

メタデータのオープン化等検討ワーキンググループ(第2回) 議事概要

日時：平成28年10月11日(火) 15:30～17:35

場所：中央合同庁舎4号館 共用1202会議室

【議事】

1. 文書館におけるデジタルアーカイブの現状と課題
2. デジタルアーカイブの利活用のためのメタデータのあり方
3. ガイドライン(素案)の構成

【概要】

1. 文書館におけるデジタルアーカイブの現状と課題

○森本構成員より、資料1に基づき説明。

○質疑の内容は、以下の通り。

(国立歴史民俗博物館 後藤准教授)

- ・APExについて、検索結果から組織へリンクする際に、リンク先はどのように表示されるか。データベースの検索結果がそのまま出てくるという理解か。

(高野主査)

- ・間に集約したAPExの検索結果があって、それをクリックすると、APExから来たクエリを自館のほうで受け止めるページがこのように用意されているということであろう。

(国立歴史民俗博物館 後藤准教授)

- ・Europeanaからの検索はどうなっているのか。

(高野主査)

- ・Europeanaもおそらく同じ考え方である。頭出しはEuropeanaで集約し、そこでヒットしたらプレビューが表示される。その詳細情報の確認には、データを元々持っているサイトに遷移するという形の連携である。個々のサイトがどこまでEuropeana側にプレビュー情報を出しているのか、そこから来たクエリに対してどこまで自館のコンテキストで受け止めるのかというのは個々のサイトのポリシーだと思う。

(東京大学 生貝客員准教授)

- ・文書館は、資料ごとにパーマネントなURLがあって、コンテキスト情報がそれぞれに付与されている。図書館や美術館はこのような情報を意図的に出さないことがあ

る。連携していくためには、文書館のようにコンテキスト情報を付与していかなければならないのではないかと。

- ・コンテキストを与えるときに組織単位で記述することが多いと思うが、ポータルで複数の組織を横断的に見る場合に、個々の組織の外にあるものを取り込むような取組はあるのか。

(東京大学文書館 森本准教授)

- ・国内ではまだ全然できていない。デジタル化以前の問題だが、標題・作成年月日・作成者を記述するというような国際的な標準に則って目録を作るようにはなっていない。標準化は進んでいるので、デジタル的フォーマットを統一すれば、つながっていくと理解している。

(高野主査)

- ・ある史料が所属しているパス、コンテキスト、ディレクトリなど、それらを全部集めてくるとその史料のメタデータになる。当然パスがないと史料がぶら下げられないことから、必須メタデータのフィールドを構成しているともいえる。ある意味では、すべての史料には、そのようなアーカイブ毎に決められた必須項目を持つメタデータスキームが付随していると考えられなくもない。その必須項目は統一されていて、それが一個でも抜けている史料はどこにも取まりようがないというように、アーカイブの世界はかなり厳密に考えられ、作られてきたと感じている。

2. デジタルアーカイブの利活用のためのメタデータのあり方

○神崎構成員より、資料2に基づき説明。

○質疑の内容は、以下の通り。

(東京大学文書館 森本准教授)

- ・東京大学文書館で構築中のデジタルアーカイブでも、リンクドデータを採っている。アーカイブズでは「まだうちに来ていない資料」の情報もつくることができる。たとえば、本日の議事録が近日にできると思うが、こうした会議が開かれたという情報を国立公文書館がキャッチすると、その時点でメタデータを記述することができる。そうすることで、実際に文書が到着したら、即座に公開することができる。デジタルアーカイブの世界も、今後このように動いていくと考えられる。

(高野主査)

- ・Europeana がホストしている、元々アグリゲーターが付与してきたものとは全く別に、もう一度整理しなおしたり追加したりすることは、アーカイブを集めて、別の軸で整理しなおし提供することに近いことかと思う。

(ゼノン・リミテッド・パートナーズ 神崎代表)

- ・Europeana はプロバイダのものには関与せず、プロバイダが持つプロシキのメタデ

ータ項目は、Europeana が独自に付与したメタデータとは分けて考えるという思想で動いている。

(国立歴史民俗博物館 後藤准教授)

- ・博物館で難しいのは、想定外のものがコレクションに入ってくる。そこをどう記述していくべきか、今後検討できればと思う。

(京都府立図書館企画総務部企画調整課 福島副主査)

