

JEPAセミナー
2022年2月2日

国のデジタル情報基盤の充実に向けて 国立国会図書館ビジョン2021-2025の取組から

国立国会図書館 電子情報企画課
福林 靖博

本日の内容

- 1 国立国会図書館のビジョン
- 2 所蔵資料デジタル化の推進
- 3 デジタル化資料の個人への送信 ※公衆送信ではありません
- 4 デジタル化資料の全文テキストデータ
- 5 デジタル情報の長期保存
- 6 ジャパンサーチ

1. 国立国会図書館のビジョン：全体像

「デジタルシフト」



- ①ユニバーサルアクセスの実現
- ②国のデジタル情報基盤の拡充

5年間の重点

情報資源と知的活動をつなぐ7つの重点事業

「基本的役割」



- ① 国会活動の補佐
- ② 資料・情報の収集・整理・保存
- ③ 情報資源の利用提供
- ④ 各種機関との連携協力

事業の全体像
(4つの領域)

1. 国立国会図書館のビジョン：7つの重点事業

ビジョン 2021-2025

国立国会図書館のデジタルシフト

— 情報資源と知的活動をつなぐ7つの重点事業 —

ユニバーサルアクセスの実現



国のデジタル情報基盤の拡充



情報技術によって私たちの社会活動やコミュニケーションのあり方が変容し、さらに新型コロナウイルスが大きな影響を及ぼす中で、社会のデジタルシフトが加速しようとしています。この状況を踏まえ、国立国会図書館は、情報資源と様々な知的活動を確につなげていくために、今後5年間で「国立国会図書館のデジタルシフト」推進期間と位置付け、7つの事業に特に重点的に取り組みます。これらの重点事業は、将来にわたる全ての利用者に多様な情報資源を提供するユニバーサルアクセスを実現する事業と、そのための恒久的なインフラとなる国のデジタル情報基盤の拡充を図る事業から構成されます。

2. 所蔵資料デジタル化の推進：経緯

2000年	<ul style="list-style-type: none"> 資料デジタル化を開始。著作権処理を行いインターネットで公開（2～4万冊／年）
2008年	<ul style="list-style-type: none"> 資料デジタル化及び利用に係る関係者協議会発足
2009年	<ul style="list-style-type: none"> 著作権法改正（NDLで保存目的のデジタル化が明確化）
2009～ 2011年	<ul style="list-style-type: none"> 大規模デジタル化事業実施（2009年度、2010年度補正予算） 図書66万点、雑誌22万点、古典籍7万点、博士論文14万点、官報等のデジタル化実施。
2012年	<ul style="list-style-type: none"> 著作権法改正（図書館等への絶版等入手困難な資料の送信が可能に）
2014年	<ul style="list-style-type: none"> 図書館向けデジタル化資料送信サービス（図書館送信）開始
2015年	<ul style="list-style-type: none"> 災害関係資料のデジタル化（2014年度補正予算） 震災・災害関係の図書約6万点、雑誌約2万点のデジタル化実施
2018年	<ul style="list-style-type: none"> 著作権法改正（外国の図書館等へも絶版等入手困難な資料の送信が可能に）
2019年	<ul style="list-style-type: none"> 外国の図書館等にも図書館向けデジタル化資料送信サービスを拡大 デジタル化内製の実験プロジェクト開始
2021年	<ul style="list-style-type: none"> 国内刊行図書のデジタル化（2020年度補正予算）、資料デジタル化推進室の設置 「資料デジタル化基本計画2021-2025」の策定 著作権法改正（特に第31条第3項：絶版等資料の個人（家庭）への送信） 国内刊行図書のデジタル化（2021年度補正予算）
2022年	<ul style="list-style-type: none"> 個人向けデジタル化資料送信サービス（個人送信）開始

2. 所蔵資料デジタル化の推進：現況

資料	これまでの取組	インターネット公開	図書館送信	NDL館内限定	合計	今年度からの取組
図書	明治期以降、1968年までに受け入れた図書 震災・災害関係資料の一部（1968年以降に受け入れたものを含む。）	36万点	55万点	8万点	99万点	2000年までに刊行・受入したもの（対象：約170万冊） ※官庁出版物は2000年以降も含む ※5年間で100万冊以上のデジタル化を目指す
雑誌	明治期以降に刊行された雑誌（刊行後5年以上経過したもの）	2万点	82万点	51万点	134万点	刊行後5年以上経過した雑誌 →学協会等からデジタル化要望があるものを優先する。
博士論文	1990～2000年度に送付を受けた論文	2万点	13万点	2万点	16万点	1989年度以前に送付を受けたもの。
新聞	（未実施）	-	-	-	-	試行（日本新聞協会と合意したもの等。ただし、商用データベース等で提供されているものは対象外。）
その他	古典籍、地図、官報、録音・映像資料、憲政資料、日本占領関係資料等	18万点	4万点	9万点	30万点	（継続） 新たに、日系移民関係資料を追加
※2022年1月時点の提供点数		56万点	153万点	71万点	279万点	

