

## メタデータのオープン化等検討ワーキンググループ(第3回) 議事概要

日時：平成28年11月14日(月) 15:00～17:10

場所：中央合同庁舎4号館 共用1202会議室

### 【議事】

1. オープンデータ政策との関係からみたデジタルアーカイブのオープン化
2. 仏教系デジタルアーカイブの事例報告
3. メタデータのオープン化等に関するガイドライン(素案)の骨子について

### 【概要】

1. オープンデータ政策との関係からみたデジタルアーカイブのオープン化

○大向参考人より、資料1に基づき説明。

○質疑の内容は、以下の通り。

(高野主査)

- ・メタデータが誰のものかという主体の欠如というのは大きな問題である。文化遺産オンラインでも第三者にデータを渡すことが前提になっておらず、提供側も意識していない。そのため、後付けで整備するのは非常に難しい。最低限のラインについて、ガイドラインを作る過程で議論していきたい。

(ゼノン・リミテッド・パートナーズ 神崎代表)

- ・資料1「オープン化への期待と課題」の中にあるメタデータ項目ごとの階層的なライセンス設計についてももう少し伺いたい。

(国立情報学研究所 大向准教授)

- ・具体化項目にもいろいろあるが、本であればタイトルなどになりあまり議論はないが、論文では抄録をどうするか、著作権がありうるものとなるとライセンスを付けて配布することが難しくなるといった議論がある。デジタルアーカイブであればサムネイルであればこのような取扱い、コンテンツそのものはこのような取扱いというそれぞれの取扱いとライセンス上の取扱いを変えていくことを指している。

(ゼノン・リミテッド・パートナーズ 神崎代表)

- ・テクニカルにそのような仕組みはあるか。やるとするとデータセットから分けることになるか。

(国立情報学研究所 大向准教授)

- ・仕組みはないので、最終的にはそうなるか。

(東京大学 生員客員准教授)

- ・ガイドラインを作る際には、アーカイブを実際に運用する図書館や美術館といったところの人たちが、実際に、どういうところに何をどういう条件でメタデータとしてつけばよいのか、ということがわかりやすく説明されるべきと思う。
- ・Europeana では、データ交換協定で提供されたメタデータは CC0 だと規定されている。だが、先ほどのお話のように「抄録はメタデータとされる場合がある」としたとき、さすがに提供する側がそれに CC0 はつけられない。そうした微妙な区別について、提供する側が、ここからここまでは明らかに CC0 つけてもよいと判断できるようにしなくてはならない。ガイドラインでは、メタデータはすべて CC0 しなくてはいけないわけではなくて、連携に必要な部分について、第三者の著作物性等々といったことを考慮して、書かれるべき。運用する側が迷うことのないように、ある程度の場合分けのようなものをつけたうえで、書きぶりを考える必要があると考える。
- ・CC0 の法的な詳細をある程度詳しく伝える必要があると考える。たとえば、先ほど CC0 では全ての権利を放棄するということが法律制度でどのような意味を持つか、という話があった。日本では著作者人格権は放棄できないので、そのこととの兼ね合いが問題となる。この点では、CC0 の権利表記では、権利の放棄が各国の法制度上できない場合について、許諾にする、それでもだめなときは不行使にするといった形で、各国の法制度事情に対応できるような、複層した書き方になっている。そうしたことをどう説明するかも検討される必要がある。

(高野主査)

- ・メタデータ項目のうち、これは CC0 だけどこれは CC-BY とか、このフィールドは評論家の名前も入っているから変えましょうとか、そういうカスタマイゼーションとか細かい視点も出てきたときに、これとリンクトデータ (LD) のような仕掛けとの考え方になると、リンクトデータみたいにばらしてもっていくときれいに整理されたりするのか。

(ゼノン・リミテッド・パートナーズ 神崎代表)

- ・ライセンスを細かくすると逆に迷いが生じ、使いにくくなるため混乱するおそれがある。重要なポイントではあるが、検討する必要がある。

(高野主査)

