

第2回東京大学学術資産アーカイブ化推進室主催セミナー  
かわいい子には旅をさせよ — デジタルアーカイブとオープンデータ —  
パネルディスカッション記録

セミナー後半のパネルディスカッションは、永崎氏によるオープンデータの事例紹介があり、その後、事前にフロアから集められた質問を中心にディスカッションが行われました。その内容を要約してお伝えします。

事例紹介：永崎研宣氏（一般財団法人人文情報学研究所 主席研究員）

パネリスト：武田英明氏（国立情報学研究所 情報学プリンシプル研究系教授）

渡辺智暁氏（慶應義塾大学 大学院政策・メディア研究科特任准教授、  
クリエイティブ・コモンズ・ジャパン理事長）

福田真人氏（東京大学大学院人文社会系研究科 博士課程, Tokyo Digital History メンバー）

山崎翔平氏（東京大学大学院経済学研究科 博士課程, Tokyo Digital History メンバー）

永崎研宣氏（一般財団法人人文情報学研究所 主席研究員）

モデレータ：中村覚（東京大学情報基盤センター 助教, 学術資産アーカイブ化推進室員）

【セミナー概要】 日 時：2018年11月22日（木）13:30-17:00  
場 所：東京大学 福武ホールラーニング・シアター  
参加者：132名

## 事例紹介（永崎研宣氏）

### オープンなデジタルアーカイブの活用事例

- ・オープンデータの事例として、そして、この後のディスカッションの論点として、第三者が自由に再配布できるデジタルコンテンツ、そういうデータを搭載しているオープンなデジタルアーカイブを構築する利点について紹介したい。また、データの自由な再配布を禁じることは、公開者の負う責任を結局は大きくしてしまうということもお伝えしたい。
- ・現在はコンピュータシステムやウェブの技術が高度化し、他機関のシステムとの連携が容易に行えるため、デジタルアーカイブに求められるサービスのハードルが上がっている。市販のパッケージシステムでの公開では満足してもらえない、一方で業者にシステム改修をお願いし続けることは予算面から無理、といった状況がある。旧態依然としたシステムで公開し続けることで、せっかくデータそのものは良いものであっても、使い勝手の悪さなどが批判され、また実際に利用者に不便を強いることにもなっている。
- ・そこで、データだけはオープンデータとして公開することで、「使い勝手が悪い」といった批判に対しては「データは公開しているので、それぞれのコミュニティで便利に活用してください」と言うことができ、公

- 開側の責任を多少は小さくすることができる。また実際に、それらのデータが活用される場合が大いにある。
- ・図書館ではコレクション単位で画像を公開することが多く、意外なところに仏教に関する画像が含まれている。また図書館では、文庫全体の特徴は把握していても、個々の資料には精通していないことが多く、研究者がもっと詳細なメタデータを付けられるのではないかと考えた。そこで「IIIF Manifests for Buddhist Studies (IIIF-BS)<sup>1</sup>」というサイトを作り、複数の機関がウェブ上に公開している仏教学関連画像を集め、かつ研究者がメタデータを付加できるようにしている。もちろん、これらのデータはオープンデータとして公開している。
- ・オープンデータとして公開すると、それに関心を持つ第三者が、何か良いものや便利なものを作ってくれることがある。また IIIF-BS では、研究者が追補したメタデータを公開機関にフィードバックしており、いわば利用者が自主的に行っていることで公開元のデータが豊かになる一事例としてもこれを紹介したい。

<sup>1</sup> <http://bauddha.dhii.jp/SAT/iiifmani/show.php>

## パネルディスカッション（以下、敬称略）

**中村：**では、会場からの質問を中心に進めたいと思います。まず武田先生に、オープンサイエンスは学术论文のオープンアクセス、研究データ管理など様々なものが含まれると思いますが、その中で人文社会系分野のデジタルアーカイブはどのように位置付けられるかなど、デジタルアーカイブをどう捉えられているでしょうか。

**武田：**講演ではオープンサイエンスのサイクルの話をしました。デジタルデータは研究の入口でもあり出口でもあって、それは自然科学系も人文社会系も変わりません。また分野によって扱うデータの種類は様々ですが、どれも重要なのだと思います。"どれも重要"を前提として、アーカイブはオープンであるべきことを強調したいですね。個人情報など理由があって公開できないものは別として、データはオープンでない価値が出ないものです。また、アーカイブ構築にはコストがかかるので、成果や説明責任を求められます。同時に研究に貢献するものだとも思うので、アーカイブというのは原則としてオープンなものを対象にすべきだと考えてます。