2. 所蔵資料デジタル化の推進：計画

資料デジタル化基本計画2021-2025

<p>評価要素</p>	<ul style="list-style-type: none"> • 唯一性・希少性 • 資料の利用機会の拡大（インターネット公開や図書館・個人送信が見込まれるか） • 資料の劣化状況、保存の緊急性 • デジタル化への社会的・学術的ニーズ • 国や世界の体系的なデジタルコレクション構築への貢献
<p>対象資料</p>	<p>日本で刊行された資料（外国刊行の日本語資料・日本関係資料も含む）</p> <ul style="list-style-type: none"> • 図書（2000年までに刊行されたもの）※官庁出版物はそれ以降も含む • 雑誌（刊行後5年以上経過したもの） • 古典籍資料 • 録音・映像資料 • 博士論文 • 他（憲政資料、日本占領関係資料、<u>日系移民関係資料</u>、地図、<u>新聞<試行></u>）
<p>利用提供</p>	<ul style="list-style-type: none"> • 「国立国会図書館デジタルコレクション」で提供 • 本文テキストデータの作成を推進し、全文検索を可能に • デジタル化済み原資料は原則として利用停止 • 公開範囲：館内限定・図書館送信・インターネット公開

2. 所蔵資料デジタル化の推進：補正予算による加速

2020年度補正予算（第3号）

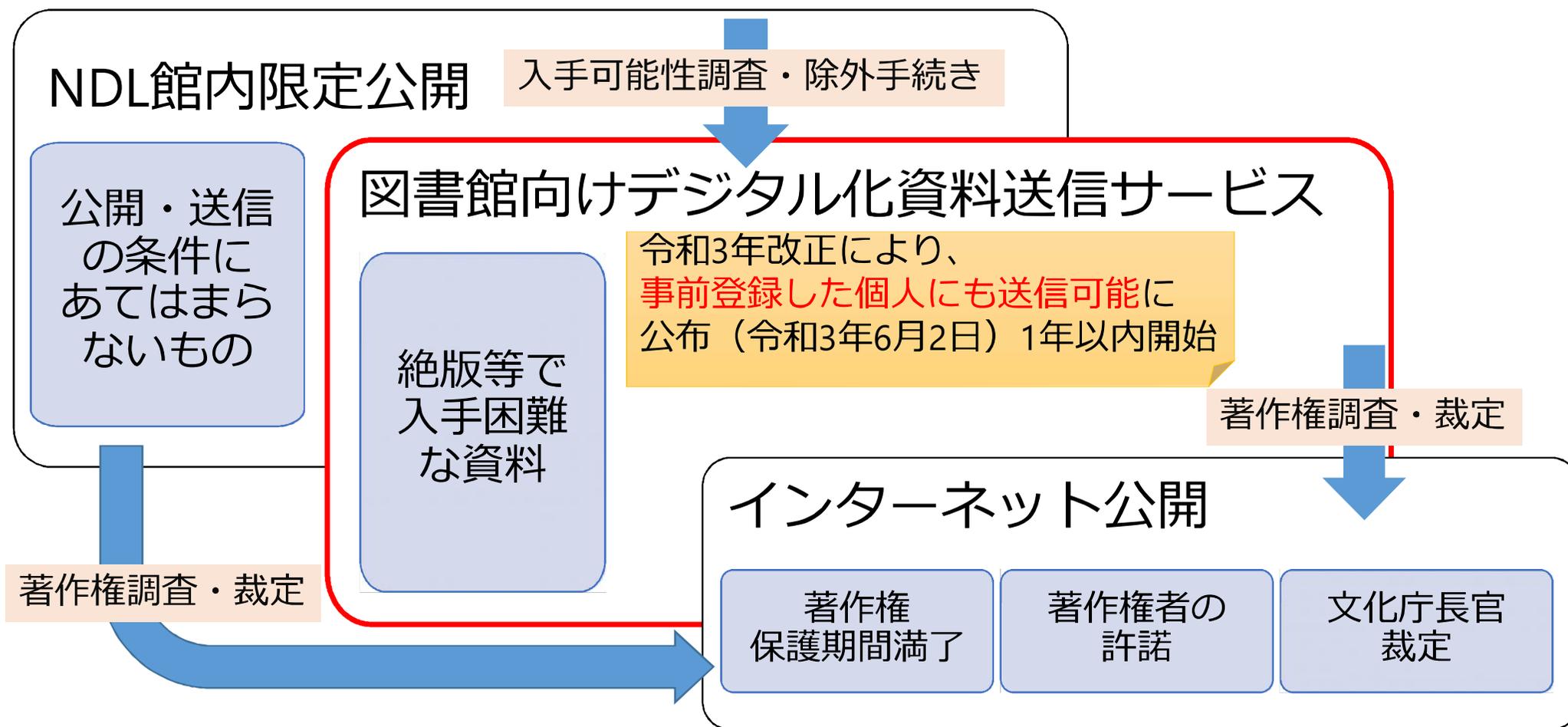
項目	概要
図書資料のデジタル化	1987年までに整理した国内刊行図書のデジタル化 約45億円 ※社会科学分野、人文科学分野の一部（大半が入手困難資料）
デジタル化設備の整備	館内で所蔵資料のデジタル化を行うためのブックスキャナ等導入
全文テキスト化の推進	デジタル化済み資料のOCRによる全文検索用のテキスト化 OCR精度向上に向けた研究開発
電子書庫機能の拡張等	ストレージ増強・国立国会図書館デジタルコレクションの改修