- ・いままでの資料を上手に使うやり方もできる。「足利義満の関係ドキュメント」というリンクができれば、従来からある文書のデータのうえに、新しい資料が出てきたときに、そこに加えることができる。

(国立歴史民俗博物館 後藤准教授)

- ・資料の認識が新たになったところでデータを追加できるのが、リンクトデータの強みである。

(東京大学文書館 森本准教授)

- ・みんなが好きに URI を設定できるという話だったが、たとえば、「ゴジラ」とはこういうものである、というコントロールされた典拠、オーソリティをつかって、そこをみんなが書き加えていく、という発想はないのか。そうではなく、千でも二千でもみんなが「ゴジラ」というデータをつかって、それらをつなげるほうがよいという考え方か。

(ゼノン・リミテッド・パートナーズ 神崎代表)

- ・既に利用可能な URI があれば、それを使った方がよい。ペストプラクティスとして例えば DBpedia に URI があるならそれを使った方が、ハブを経由するというステップが省略されるので望ましい形ではないかと思われる。だが、そこになれば諦める、というものでもない。

(高野主査)

- ・小さなオーソリティが国の活動等によりマージされてより大きなオーソリティが出てくるといったように、社会的活動により、最終的にできるだけ数が少なく、そしてより多くの人信用できるオーソリティが確立していく、そういうものをこの国の文化のために作っていくにはどうしたらいいかというのが、この会議の大変大きな目的でもある。
- ・資料の 10、11 (6 頁) の箇所は非常に整理されていて、ガイドラインに是非盛り込むべき内容だと思う。何のためにどんな情報が必要かというのが、そもそもメタデータを考えるときに必要である。例えば、FRBR (Functional Requirements for Bibliographic Records : 書誌レコードの機能要件) で分析された 4 タスクである

が、おそらくこれはそのまま生き残っている。もう一度、本よりもさらに大きな世界で、この4タスクを提示しながら、我々の行うべき活動、或いは、世の中で行われている活動を評価しなおして、我が国の知財戦略の中で取り組むべき内容を決めていくことが重要であると思う。

3. ガイドライン（素案）の構成

○国立国会図書館より、資料3-1に基づき説明。それに引き続き、後藤構成員より、資料3-2に基づき説明及び福島構成員より、資料3-3に基づき説明。

○質疑の内容は、以下の通り。

（国立歴史民俗博物館 後藤准教授）

- ・評価基準について、どこまで使われるか考えると、書き方が悩ましい。

（高野主査）

- ・評価基準がどこまで使われるかはわからないが、海外事例などと比較する場合、海外では所蔵品が外から見えるようにオープンにしている美術館等があり、その結果、例えば、入館者数の増減やそれを活用した教育実践が行われているかどうか、といった点をしっかり評価して、デジタルで開くことがもたらすメリットが実際に起きていれば、それを定量的に示したい。それによって、クローズにしている館には少しプレッシャーをかけたり、或いは、市町村レベルの自治体でもアイデアを活用してみようということになったりしないとも限らない。ここで決めたものが国の組織が拘束されるほどの影響力を持っているとは思っていない。

（国立歴史民俗博物館 後藤准教授）

- ・論文などの学術資料のオープンアクセスはポイントになりやすいが、所蔵資料のオープンアクセスはそうでもないといった実状を踏まえると、他の実務的な部分はともかく、評価基準の扱いは難しい気がする。

（ゼノン・リミテッド・パートナーズ 神崎代表）

- ・できるだけ広くデータが集まるようにすべきである。
- ・例えば、エクセルでまとめられた表データも対象とすべきである。
- ・そのために、ガイドラインには、データの活用が難しくなるような、セル内での改行はそもそもしてはならないといった、共有可能な表データの基本的な作り方に関する記述を入れてもらいたい。

（東京大学 生貝客員准教授）

- ・全体として、デジタルアーカイブ化を進展させるように背中を押すようなガイドラインにしなければならない。
- ・オープン化を進めていくにあたって、担当者がこのガイドラインを参照すれば一人で判断を行えるような書きぶりが望ましい。ベストエフォートを見込んだ書き振り

