

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、今回の特筆すべき進捗として、次が挙げられる。

- 令和 2 年度補正予算により約 30 万点のデジタル化を実施し、古典籍等を除いた国立国会図書館デジタルコレクション収録のほぼ全てのデジタル化資料 247 万点(2 億 2,300 万コマ)に対してテキスト化を実施した。
- 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から個人に拡大し、令和 4 年 5 月 19 日に送信サービスを開始した。
- 令和 5 年 1 月の有償等オンライン資料の制度収集開始に向け、国立国会図書館法等の関係法規が改正された。

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.....	1
(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.....	1
(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.....	4
(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.....	6
(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 2 年度補正予算(第 3 号)により、国内刊行図書の大規模なデジタル化を実施
- 雑誌についても、学術雑誌を中心に着実にデジタル化を実施
- 国内刊行図書 295,854 分冊(45,726,019 コマ)、雑誌 4,545 分冊(397,082 コマ)を外部委託でデジタル化
- 東京本館にメディア変換室を整備し、図書・雑誌等の内製デジタル化を本格的に開始。5,143 分冊(162,059 コマ)を同室でデジタル化
- 令和 3 年度にデジタル化した国内刊行図書を令和 4 年 5 月 31 日に館内で提供開始

- 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)

- 日本占領関係資料・古典籍等 10,514 点(2,187,908 コマ)、ソノシート 14 冊(51 枚)及びカセットテープ 187 件(615 巻)を外部委託でデジタル化
- 帝国議会議事印刷資料のデジタル化を開始
- 新聞マイクロ資料のデジタル化試行を開始
- LD の内製デジタル化を完了
- VHS ビデオテープの内製デジタル化を開始

(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)

- 古典籍等を除いた国立国会図書館デジタルコレクション収録のほぼ全てのデジタル化資料 247 万点(2 億 2,300 万コマ)に対してテキスト化を実施
- テキスト化対象資料のうち 1960 年代以降の図書資料についてはレイアウト情報付与も実施

- 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)

- 作成した全文テキストデータのうち著作権保護期間の満了した図書(28 万点)について、[次世代デジタルライブラリー](#)に搭載し、検索を可能とした。

- 書誌データやデジタル化資料を用いたデータセット 3 件、プログラム 1 件、OCR 関連事業で作成した学習用データセット 2 件を GitHub 上で公開

- 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)

- 当館資料に最適化したオープンソースの OCR 処理プログラムの開発を実施し、CC BY4.0 で GitHub から公開(令和 4 年 4 月 25 日)
- 全文テキストデータのうち、著作権保護期間が満了した図書を対象とした NDL N-gram Viewer を開発し、提供開始(令和 4 年 5 月 31 日)。その他、同形異音語の読みの改善等に係る研究、固有表現抽出に係る調査等を実施

(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 6 月 29 日時点で、33 連携(つなぎ役)機関、174 データベース、連携メタデータ数は約 2,590 万件
- 令和 3 年 9 月、「ジャパンサーチ戦略方針2021-2025」を策定

- 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は、新規に 11 機関と連携を開始
- 国立国会図書館サーチと連携済みの 5 機関について、API による連携方式への切り替えを実施
- 国立国会図書館サーチと連携する 38 機関についてデータ入替、データ項目(サムネイル URL)追加等連携維持のための取組を実施

- 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)

- 開発に向けた実施計画を策定
- サービス骨子についての調査検討を実施

- 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)

- 関係団体と協議を行い、「国立国会図書館のデジタル化資料の個人送信に関する合意文書」(令和3年12月3日)を取りまとめた。
- 令和4年5月19日に送信サービスを開始

- 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)

- 関係者間で協議を行うための「図書館等公衆送信サービスに関する関係者協議会」とテーマ毎の各分科会が設置され、制度設計に向けた検討に参画

- 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)

- 和図書及び官公庁出版物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及び権利者の許諾を得たもの 3,988 点のインターネット公開を実施

(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
- Wikidata へ、ジャパンサーチの正規化 URI を約 5,000 件登録

- 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
- 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開始
- 二次利用を促進するため、デジタル化資料のうち著作権保護期間を満了した画像を提供する電子展示会「NDL イメージバンク」を作成

- 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3年度は、デジタル資料約3.7万件にDOIを付与

(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
- ジャパンサーチで収集したメタデータについては、RDFデータへの変換を継続し約2,200万件を提供

- 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)

