これからの学術情報システムのメタデータ 収集・作成方針について (2021) 【案】

これからの学術情報システム構築検討委員会

システムワークフロー検討作業部会

(文書全体の目次)

1. はじめに	4
2. 「これからの学術情報システムの在り方について (2019)」とメタデータ	6
3. 「物理的な資料」のメタデータ	11
4. 「電子的な資料」のメタデータ	13
5. 「デジタルな資料」のメタデータ	17
5.1. 現況	18
5.1.1. 大学/大学図書館のデジタルアーカイブの現況	18
5.1.2. 国立国会図書館 (ジャパンサーチ・NDLサーチ) の現況	18
5.1.3. NII の現況	19
5.2. 見解と方針	19
5.2.1. つなぎ役としてのエコシステム構築 (IRDB 経由)	19
5.2.2. つなぎ役としてのエコシステム構築 (IRDB 以外)	20
5.2.3. 意見交換が可能なフォーラムと支援体制	20
5.2.4. 各機関での取り組み	21
6. 共同利用システムへの「見解と方針」反映の実際	21
7. 参考文献	25

5. 「デジタルな資料」のメタデータ

ここでは出版者以外の手によって電子化された「デジタルな資料」の中でも、とくに「デジタルアーカイブ」と呼ばれるもののメタデータ流通をどのように強化するか、デジタルアーカイブを巡るエコシステムモデルの構築を提案する。

5.1. 現況

「デジタルアーカイブ」の定義はいまだ確定的でない。しかし「利用規約の明示」「機械可読性の担保」「環境に依存しないデータ移行性の担保」「アクセシビリティの確保」「真正性の確保」「永続性の確保」を要素とし、社会が遺すことを選択した／すべき知識情報基盤としてのデジタルデータとそれにまつわる仕組みの総体として、万人へのアクセスが保障されていれば、コンテンツの内容を問わず、デジタルアーカイブとみなすのが妥当である。

現状、各機関のデジタルアーカイブは散在しており、とくに国内の大学等各機関にとってのつなぎ役・まとめ役が不在であるため、デジタルアーカイブのメタデータが効果的に集約・流通できていないことが最大の課題である。また各機関でのメタデータの枠組みやマネジメントもまちまちで定まることがなく、あるいはサポートが不在でノウハウが共有されていないという課題も指摘できる。

とは言え、本文書末の参考文献をはじめ、背景となる報告書やガイドラインによって、事業構築の基礎はすでに整備された。また2020年8月に正式リリースされた「ジャパンサーチ」はデジタルアーカイブの1つの集約点として着目される。ジャパンサーチでは、この集約の方法として、「分野・地域ごとの「つなぎ役」を介して、「アーカイブ機関」と「ジャパンサーチ」の間でメタデータの共有を進め、「活ユーザー」がジャパンサーチからメタデータを共有し、さまざまな用途に活用するというサイクルの構築を目指す方向性」を示している。また、「つなぎ役」とは分野・地域等のコミュニティ単位でメタデータを集約し提供等を行う機関を指し、「メタデータの標準化、用語の統制（辞書・典拠・シソーラスの管理）を行い、コミュニティにおけるメタデータ整備やデジタルコンテンツ作成への支援、長期アクセス保証のための基盤提供、活用コミュニティの形成などの役割を担う」とされている。

図書館・書籍等分野においてのつなぎ役はNDLが担っており、今後国内各種図書館の網羅的連携が期待される。

5.1.1. 大学/大学図書館のデジタルアーカイブの現況

デジタルアーカイブは多様な主体が関わる行為であり、大学図書館が大学アーカイブズや地域/分野のアーカイブズ等と連携する例も少なくない。対象コンテンツはかなり柔軟でよく、機関の経営に関わる組織文書や構成員に関わるビッグデータなども対象となる。また書籍以外の多様な形式の学術資源・文化資源全般（文書、写真・画像、音声・動画等）や、著作の部分、コレクション等の多様な単位を扱うため、そのメタデータには電子ブック・電子ジャーナルとは異なるものが多い。機関リポジトリは多くの大学で整備されているが、予算・管理面等の事情から、機関リポジトリでデジタルアーカイブを構築する機関とそうでない機関が混在する。

