

第五期国立国会図書館科学技術情報整備基本計画の進捗報告

「第五期国立国会図書館科学技術情報整備基本計画」は、第 13 回科学技術情報整備審議会(令和 3 年 1 月 13 日開催)において国立国会図書館長に提出された「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—『人と機械が読む時代』の知識基盤の確立に向けて—」を受けて、国立国会図書館が令和 3 年度から令和 7 年度までの 5 年間に取り組むべき事項を定めたものである。

ここでは、同計画において取り組むべき事項として掲げられたものについて、その進捗状況を取りまとめた。このうち、前回の審議会(令和 4 年 8 月)以降の特筆すべき進捗として、次が挙げられる。

- 補正予算による国内刊行図書のデジタル化等を引き続き進め、令和 2 年度補正予算分等約 32 万点を送信サービス対象としたほか、令和 3 年度補正予算分約 30 万点を館内公開した。
- 国立国会図書館デジタルコレクションをリニューアルし、令和 2 年度までにデジタル化した資料 247 万点の全文テキストデータの検索を可能とした。
- 障害者用資料検索サービスである、みなサーチの β 版を試験公開し、デジタル化資料約 247 万点の全文テキストデータを視覚障害者等向けに送信開始した。
- 有償等オンライン資料の制度収集を開始した。

目次

「人と機械が読む時代」に向けて取り組むべき事項

1 利活用促進のための取組.....	2
(1) 利活用の基盤となる所蔵資料のデジタル化・全文テキスト化の推進.....	2
(2) 情報へのアクセス環境の改善.....	3
(3) データの組織化・オープン化.....	5
(4) データの利活用促進のための環境整備.....	5
2 恒久的保存のための取組.....	6
(1) 資料収集の強化.....	6
(2) 未収資料・データのデジタル収集.....	7
(3) デジタル資料の長期保存.....	8

「人と機械が読む時代」に向けて取り組むべき事項

1 利活用促進のための取組

(1) 利活用の基盤となる所蔵資料のデジタル化・全文テキスト化の推進

(a) 所蔵資料デジタル化の戦略的な推進

- ① 「資料デジタル化基本計画 2021-2025」(令和 3 年国図電 2103012 号)に基づき、所蔵資料のデジタル化を加速する。特に、和図書については対象年代を拡大し、2000 年刊行までのものを対象として、集中的にデジタル化を行う。また、和雑誌については、刊行後 5 年以上経過したものを対象範囲とし、社会的ニーズ及び学術的ニーズを踏まえ、雑誌記事索引採録対象誌及び学協会等からのデジタル化要望があるもののデジタル化を優先的に行う。

(進捗)

- 令和 3 年度補正予算(第 1 号)により、国内刊行図書 301,780 分冊(45,487,848 コマ)のデジタル化を実施し、令和 5 年 5 月までに国立国会図書館デジタルコレクションにおいて館内限定公開した。
- 雑誌についても、学協会から要望のあったものを含む学術雑誌 9,933 分冊(716,128 コマ)のデジタル化を実施した。
- 図書・雑誌 3,892 分冊(382,248 コマ)を内製によりデジタル化した。
- 令和 5 年 6 月時点で、約 353 万点のデジタル化資料を提供しており、うち約 59 万点がインターネット公開、約 184 万点が送信対象、約 110 万点が館内限定公開となっている。
- 令和 5 年度は、令和 4 年度補正予算(第 2 号)による国内刊行図書約 40 万点等のデジタル化に着手している。

- ② 博士論文、録音資料・映像資料、古典籍資料、日本近現代政治史料、国会情報等の各資料群のデジタル化に引き続き取り組むほか、新たに新聞のデジタル化にも着手する。

(進捗)

- 帝国議会議事印刷資料・地図・日本占領関係資料・古典籍・博士論文・児童書・新聞等 79,884 点(3,920,742 コマ)及びカセットテープ 416 件(609 巻、13,604 コマ)を外部委託によりデジタル化した。
- VHS ビデオテープの映像 2,018 点、その他資料群 9,223 分冊(430,194 コマ)を内製によりデジタル化した。

