

メタデータのオープン化等検討ワーキンググループ(第1回) 議事概要

日時：平成28年9月12日(月) 14:00～16:10

場所：中央合同庁舎4号館 共用123会議室

【議事】

1. 本ワーキンググループの運営と進め方について
2. 実務者協議会の検討状況及び検討課題について
3. 海外の事例報告
 - (1) 欧州の事例報告
 - (2) 米国の事例報告
4. 「メタデータのオープン化等に関するガイドライン(素案)」のスコープについて

【概要】

0. 開会

○高野座長より挨拶

- ・このワーキンググループは、ガイドラインの内容を実務的に詰めていくためのもの。デジタルアーカイブの構築・連携だけでなく、アーカイブが長続きするものとなるよう、利活用に必要なことも検討していきたい。

1. 本ワーキンググループの運営と進め方について

○事務局より、資料1-1、資料1-2及び資料1-3に基づき説明。

○質疑の内容は、以下の通り。

(ゼノン・リミテッド・パートナーズ 神崎代表)

- ・ガイドラインの読者としてどのような人たちが想定されているのか。

(知財事務局)

- ・まずアーカイブ作成者が対象。その他、利用者の立場からも有益になる情報を盛り込みたい。

(ゼノン・リミテッド・パートナーズ 神崎代表)

- ・メタデータの観点となると一般的な利用とは切り口が異なる。アーカイブ作成者がいきなり自分でメタデータを設計することは想定しにくい。

(高野主査)

- ・訴えたい先は、デジタルアーカイブを作ろうと決める地位のある人であり、その人にインセンティブを与え、アーカイブの必要性や有効性が伝わることで、その協議会を含めたこの会合の意義である。

- ・タイトルはメタデータのオープン化「等」としているが、ここで議論したいのはメタデータに限定するものではなく、デジタルアーカイブをより利用しやすい形のものに作り上げるには、どういうところに留意すべきか、ということである。その最も重要なところはメタデータのオープン化かもしれないが、世界的にみてもメタデータのみを公開してそれで終了というサービスはないのであるから、プラスアルファで、商業的なサービスとも繋ぐなど、真にユーザーに活用してもらえそうなサービスに仕上げるためには何が必要なかを議論していきたい。
- ・本日お呼びした二人のスピーカーに成功事例を紹介いただくが、それはそのサービスの中でメタデータがどうなっているかではなく、プラスアルファで何をパッケージしているかであり、その結果がインパクトのある仕事になっているかどうかを伺いたいと思っている。
- ・したがって、メタデータのオープン化に厳密に拘ってスコープをかけるのではなく、より広げていく方向で議論できたらと考えている。

(ゼノン・リミテッド・パートナーズ 神崎代表)

- ・ということは、このワーキンググループは、メタデータの話だけに限らず、デジタルアーカイブ全体の問題を取り扱うということか。

(高野主査)

- ・アーカイブ全体について、社会的に意義のある形にしていくためには、メタデータを含めたデータをどのようにしていけばよいか。例えば、国際的な標記にすべきとか、利用の権利をCC0にすべきとか、そういう色々なガイドラインが世界の知識と繋がるために必要である、という議論ができるとうれしい。

(京都府立図書館企画総務部企画調整課 福島副主査)

- ・「等」は、議題4にあるスコープの話と連動すると考える。高野先生が言われたように、メタデータからスタートするものだと思っているが、最終的にデジタルアーカイブ全体も視野に入れて作るということか。

(高野主査)

- ・そのつもりである。但し、事務局はアウトプットとしてタイトルに合うものを出したいとの意図はあろうとは思う。

2. 実務者協議会の検討状況及び検討課題について

- 事務局より、資料2-1、資料2-2、資料2-3及び資料2-4に基づき説明。
- 質疑の内容は、以下の通り。

(高野主査)

- ・このワーキンググループを進める立場としては、メタデータのオープン化等といっ

たときに、メタデータのオープン化以外に何をやっていけばアーカイブ全体が盛り上がっていくか、他のアーカイブはどういう工夫をもってそれを実現しているのか、この会議で例示や提案を示し、その結果、海外でも進みつつある一種の標準化を採用することで一気に進むことになれば、皆さんに集まっていた意味があると考ええる。

3. 海外の事例報告

(1) 欧州の事例報告

○生員構成員より、資料 3-1 に基づき説明。

○質疑の内容は、以下の通り。

(高野主査)