にすることが重要である。

- 例えば、権利制限に関する記述では、正確に記載をするとハードルが高くなるが、分かりやすく、間違いが起きないようにこういうケースであればこうあるべきといった場合分けで事例集を提示できるとよい。

(高野主査)

- かつて総務省で「震災関連デジタルアーカイブ構築・運用のためのガイドライン」をまとめる会の座長も務めたが、このときは、データに触ったこともないような人が必死に何かを残そうと現場に赴いて撮ったものを、何とか死蔵にならないような形で国立国会図書館の東日本大震災アーカイブ等に収蔵していけないかといった、現場の思いに主体を置いたレポートをつくった。
- 今回は、そうではなく、すでにアーカイブに取り組んでいるが、公開されたコンテンツについて、国民に利便を提供しないといけない組織が、海外と比べると遅々として進んでいない。そういうことにハッパをかけることが、このワーキングにおいてプライオリティが高いものである。
- これは、組織のセキュリティポリシーとか、組織がコンテンツに誇りに思っている無謬性が利活用の面で妨げになっている面がある。公開することにより間違いが発覚すると困るから公開しない、というような自己矛盾のようなことが組織を委縮させている。そんなことを国民は求めている。組織が正しいと思っているのだからある意味では正しい。しかし、他の人がもっと正しいことに気づいたときに直すことができればそれでよいのではないか。
- 世界の端っこで北斎といったときに、日本の北斎というオーソリティにちゃんと辿り着いたり、一番豊かな北斎の情報に繋がっていく入口になったり、そういう世界にとっての、少なくとも日本の文化についてのオーソリティというものをこの国はたてないといけない。そういう議論の出発点にできたらよい。そういう思いを持っていても、組織ではなかなか戦いきれず公開に至らないような人たちの背中を押してあげたいと思う。せつかく内閣府の下で公式な報告が出るのであれば、公的組織がなんとなく勘違いで危ないかなと思っていることを実は危なくないですよと、そちらにこそ進んでいくべきですよと説得的に書かれるようなガイドラインにできるとよい。

(国立歴史民俗博物館 後藤准教授)

- 国文学研究資料館が画像のオープンデータセットの公開範囲を大きく広げたが、画像はCC BY-SA としている一方、目録についてはそう書いていない。古典籍総合目録だけは絶対に自分のものとしてオープンにしていない。本来はそういうところこそオープンにすべきで、意識改革の必要性を感じる。

(東京大学文書館 森本准教授)

- スリムモデルとノーマルモデルという整理はとてもよいと思う。スリムモデルでよ

いから、どんどん情報を出していこうと後押しができる。このとき、これからあらたに撮影する資料と、既存のものは切り分けて考えるべきだろう。これから撮影していくものについては、これがベストの形だけれど、他方で、昔撮った白黒のマイクロフィルムなどは、それでもいいから出していくように促したい。この両方を発信していくべき。

(京都府立図書館企画総務部企画調整課 福島副主査)

- ・おっしゃるとおりだ。そこは書き方だろう。これから新たに出すものの最低ラインを定めながら、それらを各機関が積み上げてきて評価が定まっている既存のものと組み合わせていくということだろう。

(高野主査)

- ・ゴジラのスチール写真を例に挙げると、手元に写真がなく、本で見たことがあるというような体験だけでもその人は十分語る事ができて、そういう人同士が誤りなく同じものを前提に語る事ができる。それについて、将来的に本当の写真、精度の高い写真が話を深めていくというようなことが起きてもいい。我々が知識を貯めていくための枠組みのようなものについて、手元にデジタル物があるかないかにあまりとらわれずにきちっとそういう作業を進めていけるとするのは非常に重要なことである。

(国立歴史民俗博物館 後藤准教授)

- ・オープンデータの技術とオープンアクセス的な技術は、切り分けて整理した方がよい。例えば、IIIFは画像の相互運用にこそ意味があり、各機関の画像をコピーして集約しなくても、一つのビューアで見られるところに特徴がある。オープンデータ前提の技術ではない。IIIFの技法とオープンデータでコピーして自分のところで表示させようというような技法とを混ぜたガイドラインのつくりにしてしまうと、全体像がよく分からなくなるおそれがある。

5. その他

- 次回の実務者協議会は、11月14日(月)15時から17時まで、次々回の実務者協議会については、12月12日(月)15時から17時までの予定で開催する。

以上