(b) デジタル化資料の全文テキスト化等の実施

- ① 全てのデジタル化資料(古典籍資料、録音資料・映像資料等を除く。)の全文テキスト化に取り組む。

(進捗)

- 令和 3 年度に外部委託により OCR 処理プログラムを用いて作成したデジタル化資料約 247 万点の全文テキストデータについて、読み上げ順序の改善等の品質向

上を行った。

- 今後は、令和 3 年度に開発した OCR 処理プログラム(NDLOCR)を用いて、令和 3 年度以降にデジタル化した資料の全文テキストデータを作成する予定である。

- ② 生成されるテキストデータは、著作権、個人情報、プライバシー等に配慮した上で検索利用(単なる文字列検索にとどまらない高度な検索も想定する。)に供するほか、AI の学習用データセットとしての提供及び視覚障害者等への提供にも取り組む。

(進捗)

- 令和 4 年 12 月に国立国会図書館デジタルコレクションをリニューアルし、令和 3 年度に作成したデジタル化資料約 247 万点の全文テキストデータの検索を可能とした。
- 令和 4 年度に実験開発した古典籍専用 OCR 処理プログラム(NDL 古典籍 OCR)により、古典籍資料約 8 万点の全文テキストデータを作成し、1 月に実験サービスである次世代デジタルライブラリーに搭載して著作権保護期間の満了した図書約 28 万点とあわせて検索可能とした。
- 令和 5 年 3 月に障害者用資料検索サービスである、みなサーチの β 版を試験公開し、デジタル化資料約 247 万点の全文テキストデータを視覚障害者等向けに送信開始した。
- 書誌データやデジタル化資料を用いたデータセット 4 件、ソースコード 5 件を GitHub 上で公開した。

- ③ 国立国会図書館内の開発研究体制を整備し、外部のオープンソースコミュニティや有識者等の知見も取り込み、レイアウト認識処理を含む光学式文字認識(OCR)の精度向上や、全文テキストデータの構造化に向けた調査研究に取り組み、その成果をできる限りオープンな利用条件で提供する。

(進捗)

- NDLOCR 及び NDL 古典籍 OCR のプログラムを、オープンソース(CC-BY)として公開した。
- NDLOCR について、視覚障害者等の読み上げ用テキストの作成を可能とするための機能改善及び精度改善を行った。令和 5 年度も引き続き改善に取り組み、令和 5 年 6 月に公開した。

(2) 情報へのアクセス環境の改善

- ① 内閣府を始めとする関係府省・機関と協働し、我が国の幅広い分野・地域のデジタルアーカイブと連携する「ジャパンサーチ」の一層の拡充及び海外も含めた活用促進に取り組む。

(進捗)

- 令和 4 年度には、新たに 6 連携(つなぎ役)機関、32 データベースと連携し、令和 5 年 6 月 15 日時点で、40 連携(つなぎ役)機関、208 データベース、連携メタデ

ータ数は約 2,850 万件となった。

- ② 書籍等分野のデジタルアーカイブのつなぎ役である「国立国会図書館サーチ」、蔵書検索・申込システムである「国立国会図書館オンライン」、国立国会図書館のデジタル資料の保存・提供の基盤である「国立国会図書館デジタルコレクション」等について、データ面及び機能面を拡充する。

(進捗)

- 国立国会図書館サーチにおいて、新たに 16 機関と連携を開始するとともに、連携済みの 1 機関について API による連携方式への切り替えを実施、また同じく 54 機関についてデータの入替・追加等を実施した。これにより、令和 5 年 6 月時点で、95 機関、連携メタデータ数 1 億 3 千万件となった。
- 令和 4 年 12 月に国立国会図書館デジタルコレクションをリニューアルした。(再掲)

- ③ 各情報資源への適切なナビゲートを実現するため、次世代の統合的なオンラインサービスを整備する。その際、視覚障害者等へのインクルーシブな対応も強化する。

(進捗)