- 令和3年8月、文部科学省と連携したイベント「こども霞が関見学デー」でジャパンサーチの利活用機能を体験するプログラムを実施
- SNSや公式ウェブサイトで利活用事例の共有や使い方動画の公開など、ジャパンサーチの利活用に関する広報を実施

- 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)

- 令和3年12月及び令和4年3月に連携機関やジャパンサーチの活用者を対象に意見交換会を開催し、データ利活用に係る知見や経験を共有

- 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「メタデータ流通ガイドライン」(ドラフト版)を公開

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)

- 有償等オンライン資料の制度収集開始に向け、著作権者団体、出版関係団体への説明を行い、制度の趣旨について一定の理解を得た。あわせて、国立国会図書館法が改正され、令和4年6月1日に公布された。収集フロー、収集対象から除かれるリポジトリの認定・メタデータ連携等について館内外の関係者との調整を行う等、令和5年1月の制度施行に向け、着実に準備を進めている。

- 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4年の外国雑誌購読タイトル数は、冊子262タイトル、電子ジャーナル1,889タイトル
- 英国機械学会(IMechE)の会議録等を収録したIMechE Complete CollectionやIET Journals Archiveを導入
- ScienceDirect等のバックファイルを購読
- Journals@Ovidの購読タイトル見直しを実施
- 科学技術情報整備に係る有識者ヒアリングを3回開催

- 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)

- 学協会を対象としたアンケートを実施
- 海外刊行国内学協会誌の刊行状況を調査し、新たに Neuro-Oncology(日本脳腫瘍学会)、Pancreas(日本膵臓学会)等、冊子5タイトル、電子2タイトルを収集

(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)

- 支部図書館のデジタル資料の試行収集、デジタルデータ提供の意向に関するヒアリングを実施
- 主に国内刊行の図書・雑誌を想定し、国立国会図書館で未所蔵であること、絶版等の理由により入手困難な所蔵資料であることを条件に大学・公共図書館等のデジタル化資料を受け入れる「国立国会図書館未入手困難資料のデータ収集事業」を本格的に実施するため、令和4年7月1日から公式ウェブサイトでの案内を開始

- 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、ジャパンサーチにおける書籍等分野のデジタルアーカイブのメタデータ収集を実施
- 令和3年度末時点で、当館提供分以外の書籍等分野のメタデータ約270万点をジャパンサーチで提供

- 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)

- 講師派遣型研修「図書館資料デジタル化・公開に伴う権利処理」等の研修事業を実施

(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)

- 茨城県東日本大震災デジタルアーカイブのデータを承継し、公開

- 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)

- 国立情報学研究所及び科学技術振興機構と協力してメタデータの相互運用性に関する検討ワーキンググループを開催し、3機関のシステムのメタデータ連携についての現状整理や将来像検討等の実務的な検討を進め、その成果を、第9回NII・JST・NDL連絡会議で共有

- 作成した「メタデータ流通ガイドライン」(ドラフト版)を公開

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、社会的に注目されるイベント・テーマに関するものや消失しやすく時限的なものを収集するなど、資料収集方針書の改定に当たり、収集対象の記述を見直した。

(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 年国函電 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)

- 内製によるマイグレーションの本格実施を開始
- 実績値は、FD(3.5 インチ)6,553 点(うち成功点数 6,546 点)、USB メモリ 44 点、カード型 CD95 点(うち成功点数 90 点)、RW 型光ディスク 1 点
- マイグレーション成果物のうち、約 300 点をデジタルコレクションへ登録し、令和 4 年 3 月 15 日から利用提供を開始
- 5 インチ FD の環境構築を行い、約 200 点に対してマイグレーションを試行
- FD のマイグレーション技術及びマイグレーション後データの再生環境に関する調査、長期保存に係る国内機関の実態調査、並びに光ディスクの検査手法に係る調査を実施

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 への移行作業に着手
- 光ディスク 10,087 枚、HDD147 台を LTO に移行
- WARP については、機器更新を行い、十分な容量の電子書庫を確保
- サービスごとに館内配置機器とクラウドを使い分けること等、電子書庫の将来的な方向性の検討を継続
- プロトタイピングの手法を活用しながらクラウドを活用した次期国立国会図書館デジタルコレクションの基本設計を実施
- 諸外国の国立図書館等の電子情報保存システム及び製品等の文献調査を実施

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)

- 長期保存に係る国内機関の実態調査を実施