武田英明氏

**中村：**では次に、ライセンス関連の質問が多数ありましたので、それについて渡辺先生お願いします。

**渡辺：**繰り返しになりますが、私は弁護士や法学の専門家ではないので、その点をご了解ください。まず「メタデータは著作物になるのか？」という質問について。

データベースそのものに創意工夫があればデータベースの著作物性が認められるケースはありますが、データベースの構造をコピーすることなくメタデータを抽出して利用するだけの普通の利用だったら、これは関係ないでしょう。メタデータ中に解題等もなく事実情報だけならば、天気や株価情報に著作物性がないのと同じで著作権はないものと考えます。



渡辺智暁氏

**武田：**著作物性のないメタデータにCC0（ゼロ）<sup>2</sup>を付けて公開する意味って、ありますか？

**渡辺：**あまり無いと思いますが、ただCC0は、「私はこちらに関する権利を一切放棄します」「形式上放棄できないものについては行使しません」という意思表示を含んでいます。そこで、どこかの国で、意図せず何らかの権利が発生するかもしれないと思うときは、CC0を付けておくのは良いかもしれないです。例えば米国の連邦政府刊行物はパブリックドメインとされますが、CC0を付与して公開しているよい例が議会図書館のアーカイブにありました。パブリックドメインのものにライセンスを付与しなくても一見思いますが、日本を含め他国での利用の文脈ではその国の法律等からパブリックドメインにならないことも可能性としてあり得ますので、その点でCC0を付ける意味はあるかと思います。

**中村：**永崎先生にもお聞きしたいのですが、特に人社系では研究に基づく翻刻テキストには作成者の権利、つまり著作物性を示すことも必要という意見があるように思いますが。

<sup>2</sup> <https://creativecommons.jp/sciencecommons/aboutcc0/>

**永崎**：著作権は、創作性の有無が非常に重要なポイントとしてあります。一方で翻刻や校訂テキストは、原本に忠実であることが求められます。つまりそれは創作性が無いことを意味し、オリジナリティは主張しにくいものです。そのためこういったものは、著作権ではなく専門家コミュニティできちんと評価のルールを作ることが大事だと思っています。



永崎研宣氏

**武田**：そういえば自然科学系分野では、データジャーナルを刊行してデータそのものを論文と同じように評価する仕組みがあります。ピアレビューが行われ、そのデータはデータリポジトリに置かれます。そのデータを利用するときは、当該ジャーナルを引用するのですが、人社系でもそういった動きはありますか？

**永崎**：最近、オランダの出版社 Brill が人文社会系分野のデータジャーナルを出版しました<sup>3</sup>。今後そういった評価は人社系でも可能ではないかと思います。またメタデータの著作物性についての補足ですが、古文書や古典籍では来歴情報が重要で、例えばケンブリッジ大学の図書館のメタデータには来歴情報があり、CC ライセンスが付与されていることもあります。調査結果としての来歴情報のようなものがあると、メタデータにも著作権が発生するのかなと考えています。

**渡辺**：それで言うと、新聞記事は事実を伝えるものですが、情報の取捨選択や配列にも書き手の思想や感情が表現されうると考えられています。同じようにメタデータも、独自のフォーマットで、そのデータを作成した人の知的な判断が反映されているような取捨選択

と配列があるなら著作物性があり得るとも考えられます。ただ、そういうことからメタデータの著作物性が認められ、メタデータが表示されている画面を出力したり送信したりすることも躊躇してしまうといった世界は、恐らく誰もハッピーではないですね。著作権法は相当いろいろな"お荷物"がくっついてくるものなので、簡単に適用範囲を主張する・広げるのは控えた方がいいのではと考えています。

**武田**：会場からの質問の中に、ダウンロード可能なデータを論文に使ったら、公開元から不正使用の申し立てがあったという事例が書かれていましたが、まさに「ライセンスを付けて公開してくれ」ということですね。利用条件が明示されていて、初めて安心して使うことができます。逆に言うと、ライセンスが付与されていないと使えないというのが今のデータの世界です。どんなかたちでもライセンスは付けてほしいですね。

**中村**：少し質問を変えて、福田さん、山崎さんにお聞きします。ウェブ上のデータを扱うときはライセンス確認などのリテラシーが求められます。また ToDH では様々な技術を使った取組があると思いますが、そういうスキルを学ぶ機会があったのでしょうか？ また、そういうリテラシーを活かして研究することの難しさなどがあればお聞かせください。

**山崎**：まとまって技術を学ぶ機会は、特になかったです。ライセンスについても、とにかくしっかり読み込むなど、その程度の対応です。



山崎翔平氏

<sup>3</sup> <http://current.ndl.go.jp/node/29713>



**福田：**一般的には山崎さんがいう通りで、系統的に学んだことはないです。先ほど紹介した、私が使った公文書館のデータは著作権がないものですが、分かりやすいライセンスが付与されていないので、確かに利用者側としてはそこを明確に表示してほしいと思います。