- 現在の国立国会図書館サーチ、国立国会図書館オンライン及びリサーチ・ナビを統合した「国立国会図書館サーチ」の公開(令和 6 年 1 月予定)に向けて開発作業を進めている。
- 令和 5 年 3 月に障害者用資料検索サービスである、みなサーチの β 版を試験公開した。(再掲)正式版の公開(令和 6 年 1 月予定)に向けて準備を進めている。

- ④ 著作権法改正の動向を踏まえた上で関係団体・府省等と協議を行い、デジタル化資料のうち絶版等資料に係るデジタル化資料送信サービスの提供範囲を図書館等から個人に拡大する。

(進捗)

- 令和 4 年 5 月に開始した個人向けデジタル化資料送信サービスについて、令和 5 年 1 月にプリントアウト機能の提供も開始した。なお、令和 5 年 6 月時点で利用規約に同意した利用者の累計は約 13.4 万人である。
- 令和 5 年 2 月から、令和 2 年度補正予算(第 3 号)によりデジタル化した国内刊行図書等のうち入手困難と判定された約 32 万点の送信を開始した。なお、令和 5 年 6 月時点の送信資料点数は約 184 万点である。
- 海外居住個人へのサービス拡大についても検討に着手している。

- ⑤ 図書館資料の複写物の送信サービス(複写物の電子ファイルでの送信)については、著作権法改正の動向を踏まえた上で関係団体・府省等と協議を行うとともに、制度設計及びその実施に取り組む。

(進捗)

- 関係者と協議を行ってとりまとめたガイドライン等に基づき、情報システムの改修や

規則類の改正等の準備を進めている。なお、サービス開始は令和 6 年 6 月を想定している。

- ⑥ デジタル化資料の著作権処理を加速し、インターネット公開資料の一層の拡大を図る。

(進捗)

- 和図書、雑誌及び博士論文のデジタル化資料について、保護期間満了を確認したものの、または権利者の許諾を得たもの約 4,500 点をインターネット公開した。

(3) データの組織化・オープン化

- ① 「識別子に係る実施方針」(令和 2 年国図電 2001067 号)に基づき、関係機関と協力して、国立国会図書館が作成する書誌データの相互運用性の向上や、「ジャパンサーチ」による文化情報資源のリンクデータ基盤形成への貢献等に取り組むとともに、これらの取組の一層の拡充に向けた検討を行う。

(進捗)

- VIAF(バーチャル国際典拠ファイル)に約 3.4 万件のデータを提供し、国内外の機関の典拠との関連づけを促進した。
- ジャパンサーチの正規化 URI 約 1,800 件を Wikidata に登録した。

- ② 国立国会図書館がデジタル化した資料の画像データや全文テキストデータの二次利用の枠組みを整備するほか、書誌データのオープンデータセットの継続的な提供にも取り組む。

(進捗)

- デジタルコレクション書誌情報等の既存オープンデータセットの更新を実施した。
- 転載等を目的とした利用に限って、国立国会図書館内・図書館送信限定公開デジタル化資料の画像データの提供の試行を継続した。
- デジタル化資料から著作権保護期間を満了した画像を切り出して提供する電子展示会「NDL イメージバンク」に随時コンテンツを追加し、令和 5 年 6 月時点で約 7,700 点の画像を公開している。

- ③ 国立国会図書館が保存するデジタル化資料、オンライン資料等のデジタル資料へのデジタルオブジェクト識別子(DOI)付与に、引き続き取り組む。

(進捗)

- 令和 4 年度はデジタル資料約 60 万件に DOI を付与した。

(4) データの利活用促進のための環境整備

- ① 若手研究者等を対象とする共同研究プログラム等により、デジタル化資料から抽出した大量の画像データや全文テキストデータ、「国立国会図書館インターネット資料収集保存事業」(WARP)から抽出したデータ、「ジャパンサーチ」で収集したメタデータ等、国立国会図書館が作成又は収集したデータを外部の機関や有識者等が利活用し、協働できる環境の整備に取り組む。

(進捗)