- ・LD 的なアプローチが必要だというのは、全てスライスしてシュレッターしてしまえばファクトしか書いてないので著作権も何も切れてしまう。なのでフリーにしまっただけで、集めたらきれいになったものが復元できましたラッキーみたいな乱暴な話はないか。

(ゼノン・リミテッド・パートナーズ 神崎代表)

- ・リンクトデータでライセンスを使い分けるなら、ライセンスごとにデータを異なるグラフに分け、それぞれのグラフに対して適用ライセンスを記述するなどの方法になるが、説明が難しくなってしまう。

(国立情報学研究所 大向准教授)

- ・CC0 にしてよいファイルの一つ作り ID を付け、抄録と ID しかついていないもう一つのファイルを作りこちらは CC-BY ですよ、とすることで、それらの再構成により元のデータに戻すことは可能かもしれないが、実は大変な作業になる。

(国立歴史民俗博物館 後藤准教授)

- ・ライセンスの切り分けについて、博物館だと、フィールド等のデータレベルで分けられないのではないかという問題がある。特に、サムネイルで、これはOK、これはダメというコンテンツの中身レベルで議論しなければならない部分がある。技術的な問題の他、コストが見合わないというプラクティカルな問題もある。もう少し分かりやすい切り分けができるとうい。
- ・過去分についてデータの主体が明確でないという点について、歴博ではほぼ全部その通りということになる。一般論として、多くのものに権利がないことを前提として、割り切るような言い方も考える必要があると思う。

(高野主査)

- ・博物館や美術館でよく問題になるのが、所有と言っても寄託を受けているだけで所有者は別という、提供をお願いしている立場からすれば逃げをうたれているようにしか聞こえないが、博物館・美術館の立場からすれば理のあること。なので、寄託を受けるとそれについてのデジタル記録をとったり、目録に載せたり、美術館が主体的に何か活用できることも含めて寄託を受けているんだ、という整理をしておいてもらえると、活用しやすくなるのではないか。

(国立歴史民俗博物館 後藤准教授)

- ・国立歴史博物館は寄託がほとんどないため、あまり問題とならないが、国立文化財機構等は寄託が多い。とは言え、寄託の話は、事実上、最後の砦のようなところがある。むしろ、館蔵資料をきちんとやっていくモデルを作る方が先と考える。

(京都府立図書館企画総務部企画調整課 福島副主査)

- ・後藤さんの意見に同意。主査のおっしゃることもよく分かるが、まずは寄託のことは入れないところから議論をしたほうがよい。基本的な整理としては、所蔵品を主眼としておくべき。

(東京大学 生貝客員准教授)

- ・今までに作られたデータと、これから作っていくデータとを、ガイドラインの中でどう分けるかを考える必要がある。これからの部分については、寄託元や、メタデータを作成したり写真を撮影したりするような外注先などと、どのような取り決めをするのか。この点について整理しているところはさほど多くないだろう。フォーマットや手順を示す必要がある。

(国立歴史民俗博物館 後藤准教授)

- ・所有者の問題に関し、博物館が必ず持つものとしてレプリカがある。寄託を受けなくても、一旦預かってレプリカを作って返却した場合、レプリカのデジタルデータの権利を誰が持つか。現状、複製権の取り扱いに関する契約が非常に曖昧な場合が多く、整理が必要である。

(高野主査)

- ・基本的な確認だが、過去にこういうライセンスで公開したが、状況の変化に応じてライセンスのアップデートをすることで、厳しいものを緩くするだけでなく、緩くしたが厳しくするという変更も可能か。

(東京大学 生貝客員准教授)

- ・原則論で言えば、いちど許諾を得たり契約をしたりしたものについては、事後的変更をしようとしたらそれなりの手続きが必要であろう。

(国立情報学研究所 大向准教授)

- ・ライセンスの変更には同意が必要となる。ある瞬間の同意が本質ではない。意思決定におけるガバナンスが効いていることが本質である。一回で一つの形に落とし込むことはできないが、政府標準利用規約でも周りの反応を取り込んでよりよい形にしていくことが当初から作戦として予定していたはずである。ここでも外の人の声が入るにようにする作戦をもっておくべき。