2021年度補正予算（第1号）

項目	概要
図書資料のデジタル化	1987年までに整理した国内刊行図書のデジタル化 約37億円 ※人文科学分野、自然科学分野の一部（大半が入手困難資料）
全文テキスト化の推進	視覚障害者向け（読み上げ用）OCR処理プログラムの研究開発
電子書庫機能の拡張等	ストレージ増強・視覚障害者向けの全文テキストデータ提供機能の拡張

⇒成果の提供は2022年から段階的に。

3. デジタル化資料の個人への送信：提供範囲



3. デジタル化資料の図書館への送信

2012年	著作権法改正（第31条第3項新設）→図書館等への絶版等入手困難な資料の送信が可能に	（2014年～）図書館送信サービス開始
2018年	同改正（第31条3項）→外国の図書館等へ絶版等入手困難な資料の送信が可能に	（2019年～）外国の図書館等に拡大
2021年	同改正（第31条3項ほか）→個人（家庭）への絶版等入手困難資料の送信が可能に	（2022年5月）個人への送信サービス開始（予定）

対象資料	<ul style="list-style-type: none"> 絶版等の理由で入手困難な資料 NDLが入手可能性調査を行い市場で流通していないことを確認 ※ただし、漫画・絵本・商業出版による雑誌・管理委託著作物等は除外 出版者・著作(権)者等の申し出により一定の除外基準に該当するものを送信対象から除外
対象施設	<ul style="list-style-type: none"> 著作権法31条に規定する「図書館等」 = 公共図書館、大学図書館、国公立博物館・美術館、国公立の研究機関の図書館 公益法人立の図書館（個別指定）、公益法人立の博物館・博物館相当施設 ※司書または司書に相当する職員の配置が必要 2019年度から外国の図書館にも送信開始（2018年著作権法改正） 関係者協議での合意に基づき、NDLによる要件確認・承認が必要
参加館数	国内1,327館 + 海外4館（2022年1月4日現在）

「国立国会図書館のデジタル化資料の図書館等への限定送信に関する合意事項」に基づき運用

3. デジタル化資料の個人への送信：概要

なぜ	コロナ禍で研究者・学生等から来館せず利用できる図書館サービスへのニーズの高まり ⇒ 出版者協議会、図書館休館対策プロジェクト、日本歴史学協会等からのデジタル化資料公開範囲拡大の要望 ⇒ 「知的財産推進計画2020」「図書館関係の権利制限規定の見直し（デジタル・ネットワーク対応）に関する報告書」
いつ	2022年5月19日開始予定（改正法公布（2021年6月）後1年以内の施行）
だれが	国立国会図書館
なにを	国立国会図書館がデジタル化した資料のうち絶版等により入手困難なもの ⇒ 図書館送信と同じ範囲（全て国立国会図書館による入手可能性調査及び出版者・権利者等による除外確認を経たもの。また、3月以内に復刊予定のものも除く） ⇒ 他の図書館から提供された、国立国会図書館未所蔵かつ入手困難なデジタル化資料の提供も可能
だれに	国立国会図書館登録利用者のうち国内在住かつ当該サービスの利用規約に同意した者 ⇒ 氏名・現住所・生年月日が確認できる身分証明書による本人確認が必要。インターネット限定登録利用者除く（要切替！）。 ⇒ 利用者登録は、従来の郵送・来館に加え、2021年5月以降はインターネットでの登録も可能に。
どのように	利用者がID・PWにより「国立国会図書館デジタルコレクション」にアクセス ⇒ 国立国会図書館オンラインへのログイン時に利用規約への同意が必要。 ⇒ 当面はストリーミング（閲覧）のみ。プリントアウトは2023年1月から開始予定
いくらで	無料