- ・先ほど言い忘れていたが、最後に話のあった KPI のような、デジタルアーカイブがどの程度うまくいっているのか、どの程度努力が報われているのかというものを測るメジャーについても、この会合が提案できたら素晴らしいと思っている。

(ゼノン・リミテッド・パートナーズ 神崎代表)

- ・中間報告にも入っている話であるが、権利表記について、ユーザーが Web ページで見る画面に表示するという話とメタデータに組み込むという話の両面があると思うが、両方が並行している話か。

(東京大学 生員客員准教授)

- ・基本的に、権利表記はクリエイティブコモンズをモデルとしている。Europeana では、メタデータを表記する際に利用条件も表記することがデータ交換協定の中でも求められている。メタデータにタグが埋め込まれた状態であり、結果的にそれぞれのコンテンツページに表示されエンドユーザーにも見えることになる。

(ゼノン・リミテッド・パートナーズ 神崎代表)

- ・パブリックドメインマークという言葉は、表示することを念頭においているように見える。

(東京大学 生員客員准教授)

- ・もともとのクリエイティブコモンズの仕組みとして、マシンリーダーであり、ヒューマンリーダーであり、ロイヤリーダブルであることでそれぞれが読めるものを準備することが前提になっている。パブリックドメインマークの場合、通常、権利があることを示すものであるが、この場合は、第三者が見て、権利が切れていることを確認させるためにあるものである。

(高野主査)

- ・あるデータについて、実は権利がクリアされているのに、一般ユーザーが検索エンジンなどを通じてサイトを閲覧したときに、その権利がどうなっているのか、そこで判断できないと利活用に影響が出てくる。データが閲覧できるところには権利の情報も一緒に届けることが Europeana のポリシーでもあるし、クリエイティブコモンズのフィロソフィーでもある。

(京都府立図書館企画総務部企画調整課 福島副主査)

- ・スライドの3頁について、中身をもう少し教えて欲しい。欧州委員会の勧告でオンライン・アクセシビリティの話が出てくるが、それは今議論があったような権利表記の話なのか、それとも、もう少し手前の段階の広い意味での Web 一般のアクセシビリティ、つまりあらゆる環境であらゆる状況の人が見られるというアクセシビリティのことなのか。

(東京大学 生員客員准教授)

- ・スライド4頁の5条が対応している。デジタル化して公開していくこと、誰もがアクセスできるようにすること、アクセシビリティの観点から営利・非営利を問わず、アクセスや利用を増やそうということである。

(京都府立図書館企画総務部企画調整課 福島副主査)

- ・最初に議論になったスコープの話と関わるが、結局デジタルアーカイブは Web 経由で利用することになるので、Web でのアクセシビリティを問題としたい。
- ・総務省が「みんなの公共サイト運用ガイドライン」を出していて、地方公共団体のサイトも国のサイトも 2017 年度末には必ず対応しなさいとなっている。デジタルアーカイブは、画像がきれいに見えることとユニークな挙動をするよう作り込む方向になってしまっているが、もう少しフラットに作っておかないとアプリケーションにも投げられない上、限られた環境でしか動かず、ばたばたと死んでいっている。
- ・要するにそういう意味でのアクセシビリティが確保されることが、最初に高野先生が言われた長続きするということだと思あるので、そこはスコープに入れておかないといけない。公的なファンドを入れるときの一種の必須条件にしておき、そこで規制をかけて変なデジタルアーカイブを作らないようにする。権利表記とかを充実することは必要。しかし挙動などを過剰にリッチにして Web アクセシビリティや継続性を阻害しないという点を何らかの形で入れた方が良いのではないか。

(東京大学 生員客員准教授)

- ・まさに、アクセシビリティの議論の中で、継続性の担保というのは中心的な課題となる。元データを自分のところだけで持つと継続性が維持できなくなるおそれがあるので、共通のリポジトリの中で持つことも考えなければならない。

(高野主査)

- ・膨大なデータを大量に集めている人が公開するプロトコルも、技術として目新しいものはないがソーシャルインパクトはある。大きな画像を管理している組織が IIIF の形式でも発信しようといった途端に、世界中の貴重なものが繋がりだし、部分的に引用されるようなことが起き始める。そういうようなことを、この会合で提案していけたらと思う。

(東京大学 生員客員准教授)

- ・IIIF のように API で外部からのアクセスに対応する場合、法的にみて自由に利用できないと意味がない。IIIF のようなものに乗るという背景にはオープンにすることにつながると思う。

(国立国会図書館)

- ・4 頁目のパブリックドメイン作品の自由利用が進まないのは、何が課題となっているのか。

(東京大学 生員客員准教授)

