

第五期国立国会図書館科学技術情報整備基本計画策定に向けての提言（案）の検討経過

1 基本方針検討部会設置の経緯

令和元年 9 月 26 日に開催された第 11 回科学技術情報整備審議会において、令和 2 年度で計画期間が終了する「第四期国立国会図書館科学技術情報整備基本計画」の後継となる次期計画を国立国会図書館が策定するための提言を出すことについて、西尾章治郎委員長から諮られ、その基本方針を検討し、素案を作成するものとして、基本方針検討部会を設置することが決定された。

2 基本方針検討部会の構成

部会長	竹内 比呂也委員	千葉大学副学長
部会員	佐藤 義則委員	東北学院大学文学部教授
	生貝 直人専門委員	東洋大学経済学部准教授
	北本 朝展専門委員	国立情報学研究所教授

3 基本方針検討部会における検討経過

回次	開催日時	主な議題
第 1 回	令和元年 11 月 14 日（木） 午前 10 時～正午	<ul style="list-style-type: none"> 基本方針検討部会の構成及び運営 基本方針検討部会における検討事項及びその進め方
第 2 回	令和元年 12 月 12 日（木） 午前 10 時～正午	<ul style="list-style-type: none"> データ駆動型社会からデータ神格化社会へ [北本部会員発表] 知識基盤社会におけるデータ駆動型研究の動向
第 3 回	令和 2 年 2 月 27 日（木） 午前 10 時～正午	<ul style="list-style-type: none"> デジタルアーカイブの利活用に係る EU 等における法整備面の動向 [生貝部会員発表] デジタルデータ・デジタルアーカイブの知識基盤社会における利活用 提言骨子案の検討
第 4 回	令和 2 年 5 月 28 日（木） 午前 10 時～正午	<ul style="list-style-type: none"> 識別子とメタデータ [佐藤部会員発表] デジタルデータ・デジタルアーカイブの知識基盤社会における利活用 データ利活用社会における基盤整備 提言概要案の検討
第 5 回	令和 2 年 6 月 25 日（木） 午前 10 時～正午	<ul style="list-style-type: none"> データ利活用社会における基盤整備 提言素案の検討

第 6 回	令和 2 年 10 月 22 日 (木) 午前 10 時～正午	・ 提言案の検討
-------	------------------------------------	----------

4 第 12 回科学技術情報整備審議会における審議

令和 2 年 9 月 30 日に開催された第 12 回科学技術情報整備審議会において、基本方針検討部会が取りまとめた「第五期国立国会図書館科学技術情報整備基本計画策定に向けての提言（素案）」が議題とされた。竹内比呂也部会長及び事務局からの説明の後、審議が行われた。

委員からは、国民を巻き込んだ利活用促進のための取組の重要性、新型コロナウイルス感染症の影響を前提とした優先的な課題への対応の必要性、デジタル化・テキスト化等の取組の着実な実施等の指摘があった。