(高野主査)

- ・だんだん緩めていく方向については主体がはっきりしていれば、やり続けることは問題ないし、一種啓蒙的な活動としてやりながら実証して例題を示しながら進んでいくということが出来るのではないか。

(東京大学文書館 森本准教授)

- ・国立公文書館の取り組みはオープン化する方に進んでいる。公文書管理法では閲覧室やHPで公開しているものほどのように使っても自由であるとなっており、許諾も不要である。

(高野主査)

- ・素晴らしい気付き。実は公文書館が一番進んでいる。
- ・それをうまく活用したサイトなりサービスなりをあげていく人がいれば価値を築いていける。

## 2. 仏教系デジタルアーカイブの事例報告

○永崎参考人より、資料2に基づき説明。

## 3. メタデータのオープン化等に関するガイドライン（素案）の骨子について

○知財事務局及び国立国会図書館より、資料3に基づき説明。

○質疑の内容は、以下の通り。

(京都府立図書館企画総務部企画調整課 福島副主査)

- ・論点1の図の書きぶりについて。メタデータとサムネイルなど画像の束があって、それがAPIではき出されるという基本的なイメージはこれでよいが、メタデータをポータルサイトが集約するように描かれているところは、これはショーウィンドーのようなものと伝わるようになるとよい。現状の図だと、個別機関が一階的に書かれているが、もうすこし別のイメージになっていると、説明がしやすい。
- ・論点2の「束ね役」、アグリゲーターについて、これはそれなりに負担がかかる構造であると思われる。それだけに、アグリゲーターを担うことのメリットを示せないか。当初はアグリゲーターを名指しで決めるのだろうが、それから分野などを広げていくときに、さらに増えていかないとならない。そのとき、みんなが手を挙げやすくなるようにできないか。

(高野主査)

- ・先週横浜で行われた図書館総合展でも話題になっていたが、データがNDLに吸い上げられてしまうだけ、貢献はしたがこっちは何も変わらないというのであれば、どんどん参加しましょうとはならない。インセンティブという程ではなくても、何かフィードバックがちゃんとあって、よりリッチになっていくよ、とか、あるプラットフォームが充実するのでそのプラットフォームをプラスアルファ活用することによってやりたいことがやりやすくなるよとか、プラスアルファの部分を考えていきたい。

(ゼノン・リミテッド・パートナーズ 神崎代表)

- ・アーカイブを作るためのガイドラインとするのか。メタデータをオープンにするためのガイドラインなのか。若干話がぶれている印象である。例えば IIF がアーカイブに必須かと言われると、あればよいが、それは視点が違う。アーカイブを構築するための設計方針と機関がメタデータを出すというのも別の話である。機関の人

たちが読むためのガイドラインとすべきなのではないか。

(高野主査)

- ・三つの側面、みたいな方が良いかもしれない。集めてきちっとフィードバックをすることと、外からのファインダビリティを保証するためにメタデータ等を整備すること、発信するときにスタンダードを意識してライセンスをきちっとすること。

(国立歴史民俗博物館 後藤准教授)

- ・システムについての話と、データ（ライセンス）についての話が混在している。特に、束ね役についてはシステムの話が多い印象。束ね役と機関を分けるのに苦労しているのは、そのためかもしれない。
- ・例えば、アグリゲーターについて。人間文化研究機構では本部は自分のメタデータをほぼ持っておらず、各機関のデータを預かり上に繋ぐモデル。その際、人間文化研究機構本部が採用すべきレベルは、各機関レベルなのか束ね役レベルなのか、少し分かりにくい。

(高野主査)