- ・レポートにも各国事情が書かれているが、これまでの慣行やビジネスモデルのこともあるが、状況としては日本とそれほど変わらないと思う。

(高野主査)

- ・国立国会図書館のデジタル化資料も、保護期間が満了し本来自由な条件で配布できるようなものであっても、国立国会図書館が提供するリーダー経由で小さな画像を見ることしかできない。国立国会図書館の中で見た世界と外の図書館で見た世界とでギャップがある。これには諸事情があるだろうが、本来データはもっと自由になれるのに、実際はなっていないことは多々あると思われる。

(2) 米国の事例報告

○時実参考人より、資料 3-2 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)

- ・スライド 18~20 は画像のリンクの話、これは利活用のところモデルになっているが、日本ではできていない。
- ・
- ・スライドの 18~21 ぐらいに Wikipedia との関連が書いてある。私がいる京都府立図書館でも Wikipedia タウンを先陣きって取り組んでいる。しかし、Wikipedia タウンは非常に手間がかかるように思う。やるべきだと考えるが、向こうの実態としてどれぐらいの広がりで行っているのか。

(東京大学大学院情報学環 時実高等客員研究員)

- まだ広がっていない。ここで挙げた例は米国国立公文書館で行っているもの。米国国立公文書館でWikipediaを集めて何か新しいページを作らないかと聞いたところ、公文書館のデータ使って下さい、デジタル化もやって下さいということになった。考え方としてWikipediaタウンと似ている。Wikipediaタウンの取組については日本が一番進んでいる。

(京都府立図書館企画総務部企画調整課 福島副主査)

- 実態が分かったのが良かった。人的・時間的コストが下がって、いろいろなところでやれば良い。

(東京大学大学院情報学環 時実高等客員研究員)

- デジタルアーカイブができると、紙の運搬作業など大変な作業が減り、コスト削減につながる。

(ゼノン・リミテッド・パートナーズ 神崎代表)

- DPLAにおけるメタデータの提供方法は、API経由に限定されているのか。

(東京大学大学院情報学環 時実高等客員研究員)

- メタデータの提供は基本的にAPIである。Europeana、DPLAはいずれもAPI上で画面を作っている。そのAPIは誰でも使用できるため、例えばコンテンツの権利情報の入手も簡単にできるし、例えばパブリックドメインのコンテンツという限定でのキーワード検索という画面も簡単に自分で作ることができる。

(ゼノン・リミテッド・パートナーズ 神崎代表)

- ちょっと使ってみようと思っても、APIのキーをリクエストする必要があるため、最初からDPLAを使おうと思っている人でないと使いにくい。Europeanaの場合は、そういった制限はなく、一つのURIを与えれば、事前の了解に関係なくメタデータを取得できる。

(東京大学大学院情報学環 時実高等客員研究員)

- おそらくサーバ容量の関係ではないかと思う。大量のダウンロード要求があると、すぐにサーバがダウンする。普通にサイトを作り一般の方に公開する分にはあまり問題とならないが、専門的な分析に使用されると何百万件というレベルでダウンロードされることもある。

(東京大学 生貝客員准教授)

- Europeanaには欧州委員会の強い後押しとファンディングがある。他方で、DPLAは、民間の有志による取組ではあるが、これだけ広がっている原動力はどこにあるのか。

(東京大学大学院情報学環 時実高等客員研究員)

- ・アメリカの図書館は馬力もあり、リテラシーも高い。ネットワークを使える司書も大勢いる。各州のアーカイブは10年またはそれ以上の歴史がある。それを統合してきたということは言える。

(高野主査)

- ・すでにあった活動をこの旗の下に束ね、それが参加している館の担当者にとって、立派な仕事としてきちんと評価される仕組みがあれば回ると思われる。

4. 「メタデータのオープン化等に関するガイドライン (素案)」のスコープについて

○事務局より、資料4に基づき説明。

○質疑の内容は、以下の通り。

(高野座長)

- ・協議会での議論において、国立美術館の方から意見があり、それは「紙の時代は、我々はしっかり目録を作った。そこには写真も解説文も入っていたし、今ではメタデータと呼ばれるものも一部含まれていた。ところが、世の中がデジタルに変わり、デジタルアーカイブになった時に、メタデータは目録のほんの一部の情報で、目録に載っていた写真も解説文も出さないということになっていき、本当の意味で流通し活用する場面で昔のキュレーターが蓄積していたものが、そのクオリティでは蓄積していかなくなっているのは本末転倒である。」というものであった。したがって、この会合ではメタデータとして記録するのはここまでで、あとは切り捨てるということではない立場で、国立美術館の危惧に答えることもやるべきではないかと考えている。