委員からの指摘を踏まえ、基本方針検討部会において提言案を検討することとなった。

(参考) 各回部会における主な意見

【第 1 回部会】

- ・ 第三期で掲げた「知識インフラ」、これに続く現行の第四期の「深化型知識インフラ」では、恒久的保存と利活用促進をつなぐ役割を国立国会図書館（以下「NDL」という。）が担うとしている。AI 等の技術発展も踏まえて、これまでの方向性をどう活かすか、変えていくかを考える必要がある。
- ・ 「データ利活用社会」という言葉を越えたドラスティックな変化が生じている。次期計画期間中の 5 年間のうちに、情報や知識に関するサービスは、「デジタルファースト」が当然に求められる状況になっているのではないか。その中で、知識情報基盤としての NDL の役割や、広く図書館と社会との関係を考える必要がある。また、その先に見えてくるバリューを感じられる言葉が必要である。
- ・ 「データ」の定義はかなり広く、データサイエンスが扱う「データ」だけではない。NDL が関与するのは、主に文献の画像データやテキストデータを想定しているだろうが、「メタデータ」も必然的に含まれるだろう。データとは何かを最初にきちんと明確にしておく必要がある。
- ・ 著作権法における NDL の特別な位置付けを活かしていく一方で、新たな知識情報基盤を構築していく上で現行法では不十分な部分があれば積極的に提示すべきである。
- ・ 日本全体の知識をデータとして収集し、扱えるようにするため、ジャパンサーチは、中長期的に、総合的、計画的に推進する必要がある。NDL はジャパンサーチのハブであり、次期計画でも取り上げる必要がある。
- ・ デジタル化されていないものについて、考え方を示す必要がある。何百万件の単位で大規模にデジタル化し、全文検索できるようにすることは、民間にはなかなか難しい。NDL しかできないデジタル化と、それ以外との区分けも考える必要がある。
- ・ オンライン資料、とりわけ DRM（技術的制限手段）付きのものは、現時点では収集が不十分である。諸外国の状況も調査し、対応を考えなければならない。

【第 2 回部会】

- ・ データ偏重からデータの価値判断や倫理など人間側の問題に焦点が当たる中で、データ駆動型人文学を進めるだけでなく、AI の倫理の問題など人文学の知見を他の領域に広げることも求められている。これら二つの方向性における基盤として、NDL が過去の蓄積をいかしてどのような役割を果たせるのかは検討すべき課題。
- ・ 「デジタル化」という言葉が指す内容をはっきりさせた方がよい。画像化だけにとどまらず、テキスト化、構造化が当然に求められるのではないか。学術雑誌では、全文 XML 化も広がりつつある。
- ・ テキスト化の研究やその実現に当たっては費用が問題となる。全ての資料をデジタル化、テキスト化することが望ましいが、分野や資料群等による優先順位付けも必要になるのではないか。そのための検討には、デジタル化やメタデータの整備状況や方向性を明らかにしておく必要がある。今回の提言で方針を示し、NDL で具体化していくと考えている。
- ・ NDL の資料のデジタル化は、テキスト化したものを検索して必要な知識を得ることを考えると、ここ 2、30 年間の新しいものが重要となることを考慮すべきである。既にあるテキストデータも含め、他機関との連携・分担を図りつつ、どの時代のものまでデジタル化が必要かと併せて、利用の仕組みを考えないといけない。

- ・ HathiTrust も OCR 技術自体は研究開発しておらず、Google に頼っている。他方で、日本語資料のテキスト化の精度は十分ではなく、日本が主体的に取り組むべき課題でもある。画像データからのテキスト化を精緻化するためには、大規模かつ長期の取組が必要であり、NDL 単独では難しい。オールジャパンの体制を考えないといけない。
- ・ 研究データについては、NII や JST の取組方針を尊重し、NDL は、両機関でカバーできないもの、例えば、引き取り手のいないデータの救済、地域資料など特定の研究分野に結びつかない横断的なデータに目を向けるとよい。ただし、NDL が全てを預ければよいものでもない。「ひなぎく」のように基盤となるネットワークをどう構築するかが重要となる。
- ・ 現代の倫理観からは到底認められない内容を含む資料、事件報道や個人の住所を含む資料など、デジタル化に当たって問題となり得るものもある。単に検索結果から除外すればよいのか、またオプトアウトの手続きなど、対応を議論する必要がある。
- ・ データ駆動型社会において、NDL の役割は広く人々をエンパワーメントすることにある。データの利用者を、一般市民、専門家・学生、分野専門家に分類すると、「専門家・学生」は自らツールを活用する意欲がある人たちである。これらにはデータを利活用できるツールを提供することが考えられる。「一般市民」は、例えば、気象観測データならば、天気予報など、データそのものではなく、直接的な情報を得ることを望む人たちである。例えば、データリテラシーの涵養など、一般市民を対象とした取組も何か考えられるのではないか。
- ・ シチズンサイエンスとの関わりについては、NDL が積極的に何かを支援するというものではなく、広いネットワークの中で考えればよい。