* 「国立国会図書館による入手困難資料の個人送信に関する関係者協議会」で協議

* 2021年4月設置。文化庁と国立国会図書館共催。権利者含む関係団体と有識者から構成

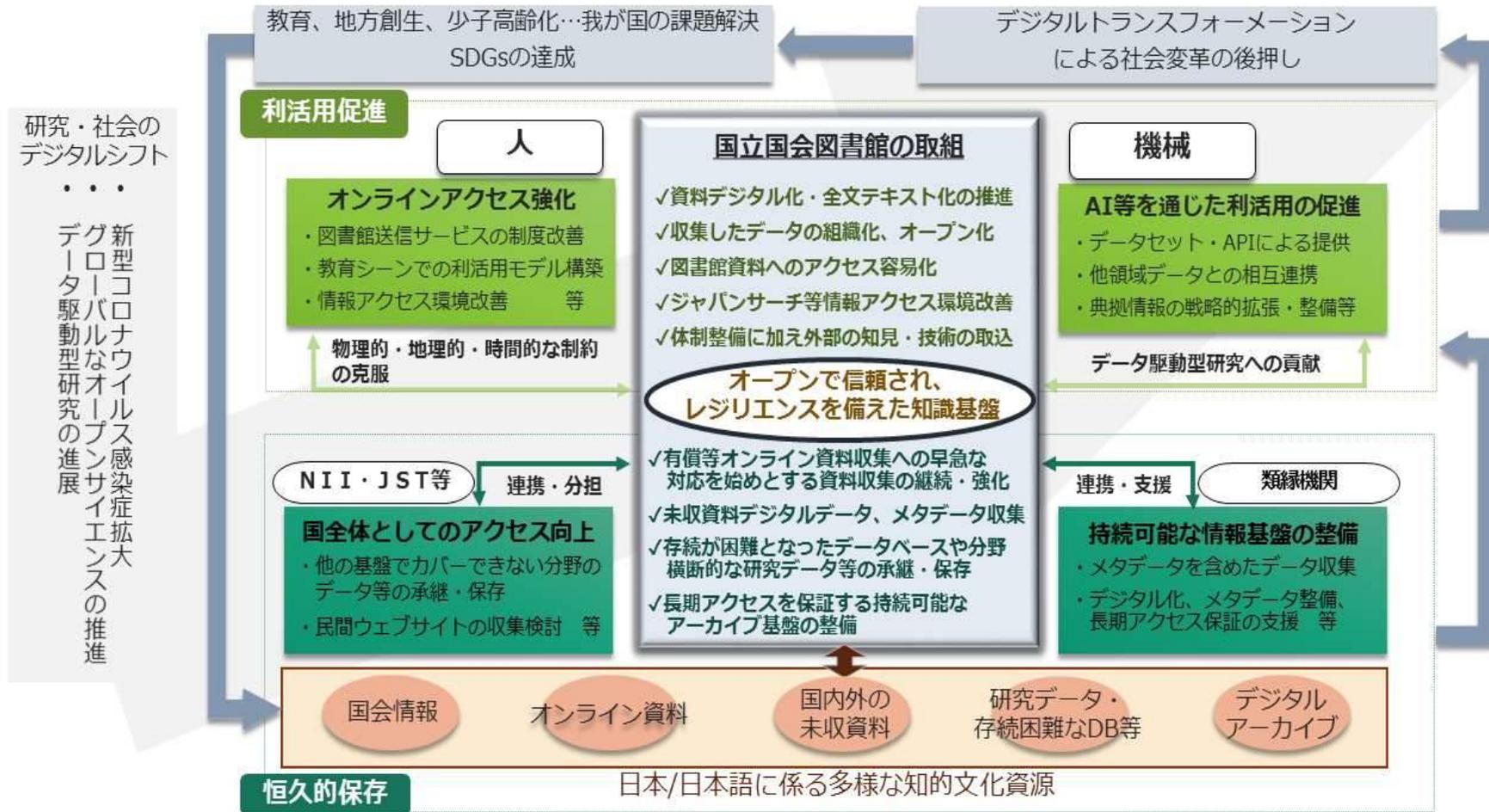
* 2021年12月に「国立国会図書館のデジタル化資料の個人送信に関する合意文書」を公表