(東京大学文書館 森本准教授)

- ・配付資料(資料2-1「『知的財産推進計画2016』(アーカイブ関連)について」)内の「日本コンテンツの主なアーカイブの現状」の中で、国立公文書館が「参考」とされていることに驚いている。アーカイブのデジタル化について議論するにあたって、国立公文書館のデジタルアーカイブ(<https://www.digital.archives.go.jp/>)は大いに参照すべき対象だと考えているが、「参考」扱いだとすると、ここで考えるべきデジタルアーカイブというものが、どのようなものなのかと悩んでしまう。

(高野座長)

- ・公文書館については、公文書管理法があって、その範囲内でフィジカルなアーカイブを作るし、それを一部保全のためにデジタルアーカイブを進めたりしていくものであり、それは利活用というよりは国のインフラである。したがって、協議会の意図としては、利活用を進めようとする各実務者の議論と混ぜてしまうと混乱してし

まうことから、シリアスアーカイブとは分けて議論しようというものであろう。但し、公文書における課題があれば、他のアーカイブと同様に利活用を開いていくことも可能であろう。

(東京大学文書館 森本准教授)

- Europeana の話があったが、ヨーロッパでの図書館や博物館などそれぞれのアーカイブスごとに関わり方は違っているはずだが、デジタル化して様々な情報を共有していこうとしているのだと思う。
- 日本でも、図書館なら図書館の情報の活かし方、資料の管理の仕方があり、博物館・美術館にもそれぞれの資料の活かし方、情報の取り方がある。それらと同様に、公文書館もそうしたものをきつと持っている。それらを横断的につなげていくところに、デジタルアーカイブのメリットがあるのかなと思っている。
- しかし、博物館や美術館がデジタル化を進めるインセンティブ、「このためにデジタル化するのだ」という目的などは、公文書館とは違うということなのだろうか。そうだとすると、デジタル化を進めていくべきだというニーズは、どこから生まれてくるのか。

(高野座長)

- 公文書館を外しているというわけではないが、公文書館はアーカイブを本業とする意味で、重たすぎるといえるのは確かにあるのであろう。オープンガバメントと公文書館の活動は当然密接に結びつくと思うが、そのオープンガバメントを知財計画で推進していく議論は、我々の力に余る。したがって、文化的な情報を中心に議論していこうという形で会議は設置されている。但し、公文書館が行っている特別展などは、美術館の展覧会とほとんど変わらない。そういうものがシームレスに繋がっていくべきという意図はある。

(京都府立図書館企画総務部企画調整課 福島副主査)

- アジア歴史資料センターが出来たとき、当時大学院生として感動したのは、引用のルールを明示していたこと。ここの画像を論文に引用するときはこう書けというのが示されていた。国立国会図書館のやっている永続的識別子に近いものであった。デジタル的な流通とともに、研究の世界で論文を書くときに引用できるデジタルアーカイブのルール決めは必要。
- もう一つ衝撃的だったのは、文書の最初の 300 文字を入れておき、引っかかるテキストの量をとにかく増やしたこと。そのおかげで、普通にとられた目録とレベルが全く違う量の文書が検索に引っかかるようになった。役所の方はよくご存じと思うが、公文書は最初に大事なことが書いてある。多少間違っていてもいいからそれだけは文字を起こしておくことで、検索に引っかかる。そういった形での単純なテキスト化によって、担当者や学芸員の研究者としての業績が傷ついたりするようなことはない。そういった知恵をもう一度リバイバルしても良いと思う。

- ・インセンティブについては、デジタルアーカイブ化が各施設の評価基準にあがらない限り誰も積極的にはやらない。国立国会図書館が手間をかけてデジタルアーカイブの利用の頻度を評価基準に入れているように、来館者数と入館者数だけの評価ではない評価基準を国の施設に対して強く求めるぐらいでないと進まない。

(高野座長)

- ・むしろそれで評価したものをガイドラインの附録につけるか。文化庁は多くの費用をかけて文化財の修復を行っているが、修復前、分解時、修復後の写真を付して報告する。普通であればなかなか写真に撮れないようなものが報告書に上がってきている。しかし、それが博物館の展示に出てくるかといえば、それは全く別の話。それを制度的にうまく繋げば、修復の費用が実はメタデータやサムネイル等共有すべきアーカイブを豊かにしていくことにきっと繋がる。そういうこともガイドラインに書き込んで、問題がなければ実行できたらと思う。