福田真人氏

**中村：**ではライセンスの話題に戻って、パブリックドメインだけでも利用時は所蔵機関名の明示をお願いしたい、しかしそれを的確に伝えられるマークがないという現状があり、渡辺先生からマーク策定についての提案がありました。これを実際に進めるとすると、どういう組織が取り組むのが良いでしょうか。

**渡辺：**お願いの標準化、お願いマークの件ですね。関連質問として「海外でも同じ問題が起こっていますか」というのもありました。把握している限りですが、国外では多くはないように思います。また、どういう機関がマーク策定を進めるべきかということについて、私はデジタルアーカイブの運営当事者ではないので僕一人が何かしようとしても駄目で、関係者の皆さんが「作りたいね」と言ってくれる必要があります。これに対する CC ジャパンのメンバーの意見は様々で、パブリックドメインに属するはずの書籍に all rights reserved などとあっても無視されるようにライセンスがあっても無視すべき時は無視するようになるのでマークの標準化も不要なのでは、という意見もあります。ただ個人的には、そういう曖昧な対応よりはマークを策定した方がいいと思っています。ちなみに「ライセンス」は、権利を持っている人がそれに対し

必要に応じて許可するものです。当事者が権利を保有していない場合に使うものは、CC では「リーガルツール」とよんでいます。お願いマークも、ある意味ライセンスとは違う何か、コミュニケーションツールの一つと捉えられると思います。そしてこれが日本特有の問題ならば CC 全体ではなく CC ジャパンの課題解決として進め、もし海外でもニーズがあるならマークを使ってもらおうという進め方が良いように思います。あるいは、CC ではなくコモンズフィア<sup>4</sup>がプロジェクトとして進める、他の団体が進めて CC ジャパンのメンバーがそこに参画するなど、いくつか体制は考えられるので、そこは今後の検討だと思います。

**武田：**私は、実は、お願いマークは必要ないと思ってます。そもそも研究は、引用が最大のリスペクトであるし、そこをお願いされたり指図されるのは、ちょっと違うかなと感じますね。

**中村：**なるほど。その一方で FAIR 原則から考えると、共通のマークがないことで現状は相互運用性を下げているという面もあるように思えるのですが。

**武田：**欧州などでは、政府や学術支援財団などが非常に強く、もはや過激と言えるほどデータのオープン化を FAIR 原則のもとで行っています。そういう流れが既にあることも念頭に置いておくべきかと思います。

**永崎：**実際に、文化資源のデジタル化にも多額の資金を提供しているアンドリュー・メロン財団<sup>5</sup>は、年間 300 億円ほどを世界各国に配分していますが、成果は原則オープンにと定めていますね。それから話を少し戻して、お願いマークの件は、是非これは機械可読のものとしてほしいです。例えば最近 Rights Statements<sup>6</sup>を採用する機関もありますが、現状の Rights Statements ではどのような法的な制限があるのかは、マークだけでは識別できません。お願いの内容が正確に伝わるマーク、つまり URL があってほしいと思います。

**中村：**ここで武田先生にお聞きしたいのですが、オープンデータが増えれば増えるほど探しにくくなる、ということがあります。そこでデータへのアプローチ、検

<sup>4</sup> <https://creativecommons.jp/about/>

<sup>5</sup> <https://mellon.org/>

<sup>6</sup> <https://rightsstatements.org/en/>

素などについてのご意見をうかがえますか。

**武田**：データに付けるメタデータの記述としては Data Catalog Vocabulary(DCAT)<sup>7</sup>が普及していて、Google Dataset Search も DCAT 相当のメタデータで提供されています。また、データを公開するソフトウェアでは、英国や日本でも採用されている CKAN<sup>8</sup>というオープンソースソフトウェアがあり、それをを用いる方法が一つあります。また当然ながら、有名なデータリポジトリにデータを置くと検索されやすくなるので、どのリポジトリに置くと注目され、活用されるかということを開示する側も意識して選んでいますね。昨今は、大学や学会、出版社など様々な機関が競ってデータリポジトリを構築している状況もあります。

**永崎**：人文社会系分野では、日本学術会議による「人文・社会科学データインフラストラクチャー構築プログラム」<sup>9</sup>が今年度からスタートしていることが大きな動きとしてあります。一方 Brewster Kahle 氏の Internet Archive<sup>10</sup>を活用したり、Wikimedia を使うことも選択されつつあります。文化資料は政治的な理由で非公開にされてしまうという危険性が常にあるので、そういう影響が及ばない場所にデータを置いておくという意味でも、一つ大事なことかと思えます。

**武田**：それに関連して、昨年「Code for 選挙：衆院選 2017 候補者データベース作成プロジェクト」<sup>11</sup>というのが作られて、選挙候補者のデータを Wikidata にも置いたという例があります。必要なデータがまとまってない、公開されていない、だったら自分達で作ろう、それを Wikidata に置こう、という面白い取組ですね。