3. デジタル化資料の個人への送信：公衆送信との違い

	公衆送信 (一部改正法第2条)	個人送信 (一部改正法第1条)
なぜ	コロナ禍で研究者・学生等から来館せず利用できる図書館サービスへのニーズの高まり	
いつ	改正法公布後2年以内の施行⇒令和5年度	改正法公布後1年以内の施行⇒令和4年5月
だれが	国内の図書館等 (国立国会図書館含む。詳細要件は今後検討) (「図書館等公衆送信サービスに関する関係者協議会」で協議)	国立国会図書館 (「国立国会図書館による入手困難資料の個人送信に関する関係者協議会」で協議)
だれに	国内在住の各図書館利用者 (詳細要件は今後検討)	国立国会図書館登録利用者のうち国内在住かつ当該サービスの利用規約に同意した者
なにを	各図書館が利用者の申込に応じ(入手困難でない)資料の一部をデジタル化したもの (権利者の利益を不当に害しないよう、対象範囲は今後検討)	国立国会図書館がデジタル化した資料のうち絶版等により入手困難なもの
どうやって	メール等によりファイルを利用者に送信 (要データ流出を防ぐ措置。詳細は今後検討)	利用者がID・PWにより「国立国会図書館デジタルコレクション」にアクセス
いくらで	実費＋補償金 (補償金の額等は今後検討)	無料

4. デジタル化資料の全文テキストデータ : 背景

「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—」の概要



4. デジタル化資料の全文テキストデータ : 取組

(2021年度の取組)

- 「国立国会図書館デジタルコレクション」搭載のほぼ全てのデジタル化資料（約2億2300万画像コマ）の全文テキスト化（検索用）
- オープンソースとして公開可能なOCR処理プログラムの研究開発

(今後の取組予定)



1. 全文検索サービス（スニペット表示も）の提供

- NDLラボ上の「次世代デジタルライブラリー」で実験後（先週、著作権保護期間満了のデジタル化資料7万点の全文検索データを追加）、「国立国会図書館デジタルコレクション」で提供（2023.1～）

