な理解が期待

できるのではないかと。さらに、先ほどの（金子先生のお話にもありましたが、必ずしも言語を使わなくても、オブジェクト同士の関係性を記述することができる。もちろん言語によるタグ付けもできますし、そうしなくてもカタログ化することができるという特徴があります。それはちょうど HUMI プロジェクトのアーカイブに、もってこいだなと思っていて、現在いろいろ試しているところです。

この MoSaIC を用いたカタログには 3 つのアプローチがあります。1 つは現状をそのままカタログ化していくこと、次に、機械に委ねた機械的なカタログ化、そして 3 つ目が、こちらを今私が進めているのですけれども、人的なカタログ化というのがあります。それぞれご説明いたします。

まず 1 つ目の現状のカタログ化ですが、HUMI プロジェクトから受け取ったデジタルデータを、現状のディレクトリ構造そのままに保存しようということです（図 1）。（スライドの）右にありますのが、HUMI プロジェクトからデータを受け取ったそのままのディレクトリですが、いろいろなディレクトリ名がありまして、例えば、「BSB 完成品 Work」ですとか、「NSMsave」ですとか、私はこれらが何かは分かるのですが、多分他の人たちは、なんのことやらさっぱり分からないでしょう。このカタログ化は、現状の保存が最大の目的となっていて、これを作った使用者、データやディレクトリの生成者の視点を、そのまま再現したカタログですね。だから見る人が見たら面白いかもかもしれませんが、ほとんどの人にとっては、データの利活用という点では不向きです。

次に機械的なカタログ化というのは、デジタルデータのプロパティ情報を抽出して、それを基に機械的にカタログ化を行うということです（図 2）。プロパティ情報というのは、ファイルについているものでして、ファイル名もそうですが、ファイルサイズ、撮影したカメラのモデルなど、いろいろな項目があります。現在プロパティ情報を機械的に抽出してもらうように、



図2 機械的なカタログ化

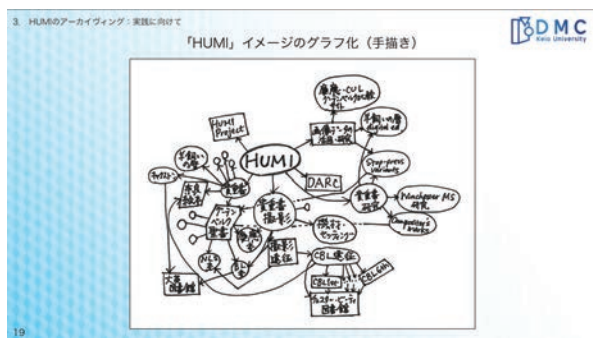


図3 「HUMI」グラフのイメージ化

金子研究室の学生さんに取り組んでいただいております。これを基にカタログ化をしてみようというものです。そうするとやはり、手作業では困難な、膨大な数のデータを、正確にカタログ化することができる。あとは人間の視点というものが入らないので、これが人的カタログ化とどのように違ってくるのか、非常に興味深いところだなと思っています。

そして最後に、私が今取り組んでいる、人的カタログ化ですが、これはまさにもう手作業です。1つ1つのデータを、一生懸命カタログ化していくのですが、ある意味これは、データのキュレーションと呼べるもので、自分がこういうアレンジをしたり、こういうふうに整理していきたい、というものを、そのまま再現していこうというものです。ですから现阶段では、私の視点を再現したものということになりますが、上述の現状のカタログ化などに比べると、利用者のニーズに合わせていくとか、多面的なカタログを作成することが可能かなと考えております。

現在実際に、HUMIのデータを用いて、人的カタログ化を作成しておりますが、その出発点としては、私から見た「HUMI」というのは一体何なのかということ、このカタログを用いて明らかにしたい、これが私のHUMIのカタログ作成の出発点となっております。

このカタログ化にあたって、HUMIとは何か、いろいろ思い浮かべてみました。いろいろなゲーテンベルク聖書がありますし、撮影の風景もありますし、シンポジウムもある。いろいろな要素が重なってHUMIというイメージを作り上げているんですね。これらのイメージを、まずどうするかというと、非常にアナログに手で書いていくわけです(図3)。これとこれはこういう関係、これとこれはつながって、という具合に、1つ1つのオブジェクト—それをノードと呼ぶのですが—に関係するデータをどんどん線でつなげていく、と教えてください。それを今度は、まだカタログのエディタが完成していないので、既存のWindowsあるいはMacのフォルダを使ってグループ分けして