- ・たぶん図は相当工夫する方が良い。階層が見えて子が親に全部出すと見えなくもない。NDLサーチはメタデータしかとっていかない、と僕らは読むので丸々とっていくわけではないと分かる。
- ・NDLサーチはメタデータに限定されたアグリゲーターだが、もっと違う、地域に関連したものばかりを写真も含めて深く集めるといったアグリゲーターがNDLサーチと同じ様に立っても構わないという立場だと思うので、発信の仕方を適切にしておく色々な組み合わせで、国や地方といった様々なレベルで集め直したり集約したりすることができる、そういう自由度が担保されたプラットフォームができるというのが美しいストーリー。絵を描くのは難しいかもしれないが、もう一つぐらい階層があってもいいかもしれない。
- ・ここでメタデータ、サムネイル/プレビュー、デジタルコンテンツといている部分はあたかも明確な区分があってそういう階層があると。最初ほど情報量が少なくてだんだん情報量が増えていく。しかし現物はもう少し向こうにあって。区切りというのは報告書では名前を付けて区切りをつけなくてはならないかもしれないが、分野によっても全然呼び方が違うので、区切りは技術が変えていくということも考えると、分野を越えてやろうと思うとこの辺も揺らいでいくのではないか。
- ・ここは今何か頼りになる区切りはあるのか。ここまではコンテンツデータ、ここまではメタデータといった。今流通しているデジタルデータは全て現物じゃないからメタデータといってもいいのではないか。

(ゼノン・リミテッド・パートナーズ 神崎代表)

- ・メタデータもデータであるが、どういう視点でデータを活用するのかということ

名称は変わる。そこまで気にしなくてもよいのではないか。

(高野主査)

- ・この報告書のメタデータといったときの範囲は、デジタルアーカイブに入っているデータは全てメタデータといった主張もまんざら成立しないことではないのではないか。
- ・メタデータの名前に引っ張られて悩んでいたが、解決するかもしれない。

(ゼノン・リミテッド・パートナーズ 神崎代表)

- ・ナショナルアーカイブだけでなく、色々なところで作られるデジタルアーカイブもこのガイドラインの視野に入っているのか。

(高野主査)

- ・アーカイブと呼ばれるものを作ろうとしている人に一定の知見を与えるレポートみたいな感じだと思っていたが違うのか。

(国立国会図書館)

- ・それもあるが、どちらかというところとアーカイブ連携を促進するために必要な要素が重要だと思っている。連携を促進するために必要な要素プラス利活用のためにはさらにデータが自由に使えることということで、アーカイブ利活用といった側面から連携するためにどう構築すべきという話と、ちゃんと連携を促進しましょう、連携促進した暁にはもっとデータが自由に使えるようにきちんとオープンにしましょう、という話がセットでこうあるべきというガイドラインになれば良いと考える。

(高野主査)

- ・主に充実した連携を引き出してこの国のそういったところを豊かにしていこう。そのために個々のアーカイブ作ろうという人はこういうところを注意してね、これはよくできたアーカイブですよと紹介できればいいかなと思っている。
- ・完全にこれだけに従っていればいいわけではないが、こういう要素に気をつけて作ってねという感じか。

(京都府立図書館企画総務部企画調整課 福島副主査)

- ・京都府内の各機関に持ち回って説明することを考えると、最低限の要素を抜き出したような付録があると、やりやすいと思う。そこに図が入っているイメージ。すごく平易なものでよい。こうしたガイドラインを全部読めと言っても、なかなか難しいだろう。

(国立国会図書館)

- ・それは10頁の評価基準になっているところか。こういうレベルがいいよね、とい

う話も欲しいし、かつ前回後藤先生から指摘のあったよくある質問もまさにそのためのもではないか。FAQ と評価基準をセットで示せばよいというイメージ。この辺もご意見頂きたい。

(高野主査)

- ・デジタルアーカイブを色々な角度から採点するような評価基準というかメトリクスがあって、是非つながったらいいいよね、という感じか。採点して歩く気はないが、いいものを引き上げる基準としては使えそう。分野が違って知らなかったけどこんな良い特性があるなら是非つなげてみようと思えるような判断基準が出来れば良い。連携しやすいとか主体がしっかりしている、そもそものコンテンツが素晴らしいとかそういった次元になると思う。
- ・森本さんが他の人が作ったアーカイブを研究なり調査なりに使おうと思ったときに、外形的にこれは使えそうとかこれは構造が全然違うよね、とかそういった判断基準はあるのか。目的を達成するためには多少の構造の違いは乗り越えていかないと、という感じか。