2. 視覚障害者向け（読み上げ用）のOCR処理プログラムの研究開発

- 市場で流通しているものを除外のうえ、視覚障害者等用データ送信サービスから提供

3. プログラム&データセットの公開（NDLラボ/GitHub）

- CC BYでOCR処理プログラムを公開 ⇒OCRの更なる精度向上へ
- 膨大なテキストデータの提供 ⇒様々な研究活動へ

4. テキストデータの効果的な利用提供方法の調査研究

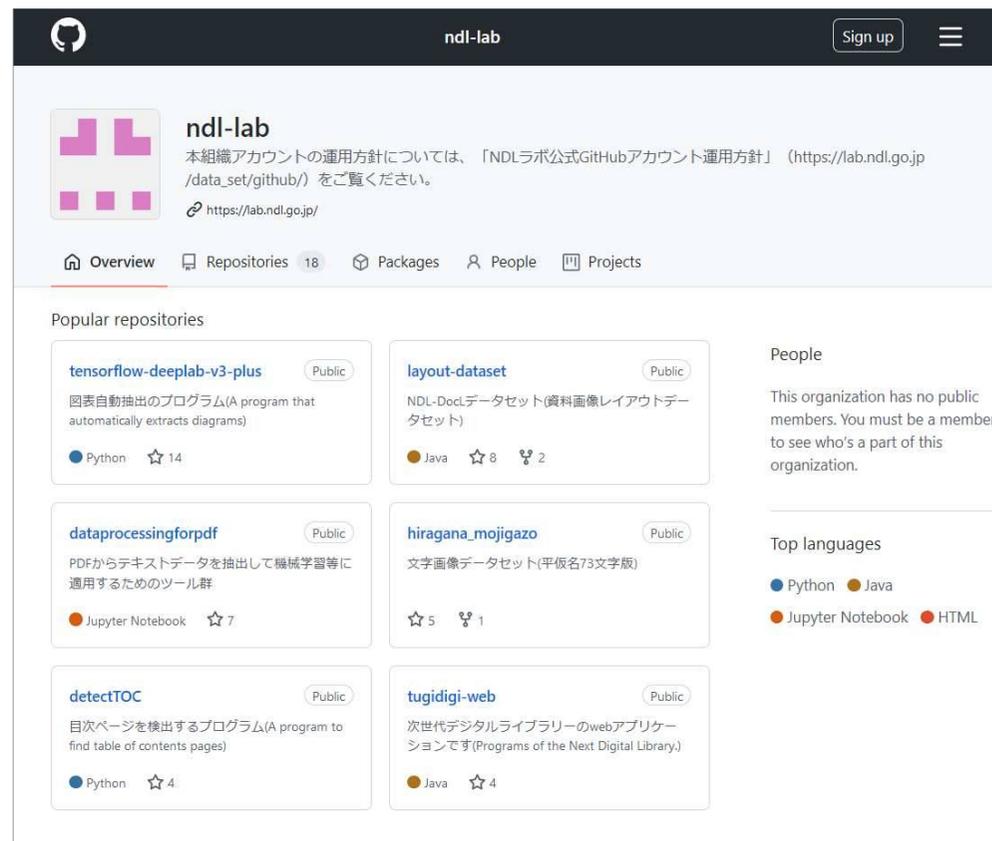
- 資料中の図版キャプションの自動抽出による「キャプション検索」、1行単位の「類似文章検索」など
⇒NDLラボ上で実験公開

4. デジタル化資料の全文テキストデータ : 参考①

(プログラム・データセットの公開)

<https://github.com/ndl-lab>

- GitHubで公開
GitHub : 世界で最もよく使われているオープンデータ・プログラムの公開・共有サイト
- 2019年8月 アカウント開設
- 次世代室が調査研究中の技術情報を公開
- プログラムはCC BY、著作権保護期間満了のデジタル化資料から作成したデータはパブリックドメインマーク (PDM) で提供
- 外部の優秀なエンジニアとの技術的な検討の場の創出へ



The screenshot shows the GitHub profile page for the organization 'ndl-lab'. The header includes the GitHub logo, the organization name 'ndl-lab', and a 'Sign up' button. Below the header, there is a bio section with a profile picture and a link to the organization's website. The main content area is divided into several sections: 'Overview', 'Repositories' (18), 'Packages', 'People', and 'Projects'. The 'Popular repositories' section lists several public repositories, including 'tensorflow-deeplab-v3-plus', 'layout-dataset', 'dataprocessingforpdf', 'hiragana_mojigazo', 'detectTOC', and 'tugidigi-web'. Each repository card shows the repository name, a 'Public' badge, a description, the programming language, and the number of stars and forks. The 'People' section on the right indicates that the organization has no public members. The 'Top languages' section shows the most used languages: Python, Java, Jupyter Notebook, and HTML.

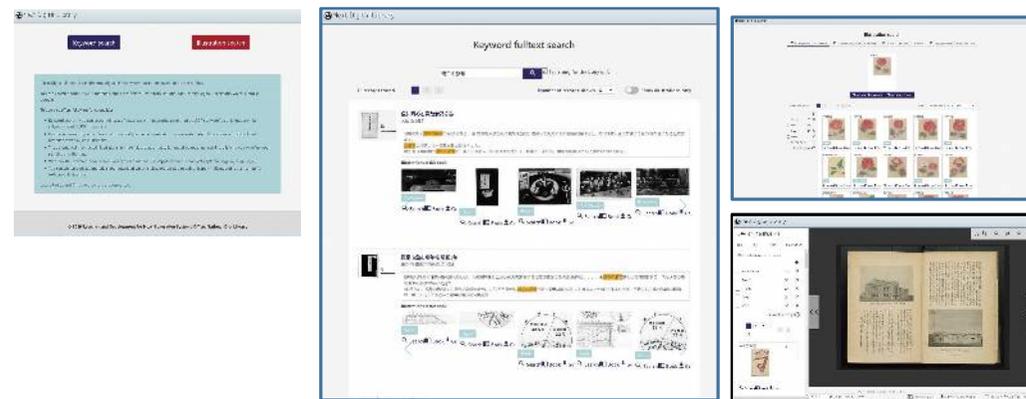
4. デジタル化資料の全文テキストデータ：参考②

(公開しているプログラム・データセットの例)

(次世代デジタルライブラリー)