これらの多様性を包括し失われたり取りこぼしたりすることのないようアグリゲート・流通させ、最終的に学内外での運営側・活用側の境界を溶解させるべく、大学図書館が主体となって活動する必要がある。

5.1.2. 国立国会図書館（ジャパンサーチ・NDLサーチ）の現況

今後デジタルアーカイブの設計においてジャパンサーチを「所与の条件」として検討することで、多くのメリットが想定される。ジャパンサーチに対応することで、他のシステムへも一定の互換性が担保される。提供機関自身のデータベース・ウェブサイト構築によらず、サーバにコンテンツとメタデータを格納することで一定の持続可能性が担保できる（例：南方熊楠記念館）。ウェブスケールディスカバリーがなくても学内の多様な情報資源の統合検索が容易になる、等である。

現在、NDLではDC-NDLをデジタルアーカイブに適したものに再設計が検討されており、今年度中に「デジタルアーカイブメタデータガイドライン」が整備される予定である。ガイドラインではDC-NDLとJPCOARスキーマとに互換性が持たれ、これによりIRDB経由でのメタデータ交換も可能となる。国内の大学図書館システム等で実装が期待されるJPCOARスキーマで各大学図書館等のデジタルアーカイブをNDLサーチに直接登録することが可能となり、メタデータ交換が飛躍的に進むと考えられる。

なお、NDLは各機関との直接のメタデータ交換にも積極的に取り組んでいる（例：東京大学学術資産等アーカイブズポータル、琉球大学琉球・沖縄関係貴重資料デジタルアーカイブ、立命館大学アトリサーチセンター、奈良県立図書情報館等）。またNDLサーチからジャパンサーチへのメタデータの収録については、現状、網羅的ではなく選択的である。

5.1.3. NIIの現況

JPCOARでは現在デジタルアーカイブ対応に伴うJPCOARスキーマの再定義が検討されている。JPCOARスキーマ準拠のメタデータはIRDB経由で流通させることが可能であることから、機関リポジトリに登録されている/いないに関わらず、デジタルアーカイブのメタデータがIRDBに登録できればその流通が進むと考えられる。

また古典籍等のデジタルアーカイブのデータをNCIDとともにERDB-JPに入れることで、CiNii BooksやCiNii Research上での表示も可能である。

5.2. 見解と方針

現況をふまえて、本文書では以下の在り方を目指すことを提案する。

- (1) Web上での検索ポータルとして総合的・代表的であり、国内外の多くのユーザの目に触れると評価できるサービスを、流通のための外部連携先として想定する
- (2) 具体的に想定する外部連携先として、NDLサーチおよびジャパンサーチを柱とする。加えて、CiNii Research、各種ウェブスケールディスカバリー、Google等の一般的な検索エンジンを射程に置く
- (3) ジャパンサーチに対するつなぎ役としての体制の整備を含め、そのための各機関のメタデータの集約と支援の仕組みを整備する

加えて、これらの目指すべき在り方の実現に向けて、各機関の現状・動向等をふまえ、以下のような方策を提案する

5.2.1. つなぎ役としてのエコシステム構築（IRDB経由）

機関リポジトリでデジタルアーカイブを構築する機関を念頭に置くと、IRDBを介してNDLサーチへとつなぎ、そこからジャパンサーチへとという流れを確固たるものにする必要がある。

① 各機関から IRDB への流通

NII の現況で述べたように、機関リポジトリでデジタルアーカイブを構築している場合、(再定義された) JPCOAR スキーマを介して IRDB へとメタデータが集約されることが期待される。一方で、機関リポジトリでないシステムをベースにデジタルアーカイブを構築している場合には JPCOAR スキーマでの流通が困難な場合がある。