【第 3 回部会】

- ・ 日本における絶版等にあたる「out-of-commerce」資料について、欧州では、デジタル単一市場における著作権指令や権利者団体とのガイドラインの策定など、公衆に利用可能とするための取組が進んでいる。他方、日本では、著作権法第 31 条第 3 項を根拠とする NDL による図書館向けデジタル化資料送信サービス（図書館送信）以外は手つかずの状況にある。
- ・ 大規模デジタル化（平成 21、22 年度）以降、NDL のデジタル化は外部からは進展が見えにくい。我が国唯一の法定納本図書館である NDL には、日本最大の知識が集積されている。「機械が本を読む時代」を迎えた現在、基本的には全てデジタル化し、データとして扱えるようにすべきである。
- ・ 日本語の OCR（光学文字認識）の精度も格段に向上している。全文テキスト化、また検索可能性の確保・向上は、日本の知識基盤の整備の基本方針として掲げる必要がある。
- ・ 権利者との調整では、データ利用の優先権を与えるなど権利者にもメリットがある仕組みも考えられる。事例を積み重ねて調整を図っていくことが求められる。
- ・ NDL は、唯一の法定納本図書館として、他機関の所蔵資料も、イニシアチブをとってデジタル化を推進すべきである。図書館送信の枠組みも活用できれば、より期待は高まるであろう。また、NDL に所蔵がない国内資料について、デジタルデータで補完する方針を明確にしてもよいのではないか。
- ・ 忘れられる権利など、個人情報・プライバシー等について、検索可能性などテキスト化

ではデジタルと紙の違いも踏まえる必要がある。完全な基準を設けることは難しく、個別事例を積み重ねるほかないのではないか。

- ・ 公的機関の場合、検閲的なものと受け取られないようにすることが重要である。手続きと基本的な判断基準を検討し、ルール化することから始めるべきである。検索結果の中立性（neutrality）は、知識基盤を担う図書館にとって重要である。
- ・ OCR も完全ではなく、検索結果の数やランク付けによって、検索結果から資料にアクセスすることが事実上困難なものがあるように、全文テキスト検索も万能ではない。検索はアクセス手段の一つにすぎない。重要なことは、何らかの手段で資料へのアクセスが保障されていることである。
- ・ 中長期的には補償金制度について考える必要があるのではないか。ドイツでは、教育研究目的であれば、著作者への相当の報酬の支払いを条件として、絶版等だけでなく、新刊もその一部を複製、公衆送信することが認められている。日本でも、地方における知識基盤の確保や高齢化等を考えれば、費用負担の問題はあるが、こうしたオンラインサービスのニーズもあるのではないか
- ・ 平成 30 年改正で盛り込まれた著作権法第 47 条の 5 により、図書館に限らず、放送番組を録画したものをスニペット形式で所在を検索できるサービスを提供できるようになっている。放送番組等を含むオーディオビジュアルのアーカイブは、欧米諸国に比べて、民間を含め取組が遅れている。知識基盤の整備では、これも考えないといけない。
- ・ ヨーロピアーナ（Europeana）では、ファンディングエージェンシーに近い機能を前提としており、アグリゲータの立ち上げにも関与し、デジタル化に当たってはスケールメリットを重視している。ソフトなルールづくりと EU によるファンディングが両輪としてうまく機能している。ジャパンサーチでも、これらの強化について検討するのではないか。
- ・ 提言には、「オープン化」というキーワードを入れる必要がある。国が作成したものは、営利非営利を問わず、再利用可能とする潮流がある。その中で NDL の対応を考える必要がある。