(東京大学文書館 森本准教授)

- ・公文書分野の資料は、資料がユニークで形も多様である。だからこそ、構造が把握できるような目録が整備されている。その上で、便利な検索システムが準備されていると安心して活用ができる。

(高野主査)

- ・各国ごとに住所の表記の方法や地域の構成とか違うけど、その国に行ってそのシステムを学べば土地勘が生まれるとかそういう感じか。

(東京大学文書館 森本准教授)

- ・公文書はかたまりであるという共通認識があることで活用しやすくなっている。一点だけの資料が出てくると不安になる。

(高野主査)

- ・この委員会の司会をやりながら危惧しているのは、最終的におさめる先がNDLで図書の専門の知識の体系を持っていて、それについての絶対のオーソリティで、他の人たちはそうではないバリエーションをかなり持っているというときに、そこに寄せていくことで図書っぽいメタデータには整理されるが、博物館の人たちにとって自然な知識体系にはならないということが可能性としては起こりうる。
- ・同じものを集めても整理の仕方が色々違う。アーカイブズの人を使いやすい、なじみのある整理の仕方もあるだろうし、いかにもミュージアムっぽいコンテンツの分野みたいなものを勉強しそれこそが重要であるといったこともあるだろうし、そういったことが起こりうることを講釈だけでも書き込んでいけば色々なことを



検討しているのだとわかってもらえるのではないか。

(東京大学文書館 森本准教授)

- ・ガイドラインの構成という点で、用語の定義だけ独立させて充実させると分かりやすいのではないか。

(国立歴史民俗博物館 後藤准教授)

- ・評価項目について、URI/URLの明確性は必須レベルで入れる必要があると考える。本来やらないといけないことは、連携しやすいデータはどのようなものか、分かりやすく示すこと。そのために、数値目標等よりは、こんなデータが良いということが、もう少し出せるとよい。
- ・コンテンツの拡充に関し、大前提の総論として、自分たちの組織のものをデジタル化して表に出すことの重要性をもっとしっかり明記した方がよいと思う。誰かに任せっきりでなく、自分たちによるデジタル化の促進も考えた方がよい。
- ・これまでのデジタルアーカイブ構築事例で、デジタル化まですべて外部委託してしまい、自組織のデータなのに自由に公開できなくなってしまったものがある。そのようなことにならないような表現を検討してほしい。

(京都市立図書館企画総務部企画調整課 福島副主査)

- ・以前に自分が関わった契約で反省するところがある。契約についての最低限のガイドラインが必要。この点についての書きぶりも考える必要がある。

(高野主査)

- ・担当者がそういうところに感度が低いと散逸してしまう。デジタル化は向こうでもらってもいいのだけれど、成果物については共有するとか向こうが使うのをやめたら自由に利用できるとかあった方がよいのでは。

(東京大学 生貝客員准教授)

- ・ガイドラインそのものについてというより、位置づけについて。このガイドラインはスタティックなものでなく、今後もダイナミックに変わりゆくものであるということ念頭におく必要があると考える。たとえば Europeana も DPLA でも、標準的なライセンスは当初は CC0 ではなかったが、それが数年後に現在のように変わっていった。
- ・統合ポータルが存在することの意義は、そのポータルとしての機能においてだけでなく、ここで話し合っただけでルールが決められる、変えられるというマルチステイクホルダー・フォーラムであるということに、最大の意義があると思う。
- ・ルールを変えるときは、二次ルールを考える必要について先に述べたが、こうした動的な部分も最後のほうに入れておくべきではないか。

(高野主査)