例えば Alma はデジタルアーカイブを構築する機能を有しているが、その場合のメタデータは MARC21 や DCMI メタデータ語彙等となる。従って、このような機関リポジトリ以外のデジタルアーカイブの場合、そこで採用されているメタデータスキーマを JPCOAR スキーマに変換し、収集するゲートウェイを用意する必要がある。今後、そのための検討を制度面と技術面から行うことが望ましい。そのほか、IIIF マニフェストからの収集、Excel/CSV の登録等、異なるルートや異なるスキーマからのメタデータを IRDB で集約するための対応も求められる。

② IRDB から NDL サーチへの流通 (JPCOAR スキーマから DC-NDL への変換)

NDL にて検討中のガイドラインによる JPCOAR スキーマ・DC-NDL 間データ交換の実現に取り組むことを検討する。すなわち資源タイプ・分野・提供機関名等が流通先でどう検索・表示されるかや、資源とメタデータの多様性を考慮した標準化の要不要の切り分け等について、考慮する。

また、IRDB と JPCOAR スキーマについて、デジタルアーカイブに適した対応を進めることも必要である。具体的には、文書、古典籍資料、非書籍資料 (図像・地図・音源、著作の一部分等) で必要とされる、メタデータ項目 (書誌・形態等の特記事項、コレクション、ライセンス、サムネイル、IIIF マニフェスト URI 等) への対応等が求められる。

③ IRDB から NDL サーチ以外への流通

各種のウェブスケールディスカバリーや Google のような検索エンジンへの流通方法を検討する必要がある。ウェブスケールディスカバリーについては、各ベンダーのナレッジベースに登録されるように働きかけを行ったうえで、OAI-PMHなどを介して自動でデータ更新を行えるようなシステムとしなければならない。Google については、一般の検索エンジンで適切な検索結果が得られるよう、Schema.org などの活用なども射程に入れる必要がある。そのほか、Google Arts & Culture との連携も検討する価値がある。

5.2.2. つなぎ役としてのエコシステム構築 (IRDB 以外)

各機関で保持できるインフラや扱うアイテムの種類が異なるため、機関リポジトリの使用や IRDB へのメタデータ登録ができない場合を考慮し、それらのメタデータを集約できるよう窓口を整備することも必要となる。また IRDB において、多様なメタデータスキーマでの収集が困難であった場合や、NDL サーチ以外への流通対応などが厳しいといった場合には、共同利用システムの一部として新たなデータ連携用のサービスを構築することも考慮する必要がある。

5.2.3. 意見交換が可能なフォーラムと支援体制

つなぎ役の運営主体を中心とした闊達な意見交換が可能なフォーラムを構築し、各大学・研究機関がたどってきた歴史的経緯や各種の諸条件を前提としながら、最適解の相互提案とその

共有を行なう。また多様な主体が関わるというアーカイブの本質上、図書館のみならず各組織での教員や学生、院生、事務方なども巻き込み、学内での運営側・活用側の境界を溶解させる方向になっていくことが望ましい。

意見交換において、とくに重要と考えられる要件には以下のようなものがある。

- ① ノウハウ、ベストプラクティスの意見交換と共有
- ② メタデータのマッピング、技術的要件、流通範囲等の情報
- ③ 機関リポジトリでのデジタルアーカイブ構築や、IRDB へのメタデータ登録についての支援
- ④ NDL サーチ・ジャパンサーチと直接連携を希望する機関への支援
- ⑤ 流通先でのメタデータ表示・品質に対するケア

意見交換・議論の際に準拠できる参考文献類を本文書末に示した。これらの共有化の努力を行い、最低限の要件を常に再確認しつ、改善の余地のあるものについては意見を出すという体制を構築したい。

5.2.4. 各機関での取り組み

① 学内における調整とメタデータ集約

大学図書館は、散在するメタデータを包括的にかつ効率的に集約・流通させるため、学内におけるデジタルアーカイブに関わる多様な主体のとの“サブ”つなぎ役としての役割を担うことが必要である。