【第 4 回部会】

- ・ メタデータの提供（共有）は、識別子付与の条件ともなっている。DOI が成功例であるが、識別子を通じて、メタデータを収集できる仕組みを作ることは、長期的にも意義が大きい。
- ・ 識別子とメタデータをデジタルアーカイブの基盤として位置付け、利活用を推進するには、伝統的な図書館目録による書誌データ管理では限界があり、広げて考える必要がある。提言において、NDL が実施方針で掲げている「相互運用性」よりも具体的なイメージを示す必要がある。
- ・ 学術出版社は、識別子を単独ではなく、知識創造サイクルの各段階や要素とリンクさせることで、最終的に研究評価（Research Evaluation）の付加価値を高めることを目指していると考えられる。NDL も、単に識別子を付与するだけでなく、どのような利活用を進めていくか、例えば、教育目的の利活用を考えるならば、そのために必要なものは何かを考え、取組を進めていく必要がある。
- ・ 英国図書館は、教育目的に限定して、大量のデジタルコンテンツをオープン化している。こうした積極的な試みは利活用を進めるために重要である。しかし、新型コロナウイルス

ス感染症対応の中で明らかとなったが、オープン化されたデジタルコンテンツがあっても、特に初等中等教育の現場で使いこなすことはまだ難しい。電子展示等をより使いやすい形に作り込み、グッドプラクティスを周知することが重要となる。NDL 自らではなく、外部の専門家や機関との連携協力の中で取り組むとよいのではないかと。

- ・ データの利活用を進めるには、まず使いやすいデータにして公開しておくことが重要である。特定の目的に作り込むことは、対象者の限定につながることもあり、必ずしも費用対効果がよいとはいえない。データが使いやすい形でオープン化されていれば、何かのきっかけで利活用が広がることもあるだろう。
- ・ Europeana のアグリゲータとの対比で考えると、全文テキストの（統合）検索は、ジャパンサーチのインターフェースから提供することは考え難い。他方で、書籍等分野の「つなぎ役」として、ジャパンサーチからアクセスされる書籍（図書館資料）をどのように使いやすくしていくかを考えなくてはならない。
- ・ 新型コロナウイルス感染症によって、物理的なコレクションに基づく図書館の弱点が浮き彫りになった。デジタル化の推進により、オンラインでコンテンツが利用できるようにするなど、災害等があっても図書館の活動に支障が生じない基盤とすることが一層求められる。
- ・ 「知的財産推進計画 2020」では、絶版等資料の図書館送信の拡大を念頭に置いた制度改正について、来年の通常国会への法案提出を目指すことが示された。著作権法改正への対応とともに、学校等の教育研究機関やその構成員が、図書館を経由せず、デジタル化した絶版等資料のデータを直接利用できるようになった場合に、どのような利活用が考えられるか、モデルを示すことが求められるのではないかと。
- ・ 新型コロナウイルス感染症への対応を含めて、対象や公開方法など、ウェブアーカイブの在り方を改めて考えるときにあるのではないかと。
- ・ 外国刊行資料について、ジャーナルをはじめ、学術コンテンツのオープンアクセス(OA)の割合は、十分とまでは言えないが、増えてきている。NDL の科学技術情報整備において、外国刊行資料の収集と提供は歴史的にも重要な意味を持ち、国民のアクセスを保障してきたが、今後は、こうした環境の変化を注視しつつ、その役割の見直しについて議論が必要ではないかと。
- ・ 研究データについて、国立情報学研究所(NII)の Research Data Cloud 等の取組が進展したことを踏まえると、NDL はメタデータに焦点を合わせるとともに、データ本体については、関係組織との役割分担(長期保存等)を検討する必要があるのではないかと。
- ・ 出版物に対する納本制度とのアナロジーを考えると、ナショナルコレクションの整備を担うNDLは、データの時代には様々な組織からメタデータを受け入れることになるだろう。その観点では、ジャパンサーチは、NDL が、伝統的な図書館資料ではないものについて、メタデータを受け入れ、整理して公開する窓口とも考えられるのではないかと。

【第 5 回部会】

- ・ デジタル環境における知識生産のエコシステムを考えなければならない。知識・価値の創出のサイクルを回すためには、データを収集・整備、公開するだけでなく、利活用とともに「共有」されることで発展的な構造が作られることを示す必要がある。