中村覚氏

**中村**：ちょっと視点を変えて、山崎さんと福田さんに聞きますが、データ共有についての意識、オープン志向が広まりそうだといった感触はどうですか。ご自身の分野のことで結構ですので。

**山崎**：経済史では、そもそもデータを手入力で作るところから始まるという状況なので、データ作成のコストが高いと言えます。それを他人が自由に使うというのは、まだ少し抵抗があるかなと感じますね。研究の再現性を担保するためにデータを公開することはあると思いますが、今のところ実際には少ないように思います。

**福田**：利用する側としては、データは便利に使えるようであってほしい。でも自分のデータを公開するのは手間やコストがかかるので、オープンかということ、正直そうならないと思います。大規模な科研費の成果データなどはオープンになっているとは思いますが。逆に言えば、オープンにしても評価される仕組みが無いのも要因で、評価サイクルがうまく機能すれば状況は変わっていくようにも思います。

**武田**：永崎さんに質問だけど、仏教学分野は、どうしてこうもオープン志向なんですか？

**永崎**：そうですね、色々な理由があると思いますが、一つの理由として、海外からの圧力もあるように思います。仏教学分野は国外に大勢の研究者がいて、その人達が「もっと便利に」「使えるものを」と要求してきます。また「英文学の分野では、こういうこともできているのに」など、他分野を引き合いに言われたりもしますので、そういう要求があって、要望に答えているうちに、ということでしょうか。

**武田**：なるほど、その分野の人が外や隣を見たりして、自分たちも、となるわけですね。

**中村**：さて、そろそろ時間もなくなってきたので、最後に大学等におけるオープンデータの取組について一言ずつお願いします。

<sup>7</sup> <https://www.w3.org/TR/vocab-dcat/>

<sup>8</sup> <https://ckan.org/>

<sup>9</sup> [https://www.jsps.go.jp/data\\_infrastructure/](https://www.jsps.go.jp/data_infrastructure/)

<sup>10</sup> <https://archive.org/>

<sup>11</sup> <http://election.code4japan.org/>

**武田**：分野によってデジタル化の進行には違いがありますが、10年のスパンで考えると、どの分野も大きく様変わりしているのは確実です。もはやデジタルの力を使わざると得ないと思うし、使えばより良い研究ができるというのは当然と言えます。自分達の研究、学術の世界はそういうところにあることを見越してデータの公開に取り組むべきかなと思います。

**渡辺**：オープンデータの取組でも、積極派と懐疑派がいて、積極的な国の模範的な政府機関でも同じだったりします。少しずつしか動かないところがあります。懐疑派の中には、オープンにしないことのデメリットや、特定のデータについてのニーズやコストベネフィットを説明して、やっと動いてくれる人もいます。積極派は、説明すら不要で賛同してくれたりする。だから、うまく分かり合える人とネットワークを作っていくことが重要で、その中で外部への説得材料として、オープン化の優良事例や効果の推定情報などを収集し、より良いコンセンサスを作っていくのが大切かなと思います。

**永崎**：会場にいる人は、ある程度デジタルアーカイブやデジタルコンテンツに親しんでいる方々だと思いますが、一般の人がデジタルアーカイブを見ても「よく分からない」となりがちです。そのため、専門家がもう少し噛み砕いて説明するなど、そういう取組もうまく回していけるような仕組みが必要と思っています。Europeana や Digital Public Library of America では、公開後の活用方法を検討する人材が確保されています。残念ながら日本はそこまでの体制がなく、そうい

うポストを作るのも大変なので、急にそこまで望まなくても魅力が伝わるように画像や情報を公開し、そういったポストの必要性なども広めていって、少しでもデジタルアーカイブが広く利用される流れになると良いかと思っています。

**山崎**：データを使う側にとっては、「公開してください」としか言えないのですが。データを公開すると、提供者側が想定していなかった活用のされ方が起こり得るし、使う側はどのような形であれ公開されていれば使うものなので、とにかくデータは公開していただければ。

**福田**：永崎先生から「提供者の責任が大きくなる」という話がありました。講演ではメタデータの誤り発見の話をしたのですが、こうした誤りを指摘されるため、提供者に過度な専門性や責任が要請されて、公開が遅滞するよりは利用者がむしろ積極的にデータを修正したり、それをオープンにしていくといったシステムを作り、オープンサイエンスを提供者と利用者が共同して作り上げていくことが今後求められるのかなと思います。

**中村**：データだけでなく、利用者側の意見なども含めてオープンになり、それをまた別の利用者を使って、という良いサイクルが望まれるということですね。さて、全ての質問を拾えず申し訳なかったのですが、時間になりましたので、これでパネルディスカッションを終了とします。