- デジタルライブラリーカフェ等のイベントの開催や外部イベントへの協力により、当館データ利活用のための情報を発信した。
- ジャパンサーチで収集したメタデータ約 2,500 万件を RDF データに変換のうえ提供した。

② 関係機関や外部の有識者等とも協力の上、「ジャパンサーチ」の利活用機能等も用いて、上記のデータを活用した教育シーン(オンライン教育を含む。)での利活用モデルの構築や実践支援、一般市民の主体的な参加を促し、その知的な活動を後押しするようなプログラムの開発に取り組む。

(進捗)

- 「デジタルアーカイブフェスージャパンサーチ・デイ」や「ジャパンサーチ API ハッカソン」等のイベントや、学習教材を作成するワークショップにおいてジャパンサーチの活用方法等紹介し、教育活用の促進を図った。
- 地方自治体や学校における地域学習イベントに講師として参加し、ジャパンサーチの利活用を支援した。

③ 前 2 項の取組をコーディネートし、橋渡し役となる人材の育成に取り組む。

(進捗)

- 「ジャパンサーチの連携機関向けギャラリー作成ワークショップ」の開催を通じて、データの活用方法及び事例の共有を図った。
- SNS やウェブサイトを通じて、利活用事例の共有や使い方動画の公開などを継続している。

④ メタデータの外部提供インターフェイス(API)による提供については、新しいプロトコルへの対応、項目の追加、新規フォーマットへの対応等を行い、より利活用しやすくする。

(進捗)

- NII・JST・NDL 連絡会議に設置されたメタデータ WG の活動の一環として、令和 5 年 3 月に「メタデータ流通ガイドライン」の研究データ編、同 6 月に古典籍編を公開した。

2 恒久的保存のための取組

(1) 資料収集の強化

- ① 我が国の有償等オンライン資料の収集・保存・利用のため、制度収集を開始し、これらの基盤の整備に取り組む。また、制度収集開始までの間、引き続き任意提供による有償等オンライン資料の収集に努める。

(進捗)

- 関係団体との協議・調整のうえ、1 月から有償等オンライン資料の制度収集を開

始した。引き続き着実に収集を行うとともに、一層の周知・広報に努める。

- 令和 4 年 12 月にデジタル出版者連盟が運営するリポジトリを収集除外することとした。

- ② 外国資料については、オープンアクセスなど刊行・流通形態の変化に留意しつつ、引き続きコアジャーナルを中心とした電子ジャーナル等の契約を着実に行う。また、電子ジャーナルのバックファイルを整備するとともに、セーフティネットの観点から永続的なアクセスの確保に努めた上で、大学等とのネットワークの構築に協力する。

(進捗)

- 科学技術関係資料費による令和 5 年の外国雑誌購読タイトル数は、冊子 134 タイトル、電子ジャーナル 1,868 タイトルである。
- 数学・物理学分野等の購読タイトル見直しを実施した。
- ACS Publications、Springer Nature の購読タイトル見直しを実施し、新規導入タイトルの購読を開始した。
- 科学技術情報整備等に係る有識者ヒアリングを 2 回開催した。

- ③ 我が国の学術情報でもある海外刊行の国内学協会誌の調査及び収集に引き続き努める。

(進捗)

- 海外刊行国内学協会誌の刊行状況を調査し、新たに Digestion(日本消化管学会)を収集した。

(2) 未収資料・データのデジタル収集

(a) 未収資料のデジタルデータ及びメタデータの収集

- ① 未収(国立国会図書館が所蔵していない)の地域資料や、海外の機関が所蔵する日本関係資料等について、関係機関と連携・調整の上、デジタルデータでの収集・保存・提供に積極的に取り組む。そのため、受入れの条件や手続等を整備する。

(進捗)

- 令和 4 年 7 月に案内ページを公開し、国内の公共図書館から 249 点の収集を行った。引き続き、支部図書館を含めた国内図書館からの収集に取り組んでいる。
- 国公私立大学図書館協力委員会との間で連携に係る検討を行い、先行事例の共有や収集要件の説明、デジタル化等に係る情報共有や研修の実施を図ることとした。