- ・統合的オンラインサービスは、ポーンデジタルへの移行の中で考える必要がある。例えば、海外の文献データベースでは引用文献とのリンクが標準となっているが、日本語文献については不十分な状況にある。また、著者等の典拠データも、デジタルであれば当然可能になるが、NDL のデータベースと他のデータベースとの相互連関を実現するための基盤がまだ確立されていない。
- ・NDL は、書誌コントロールのナショナルセンターを担っているが、出版や流通のデジタル化が進展する中で、現状では商業サービスとの境界を整理できておらず、電子書籍等の有償オンライン資料の制度収集が実現しない、機能しない一因にもなっている。電子書籍等の有償オンライン資料、オンデマンド出版等を始め、NDL や他の国内図書館が所蔵していない日本の出版物の総体を対象とする、いわばデジタルを含めた全国書誌も求められるのではないかと。
- ・複製物のデジタル送信や図書館向けデジタル化資料送信サービス（図書館送信）の拡大など、図書館サービスのデジタル化に向けて、権利者の利益保護に十分に配慮しつつ、幅広く検討、取組を進めていく必要がある。NDL が、図書館サービスのデジタル化に積極的な姿勢を示すことに大きな意義がある。
- ・知識の蓄積へのアクセスを地理的、物理的に開かれたものにするのがデジタル・シフトの最大の価値であり、納税者への還元でもある。少子高齢化や地方創生の視点も重要となる。
- ・デジタルの重要性が高まる中では、民間サービスとの接続や連携も考えていく必要がある。デジタル化資料について、Google 検索結果から発見、アクセスされるものも少なくない。そこから他の文献とのつながりをうまく見せられるかが、今後のオンラインサービスのポイントになる。これまでは資料やコンテンツを保有する機関との連携協力が多かったが、それだけではないものが求められる。
- ・個人情報保護やプライバシー等への配慮は、全文テキストの検索結果だけでなく様々な場面で求められている。例えば、出版後年月が経っていない資料については、本文のインターネット公開自体やその方法の検討が必要かもしれない。
- ・NDL が技術開発を担い、推進することを明確にする必要がある。例えば、一般に公開していないデータを NDL 内部では利用できるようにするなど、魅力的な開発環境（施設設備を含む。）の整備に取り組んでほしい。また、フェローシップや共同研究等だけでなく、インハウスの技術開発も重要である。
- ・永続性は個々のデータベースの管理者が取り組むべき課題でもあるが、維持できなくなったデータベースの継承は重要であり、NDL に期待されている。サービスを継続提供するかの判断は別として、ダークアーカイブとして引き受けることはあり得る。継承の基準や技術的課題等について、検討を進めていくことが求められるのではないかと。
- ・NDL は、インターネット資料収集保存事業（WARP）の中で、許諾に基づく民間ウェブサイトの収集も拡大してきているが、引き続き継続した取組が求められる。デジタル空間の重要性が高まる中で、国のウェブアーカイブの在り方を改めて検討する必要がある。NDL がその担い手になるとは必ずしも限らないが、積極的に発言することが期待される。

【第 6 回部会】

- ・副題の「人と機械が読む」について、機械が読めることによって、人が恩恵を受けることを本文でもう少し補った方がよい。基本計画で書き込まれると思うが、そのことがイメージしやすいようデータを提供するためのアプリケーションの開発等の具体的な取組にも言及があるとよい。
- ・提言では、ウィズコロナなど、新型コロナウイルス感染症との関係に直接言及するよりも、中長期の視点に立ち、新型コロナウイルス感染症を契機として明らかになった社会の認識の変化の中での図書館の課題や役割を書くことがより重要である。
- ・機械が読めることが全ての出発点となる。したがって、資料のデジタル化、全文テキスト化を着実かつ目に見える形で進めていくようにすることが、今回の提言の重点である。