● プログラム

- 次世代デジタルライブラリーのソースコード
- 図版の自動抽出プログラム
- 画像検索インデックス作成プログラム
- 見開きページをのど元で分割するプログラム



● データセット

- NDL-DocLデータセット (資料画像レイアウトデータセット)
- 文字画像データセット (漢字300文字版)
- 文字画像データセット (平仮名73文字版)
- 次世代デジタルライブラリーで本文検索可能なテキストの全データ
- 全国書誌データから作成した振り仮名のデータセット

<https://lab.ndl.go.jp/dl/>

- 2019年3月公開、2020年3月リニューアル
- 検索対象：
国立国会図書館デジタルコレクション(<https://dl.ndl.go.jp/>)でウェブ公開している著作権保護期間満了の図書・古典籍資料、約33万6千点
【本文テキスト】 日本十進分類法(NDC)6類 (産業) のみ
【資料中の図版】 全て
- 本実験サービスにある各種機能は、国立国会図書館デジタルコレクション等での実装を目指す
- Chrome、Firefox推奨

5. デジタル資料の長期保存

◆ 「国立国会図書館デジタル資料長期保存基本計画2021－2025」 2021年3月策定

媒体のせい弱性、再生装置の入手困難化、再生ソフトウェア等技術の陳腐化
⇒デジタル資料も、長期利用のためには、紙資料と同様、維持管理が必要

【最近の取組】

- パッケージ系電子出版物（CD/DVD、USBメモリ、フロッピーディスク等）のマイグレーション（主に媒体変換）による保存対策の本格実施 **2021年度～**
- マイグレーション後データの長期利用保証のためのエミュレーション技術等調査 **2020年度～**
- デジタル化資料の保存用データ（光ディスク約19万枚等）の管理コスト削減のため、LTOへの媒体移行の試行 **2021年度～**
なお、マイグレーション後データは次期デジタルデポジットシステム（2023年1月リリース）の電子書庫（クラウド利用）へ
- 国内機関におけるデジタル資料の保存に関する実態調査の実施 **2021年度**

6. ジャパンサーチ : 位置づけ

NDLの収集対象

- 国内出版物**
図書、雑誌、新聞、パ系等
※納本制度により収集、購入や寄贈による収集もあり
- 外国出版物**
図書、雑誌等
※主に購入、国際交換等により収集
- オンライン資料**
電子書籍・電子雑誌等
※オンライン資料収集制度により民間発行のオンライン資料を収集、任意提供による収集もあり
- インターネット資料**
ウェブサイト
※インターネット資料収集保存事業により公的機関のウェブサイトを収集、民間のウェブサイトは許諾に基づき収集