いきます(図4)。今実際のものをご覧いただきたいと思いますが、こんなふうになっておりまして、上の技術展示では、これがまさに可視化されたものが見られます。例えば、私にとってのHUMIというイメージには、「DARC」、「貴重書研究」、「貴重書」とか、いろいろあるわけですし、さらに貴重書研究の中にも、「ああゲーテンベルク聖書の研究がこんないろいろなあるな」と思ったり、あるいは貴重書という項目の中でも、いろいろなグループが作れるな、というイメージを、カタログで表現していきます。通常のフォルダ構造では、同じデータがいくつものフォルダに重複して入ってしまうこともあります。同一のデータが、こちらのグループとあっちのグループ双方に所属するんだよ、というのは、既存のフォルダ構造では表現できません。それが、カタログの場合は、同一のデータが複数のグループに所属する、という状態を、矢印を引っ張って表現するんですね。その作業をしまして、今こんな感じでカタログ・ビューワで見られるようになっております(図5)。後ほどデモでご紹介できるかと思っておりますが、こういったカタログ化作業を現在行っておりまして、まだ始めたばかりですが、それなりにいろいろな課題も見えてきました。

人的カタログ化の作業での課題というのは、やはり何しろ分類や定義づけの難しさという点が、挙げられます。カタログにはタグ付けは必要ない、関係だけで記述できる、それは分かっているのですが、もともと私の研究分野では分類というのが重要ですので、どう

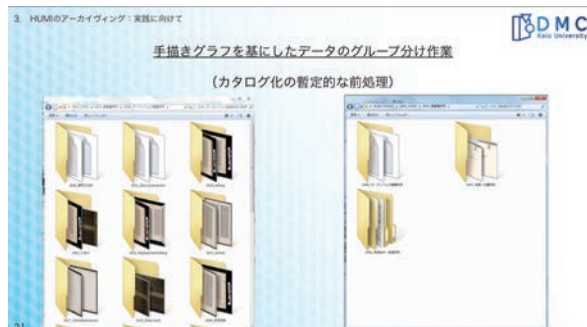


図4 フォルダを用いたグループ分け



図5 カタログビューワ (MoSalC II)

しても定義づけしたい、分類したいという方向に行ってしまうのです。しかし、それはもしかしたら、そういった考え方に慣れきっている、私自身の問題かもしれません。あるいはもしかしたら、分類したがるのは人間の本質的な思考法に基づいているのか、まだ分からないのですが、とにかくそういう課題が見えてきました。あとはもちろん、手作業によるカタログ化の限界という課題もあります。HUMIのデータは12TBあるのですが、これですとまだ扱うには、どうってことない量かなと思います。けれどもどんどんアーカイヴが大きくなっていくと、やはり人間だけによるカタログ化は難しい。また人間ですから、どうしても間違えることもあるかもしれない、そういった場合どうするのかという問題があります。ただMoSaICの場合、1つのデータに対して、いろいろな方向から関連性があるとすれば、そのうちの1つが間違っている、他の関連性からそのデータを見つけることができる。そのときに、ああこれ間違えたって気づくことができれば、大丈夫かなと。そこがMoSaICの良さかなとも考えております。とにかく、手作業の限界といった点なども踏まえると、機械と人間がどうやってうまくコラボレーションして、このカタログ化を進めていけるのかなというところが、今後考えていければと思っているところです。

カタログ可視化についても、課題が見えてきました。先ほどお見せした可視化のシステムは、MoSaIC II と言いますが(図5)、もともとデジタル・ミュージアムを想定して開発したのですが、ミュージアムとアーカイヴというのは、似て非なるものかもしれません。アーカイヴでは、ミュージアムと違って、価値があるとみなされたものだけではなく、一見そうとはいえないものも含めて、多種多様なファイルがあります。それを先ほどの可視化を使って全部表示してしまうと、とんでもないことになる。コンピュータもですが、人間がとまくなにがなんだか把握できなくなる。では何かに限定して表示するのか、そういう場合は、何を選んで表示するのか、その辺も考えていかなければならない課題です。また、先ほど嘉村先生の話にもありましたけれども、HUMIの場合でも、肖像権やプライバシーにかかわるデータもあります。ですからそれらは可視化しないように、何かロックをかけなければいけない。でもそれは、誰がどう判断するのか、その辺りについては、私よりも詳しい専門の方もいらっしゃると思いますので、お聞きできれば嬉しいなと思います。そういうわけで、これからはアーカイヴとし