- ・図書館のデジタル化は、データ駆動型研究を進めるに当たって大きな要素であり、不可欠であることをはっきりと書いてよい。
- ・学術情報コミュニケーションの変化は、プレプリントや査読のオープン化だけに留まらず、市民にとって様々な科学的知識に触れる機会の増加をもたらしている。シチズンサイエンスは、変化の流れの先にあるものとして整理すると分かりやすいかもしれない。
- ・プレプリントや査読のオープン化をめぐるっては、信頼性や真正性への懸念も示されており、新たな課題として言及した方がよい。

- ・NDL にとっての信頼性とは、アーカイブされた一つ一つの情報の信頼性を保証するというのではなく、総体としてエビデンスとなる情報を蓄積・保存し、アクセスや参照を可能にするを通じて、人々の知的活動を支えることにある。その趣旨が伝わるようにすべきである。
- ・「国立国会図書館デジタルコレクション」がウィキペディアの出典に挙げられることも増えている。インターネットを含め、流通する情報の信頼性を支える存在として NDL の役割は大きい。NDL に蓄積されているものが参照可能であることは、必要な時に情報を探し出せることの裏返しでもあると考えており、この点からも全文テキスト化は重要である。

- ・識別子は、機械が読むときに不可欠なシステムである。できる限り多くのデジタル化資料やデータに識別子を付与することが最重要である。また、識別子は、ある種のオープンリティと一体で整備、推進されるべきもので、NDL が果たすべき役割は大きい。
- ・学術情報であれば、プレプリントサーバや機関リポジトリなど、既存のプラットフォームも様々あるが、現在のところ、日本では、それらのデータを適切にリンクさせつつ、統合的に見せる利用環境は整えられていない。世界から遅れてしまわないためにも、基盤的な研究開発やシステム開発が求められる。
- ・識別子の普及やデータ利活用の促進を図るに当たって、NDL が全てを担うわけではない。エンドユーザー側のアプリケーション開発を民間に任せてしまう考え方もある。NDL と民間の役割分担、その境界領域をどうするかは考えておかねばならない。

- ・図書館資料の複製物のデジタル送信や絶版等資料の利活用が教育に資することは疑いない。学校図書館の所蔵資料も多くはなく、大学図書館も様々なので、コンテンツの提供者として NDL やジャパンサーチの役割は大きい。オンライン教育の実現を含め、教育分野の利活用促進についてより書きこむべきである。

- 例えば、修士論文のテーマとして NDL のデータを使って研究するなど、大学院生を対象とすることはどうか。国立情報学研究所のデータセット共同利用研究開発センターは、民間を含む研究利用可能なデータセットを整備しており、多くの大学院生が利用している。
- NDL 内部でしか使えないデータを使って研究し、成果を残すことができるとよい。民間の研究者に対して、必ずしも成果のオープンソース化を条件として課せられるかは分からないが、共同開発したオープンソースを基に、民間企業がサービスを作ることも考えられる。
- 外部との連携を推進するに当たっては、NDL プロパーの職員にも人材がいることが重要である。研究開発を専門とする組織が必要であることは間違いない。

- クラウドソーシングについて、一般化した言葉ではあるが、人によって受け取り方が異なり誤解を招くおそれもある。シチズンサイエンスとして整理した方がよい。なお、市民の知的活動を後押しするプロジェクト等について、欧米では public engagement という言葉がよく用いられるが、日本語でそのまま用いることはなかなか難しいかもしれない。
- Europeana 1914-1918 のように、社会や家庭に埋もれ、従来は収集対象ではなかった資料を掘り起こしてデジタルアーカイブとして利活用を促進し、草の根アーカイブを後押しできるとよい。ただし、出版物ではないものに NDL がどのように関わるべきか、ジャパンサーチを始め、ナショナルアーカイブの役割と合わせて検討していくことが必要である。
- シチズンサイエンス等も自然と進むものではない。パートナーシップの担い手となり、市民の活動を積極的に後押しできるような人材が NDL 内部に当然必要になる。