- ・全部中央で決めるとはならないと思う。アグリゲーターなりサブアグリゲーターなり信用しあっている人たちが集まったコミュニティで最終的に判断してもらい、そういうものを作っておこうねということ。誰も決めない鳥合の衆がどのレイヤーまでいっても続くというのは不幸。

(ゼノン・リミテッド・パートナーズ 神崎代表)

- ・ガイドラインの項目案についてはそれぞれの機関を尊重することを重視していると思うが、マッピングを前面に出しすぎるのは危険ではないか。マッピングすることで、元データを変更されてしまうと捉えられてしまう。そうではなく、提供された元データに新たなラベルを加えるという理解なり説明の方がよい。

(高野主査)

- ・やることは同じでも元データは書き換ええない。中身の情報を切りだしたりしてメッキをつけてあげるといったことか。

(人文情報学研究所 永崎主席研究員)

- ・インターフェイス上も実現できると思う。話が戻るが、デジタルアーカイブとは何かということをはっきりと定義すべきである。
- ・今までアーカイブズ、コレクション、キュレーション、データセットなどと言ってきたものをデジタルアーカイブとしてまとめようとする取組であると思うが、私の把握している限り専門家の間でも理解が得られていない。色々な取り組みを含めてデジタルアーカイブであると明示、押し出した方がよい。

(高野主査)

- ・森本先生が先ほどおっしゃられたように、同じことを違うニュアンスで語っている人もいるだろうし、ここで意味が定まっているように使っている言葉でも、人によってはずいぶん違うコンテキストで理解していることもある。用語集をうしろの方で書いたら良い。
- ・親会には放送の人もいるし、文化財のデジタルアーカイブに限った話をしているわけではない。全てに共通する活動としてとらえている。たまたまNDLにデータを入れるのが文化財オンラインのデータを入れてはどうかという話はしているが、この委員会の範囲としてはかなり広いものと考えている。

(国立情報学研究所 大向准教授)

- ・このガイドラインは個別機関に参照してほしいと思って作るのだと思うが、目的のところどのくらい interoperability を取り入れられるかが重要である。大きなものを作っていくために各個別機関が働くわけではない。
- ・個別機関からみて interoperability を取り入れているのであれば、負担が増えて

も納得できる。その意味で建付けは気にしなければならない。主役となる個別機関を主語として考えていかなければならない。機関からみれば、自分だけ中心にあるので、それを踏まえた図にしてもよい。評価については、非常に唐突感がある。誰が誰を評価するのか。

(高野主査)

- ・この図はこれでもよくなった。昔は一番下にあっけいかに搾取されているように見えて評判悪かった。
- ・まずは自己評価。自分たちの発信の仕方がどれくらい開くという形で準備できているか、無理のないもう一步を進められるなら進めてみましょうというもの。
- ・とはいえ内閣府の下であるので取り方はいろいろ。気を付けていきたい。
- ・ガイドラインという用語は分野ごとのバリューが入っている。

(人文情報学研究所 永崎主席研究員)

- ・評価基準については、アメリカの文学(MLA)と歴史学の学会(AHA)などが示しているものを参考にしてもよいかもしれない。

(ゼノン・リミテッド・パートナーズ 神崎代表)

- ・評価することは良いが、それが独り歩きしないものとしてチェックリスト方式の方が良い。

(高野主査)

- ・チェックリストでありそのアーカイブの特性がわかる診断みたいなことにも使えるといい。チェックリストは適切である。

(京都府立図書館企画総務部企画調整課 福島副主査)

- ・趣旨は共有した上での意見だが、デジタルアーカイブ側の発想では、「チェックリスト」まで落としてしまうと、そこまでリソースが回らなくなる。このガイドラインは、各機関がアーカイブ化のためのリソースが回っていない状況の中で、背中を押すためのものだから、入館者数・来館者数以外の指標を与えるものであるべき。

(高野主査)

- ・あまり独善的にならないように、かつ国として応援したいという活動をしたい人を応援できるような発信の仕方を考えている。

#### 4. その他

- 次回、最終回となる実務者協議会は、12月12日月曜日16時から18時までの予定で開催する。

以上