- ② 「ジャパンサーチ」、「国立国会図書館サーチ」等を通じて、未収資料や未連携のデータベース等のメタデータの収集、組織化及び提供に取り組む。

(進捗)

- 書籍等分野のつなぎ役として、国立国会図書館サーチの連携を拡張して書籍等分野のデジタルアーカイブのメタデータ収集し、ジャパンサーチに連携している。
- 令和 5 年 6 月時点で、当館提供分以外の書籍等分野のメタデータ約 324 万点

をジャパンサーチで提供している。

- ③ 資料のデジタル化やメタデータの整備が進んでいない機関に対する支援策を検討する。

(進捗)

- 職員を派遣して「図書館資料デジタル化・公開に伴う権利処理」等の研修や意見交換会を実施したほか、支部図書館向けの研修も実施した。

(b) データベース、研究データ等の収集

- ① 存続が困難となったデータベースやデジタルアーカイブの継承に取り組む。「国立国会図書館東日本大震災アーカイブ」におけるアーカイブの継承を継続するほか、その経験を踏まえ、これらを継承する際の課題や条件等を検討する。

(進捗)

- 「東日本大震災アーカイブシンポジウム」等を通じて、継承の取組に係る課題整理や周知を図った。
- 令和4年度内に閉鎖したひなぎくの連携先はなかったが、今後閉鎖が予定されている機関との調整に着手している。

- ② 研究データについては、国立情報学研究所、科学技術振興機構等との分担を前提に、地域資料や特定の研究分野に結び付かない分野横断的なものを対象とする等、国立国会図書館が担うべき役割の整理を進め、収集・保存・提供する際の課題や条件等を検討する。

(進捗)

- NII・JST・NDL 連絡会議に設置されたメタデータ WG の活動の一環として、令和5年3月に「メタデータ流通ガイドライン」(研究データ編)を公開した。(再掲)
- ③ 民間ウェブサイトの収集については、近年の動向を踏まえて課題を整理する。

(進捗)

- 民間ウェブサイトについて、諸外国における収集動向を調査し、技術面・運用面での課題とその対応事例を整理した。

(3) デジタル資料の長期保存

- ① 「国立国会図書館デジタル資料長期保存基本計画 2021-2025」(令和3年国図電 2102181号)に基づき、国立国会図書館が所蔵するパッケージ系電子出版物の長期利用が可能となるよう、国立国会図書館内の実施体制を整備するとともに、マイグレーション等の作業に取り組む。

(進捗)

- 手順が確立している媒体を対象として、内製によるマイグレーションを実施している。実績値は、FD(3.5インチ)440点(うち成功点数418点)、USBメモリ47点、MO5点、書換形(RW)光ディスク24点である。
- 光ディスクの汎用的なマイグレーションに係る技術調査等を行ったうえで、令和4

年度補正予算(第2号)による光ディスク約1.5万点のマイグレーションに着手している。並行して、マイグレーションしたデータを利用可能とするため、エミュレーションに係る技術調査にも着手している。

- ② デジタル化資料、オンライン資料等のデジタル資料に関しては、合理的かつ安定的な保存環境を整備するとともに、これらの長期保存対策に必要な技術的調査研究を行う。

(進捗)

- デジタル化資料の保存用画像(光ディスク及び HDD)の長期保存に向け、LTO への移行作業を継続しており、令和4年度は光ディスク約1.1万枚、HDD28台の移行を実施した。
- 国立国会図書館デジタルコレクションについては、令和4年12月に新システムへの移行を完了した。(再掲) WARPについては、次期システムの検討を開始した。

- ③ 前項で得られた知見を踏まえ、関係機関と長期保存に関する課題等の共有を可能とするコミュニティの形成を図る。

(進捗)

- 令和3年度に実施した長期保存に係る国内機関の実態調査結果を公開した。
- 「Digital Preservation Handbook」(第2版)を翻訳し、図書館員向けの遠隔研修教材にその内容を反映させた。