② 各機関のメタデータに求められる要件

どのような方法をとる場合でも、外部の機関・サービスと広く連携すること、国内外の広範囲にスムーズに流通させること、発見可能性を高め活用を促進することを考慮されたい。

以下、とくに重要と思われる要件を示す。

- メタデータ・サムネイルのライセンスとして、CC0 を明示する。
- タイトル（ラベル）、一意で永続的な管理番号（識別子）、各アイテムの固定 URL
- 英語等の多言語・ローマ字表記、サムネイル画像、原資料の NCID、資源タイプ、分野、ライセンス、所蔵機関名

7. 参考文献

[5章]

【1】デジタルアーカイブジャパン推進委員会・実務者検討委員会。「3か年総括報告書：我が国が目指すデジタルアーカイブ社会の実現に向けて」。2020-08-19.

https://www.kantei.go.jp/jp/singi/titeki2/digitalarchive_suisiniinkai/pdf/r0208_3kanen_houoku_honbun.pdf, (参照 2021-10-05).

- 知的資産のシェアと利活用により新たな価値を創生する社会基盤としてのデジタルアーカイブジャパン（デジタルアーカイブ社会の実現）を推進
- 幅広い知識や理解を要する人材育成には、育成環境や財政基盤等の課題がある

【2】[デジタルアーカイブジャパン推進委員会・実務者検討委員会]. 「デジタルアーカイブの

ための長期保存ガイドライン（2020年版）」.[2020-08].

https://www.kantei.go.jp/jp/singi/titeki2/digitalarchive_suisiniinkai/pdf/guideline2020.pdf, (参照 2021-10-05).

- 媒体の議論に重点を置かない
- ドキュメンテーション，管理の仕組みの構築，組織的対応による真正性の確保
- 分散保存の強い推奨

【3】 [デジタルアーカイブジャパン推進委員会・実務者検討委員会]. 「デジタルアーカイブアセスメントツール（改定版）」.[2020-08].

https://www.kantei.go.jp/jp/singi/titeki2/digitalarchive_suisiniinkai/files/assessment_tool_kaitai.xlsx, (参照 2021-10-05).

- デジタルアーカイブの自己点検のためのツール
- 標準モデル（小規模な機関で行うことが推奨される水準）／先進モデル（各機関のミッション等の必要に応じて目指す水準）／つなぎ役モデル（分野・地域コミュニティのつなぎ役の役割をもつ機関が目指す水準）

【4】 国立大学図書館協会学術資料整備委員会デジタルアーカイブ WG. 「大学図書館におけるデジタルアーカイブの利活用に向けて」. 2019-06.

https://www.janul.jp/sites/default/files/2019-07/sr_dawg_report_201906.pdf, (参照 2021-10-05).

【5】 [デジタルアーカイブジャパン推進委員会・実務者検討委員会]. 「デジタルアーカイブにおける望ましい二次利用条件表示の在り方について（2019年版）」.[2019-03].

https://www.kantei.go.jp/jp/singi/titeki2/digitalarchive_suisiniinkai/jitumusya/2018/nijiriyou2019.pdf, (参照 2021-10-05).

- 国際的に普及しているパブリック・ドメイン・ツール及び CC ライセンス。とくに，CC0, CC BY を強く推奨する。
- Rights Statements からは，著作権あり，著作権あり－教育目的の利用可，著作権なし－他の法的制限あり，著作権なし－契約による制限あり，著作権未評価のマーク。
- 日本独自の表示としては，裁定制度により利用された著作物であることがわかるマーク（著作権未決定-裁定制度利用著作物）

【6】 デジタルアーカイブの連携に関する関係省庁等連絡会・実務者協議会. 「デジタルアーカイブの構築・共有・活用ガイドライン」. 2017-04.

http://www.kantei.go.jp/jp/singi/titeki2/digitalarchive_kyougikai/guideline.pdf, (参照 2021-10-05).

- メタデータの整備，長期アクセスの保証，望ましい利用条件，データ共有，データ活用，成果物の還元，コミュニティ形成を重視。
- 「活用できる表形式のデータとは？」を付録として添付。