ての利用に供するのに都合のいい可視化がどんなものなのかということ、考えていく必要があるのではないかと思います。

最後にもう一度、「プロジェクトをアーカイヴする」ということを、私の視点で振り返ってみたいと思います。今回のHUMIのアーカイヴ化作業をやっていて実感したのは、プロジェクトをアーカイヴするのは、そもそもそのプロジェクトの営みそのものを再構築するということ、1つにはあること。これは、いってみれば考古学的な興味といえるかもしれません。それから、プロジェクトにまつわるモノや時、空間を記録し、保存し次世代につなげていこうということが、いま1つ挙げられる。ただそれだけでなく、プロジェクトを全く知らない人にとっても、価値ある何かを生み出していきたい、プロジェクトの枠を超えた活用を促していきたい、そうすることによって、新たな広がりですとか、新たな価値が生まれていくのではないかと考えております。そしてやはり最初に、プロジェクトをアーカイヴするといったときに、成果物のアーカイヴ、活動プロセス、プロジェクトの関連事象、それぞれのアーカイヴが大事じゃないかとお話しましたが、そういったカテゴリーにとらわれることなく、これらの情報をシームレスで、より有機的に関連付ける必要性というのがあるのかなと、今回感じました。それは言ってみれば、人間の認知や思考・記憶の仕方に即した、アーカイヴィングを目指したいということになるかと思います。以上で私のお話を終わらせていただきます。どうもありがとうございました(拍手)。

司会：ありがとうございました。今は少し時間がありますので、何かフロアの方から、詳しいコメントか何かございませんでしょうか。では私の方から1つだけ。今回のそのDMCシンポジウムのテーマというか、タイトルにあるMoSaICというシステムを使って、池田さんのHUMIのデータを、実際に使ってみられたということなのですが、今のところ最後の方のスライドで、今後の課題というのをいくつか挙げていただきましたけれども、現場としてその今後の課題を考える以前に、もうちょっと使ってみるみたいなフェーズが必要かなと思うのですが、まだ今使ってみるフェーズというのか、そういう段階なのか、それとももう十分使いこなした使い倒した、それでやっぱりこういう課題が見えてきている、そのあたりどんな感じか、ちょっと教えていただけませ

んか。

池田：使うというのは MoSaIC を。

司会：はい、MoSaIC を実際に使う。

池田：なるほど。使ってみての方。

司会：まだ十分その使い込みが足りないのではないかと
いう、私はどちらかというと、この技術者サイド
の立場にいますので、もうちょっと使ってみてから課
題ができるかなと思うし、どれぐらい使いこんで出
てきたのだろうと思って、ちょっと意地悪な質問か
もしれない。

池田：どのぐらい、そうですね、エディタという意
味では、全然使いこんではいないですし、それはま
だ開発中なので、まだ使い込みようがないといいま
すか。ですから、エディタは、技術的な面ではいろ
いろ課題、例えば作業の利便性というところでの課
題がありますが、ある種カタログというのは、考
え方というか、コンセプトでもあると思いますの
で、そういう意味では DMC に今年 4 月から正式に
入りましたが、4 月からずっとカタログってなんだ
ろう、とずっと考えて、最近ようやく分かったとこ
ろでして、最初はやはり私もどうしても、言葉で定
義したい、タグ付けしたい、タグ付けと何が違うん
だろうという疑問が解消されず、何度も DMC の会
議でも紛糾して、そういう話もしましたけれども、
実際に今回の作業を経たことで、矢印だけで関係性
を表せるというのが、ようやく見えてきたという
ところなんです。ですから、今回短い時間で、こうやって
MoSaIC を紹介しても、何だろうって思う方が、た
くさんいらっしゃると思いますが、理解するには最
低 5～6 カ月は考える必要がきっとあるのだら
うと思います。