(東京大学 生員客員准教授)

- ・我が国におけるデジタルアーカイブの取組を促進していくためには、インセンティブを設ける必要がある。しかし、これは法律を作って解決していく問題ではない。これから作るガイドラインでは、時実先生が資料4の45ページで説明された内容をどう実現していくかというストラテジーを示すことになるのではないかと。
- ・しかし、分野ごとに抱えている問題は異なる。例えば博物館や美術館では、図書館と比べて資料の特性上、メタデータを作ることが難しいということがある。このような問題を把握した上で、働きかけをしていかなければならない。
- ・どういったことをすればちゃんとした連携とデジタル化基盤が見えてくるのか。研究業績の反映をある程度統一的にできるようになり、税金の活用がなされ、継続的なアーカイブの運用が図られるようになるには、どのようなインセンティブを与えるべきかが論点になっていくと思う。

(ゼノン・リミテッド・パートナーズ 神崎代表)

- ・メタデータの話に踏み込んでしまうと領域ごとに抱えている問題が異なる。作者や作成日は図書館では問題にならないが博物館ではそれが簡単にはできない。さらに、文書館には、作者が存在しないような文書まで存在する。どこまで深入りした話をするのか。

(京都府立図書館企画総務部企画調整課 福島副主査)

- ・メタデータは、現場の作成レベルでの標準化は無理なので、あとは、システム側でどうマッピングするかという話だと考える。

(高野座長)

- ・世の中ではこのように克服しているとか、そういう議論に留めざるを得ない。具体

的にこの会合としてマッピングを決めるとかいう話ではない。メタデータが取りにくいものについては、生のテキストデータのトランスクリプションを一部メタデータの代わりにすることで、メタデータが意味的なものを伴っていない部分を補完する工夫があるということを書くことに留めるしかないのではないか。それよりも、工夫等を集めて列挙していく方が参考になると思う。

(国立国会図書館)

- それぞれの分野内の標準化にまで踏み込んでいくと、ガイドラインは作れないと思う。ただし、分野を横断するアーカイブ連携のために必要な要素・項目の標準化は必要であり、それをルール化する必要があるという内容が今年度のガイドラインになるのではないかと想定していた。ルール化の方法まで決めて書ければ理想的ではある。

(高野座長)

- 「ガイドライン」という言い方がミスリーディングとなるかもしれない。
- 具体的な課題のリストアップやその解決のアイデア、解決した具体策をグッドプラクティスの例として掲載し、課題解決のためのガイドラインとするのがよいのではないか。

(京都府立図書館企画総務部企画調整課 福島副主査)

- 資料 2-1 の日本のコンテンツのアーカイブ一覧について。資料にある各種のデジタルアーカイブから検討されるものと理解しているが、もし書けるのならば、地方のアグリゲーターについて考えたい。文化資源と呼んでいる、各地域の歴史資料・地域遺産を対象としたデジタルアーカイブが大量にある。各文化施設なり在野の団体なりが大量にデジタル化しているが、どこにもつながっておらず、しばらくしたら消滅して追跡もできないという現状がある。これをどうやって拾ってくるか、地域の資料館や博物館、図書館が直面しているのがこの課題。そういうことを念頭においた議論ができればいいし、実態としてそこを何とかしなければというのが本音のところである。

(高野座長)

- 取り入れればよいと思う。

(東京大学 生貝客員准教授)

- ガイドラインにしてもルールにしてもレベル感が必要である。個人的には、時実先生が資料 4 の 45 ページで示されたような 4 項目くらいの課題はクリアするようなガイドラインができればいいと思っている。
- ナショナルアーカイブがさらに評価されるために、IIIF や高画質なコンテンツがリユースできる、多言語対応といったことができればよいのではないか。

- Europeana にせよ、DPLA にせよ、CCO のメタデータ提供のルールがエンフォースされるトリガーとなるのは、統合ポータルへの参加である。

(高野座長)

- 参加館は、現実的な予算で手元の資料を電子化して活用したいという要望があって、そのための予算が Europeana に参画することで付いてくるということで動いている。したがって、予算が充実している館は Europeana に参画しないということになる。
- 本日は第 1 回目だが、このくらい幅を持った議論をあと 3 回重ねて、事務局と調整しながらまとめていきたい。

5. その他

○次回の実務者協議会は、10 月 11 日（火）15 時 30 分～17 時 30 分である。お願いする発表内容は、追って調整の上、連絡する。

以上