NDLのアーカイブ



NDLの所在情報検索

国立国会図書館サーチ

国立国会図書館オンライン

当館所蔵資料（デジタル化資料含む）、オンライン資料等を検索

NDLのアーカイブの他、全国の公共図書館、学術研究機関等が提供する資料、デジタルコンテンツを統合的に検索

書籍等分野の情報を提供

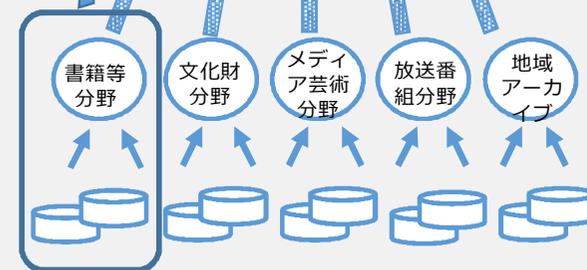
NDL以外のアーカイブ

学術研究機関/民間のリポジトリ、公共図書館蔵書/デジタルコンテンツ

国の分野横断統合検索

ジャパンサーチ

日本が保有する多様な分野のコンテンツの所在情報を提供し、オープンに利用可能なデジタルコンテンツを検索



国内のアーカイブ

図書館、博物館、美術館、文書館、企業、官公庁、地方公共団体、文化施設が保有する情報

6. ジャパンサーチ：概況

● デジタルアーカイブの検索・閲覧・活用基盤

さまざまな分野のデジタルアーカイブと連携し、我が国の多様なコンテンツの**メタデータ**をまとめて検索・閲覧・活用

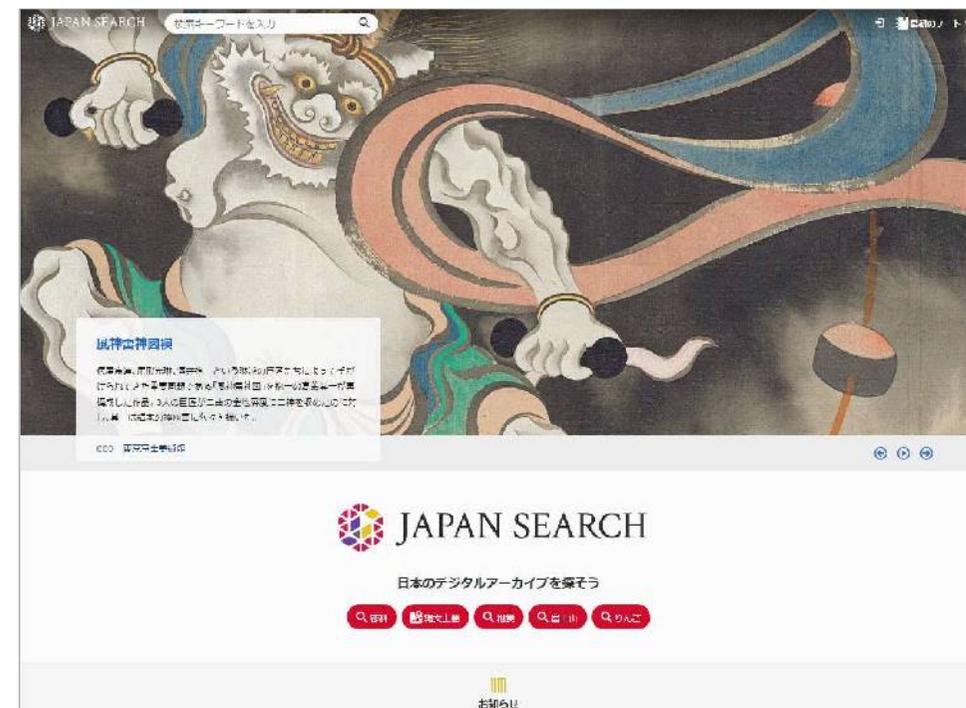
- **30**連携（つなぎ役）機関から、**148**データベース・メタデータ約**2,360万件**を検索可能（2022年1月現在）

● 国全体の取組

運用主体：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（事務局：内閣府）
 システムの開発・運用担当：国立国会図書館

● 実験的な技術の取込

類似のサムネイル画像の検索
 ローマ字読みの自動生成



ジャパンサーチトップ画面
<https://jpsearch.go.jp/>

6. ジャパンサーチ：戦略方針2021-2025



JAPAN SEARCH

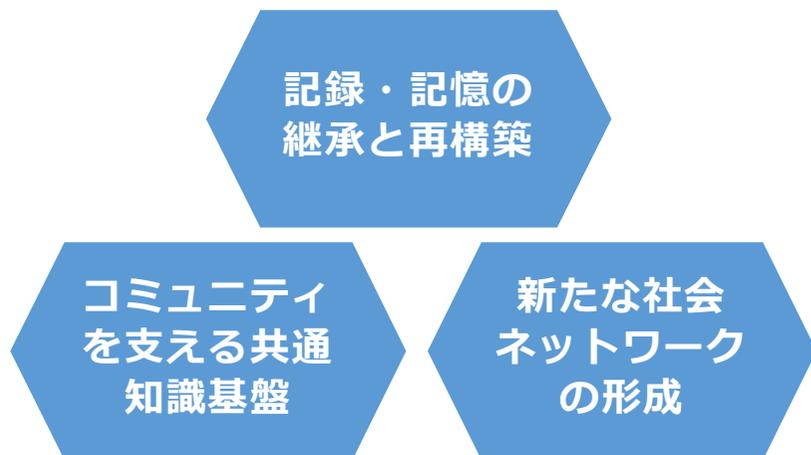
「デジタルアーカイブを日常にする」

ミッション

新しい情報技術とアーカイブ連携を通じて、日本の文化的・学術的コンテンツの発見可能性を高め、それらを活用しやすい基盤を提供することで、デジタルアーカイブが日常に溶け込んだ豊かな創造的社會を実現します

3つの価値：

デジタルアーカイブの大切な役割



4つのアクション：

ジャパンサーチを使った活動の柱

支える	学び・遊びの体験、キュレーション活動、ひとも繋ぐ、DAを長く
伝える	DAを集約・整理、DA活用方法、DA構築方法、コンテンツの魅力
拡げる	デジタル化の推進、ファン層、魅力あるDA、使いやすいコンテンツ
挑む	オープン化、新たな可能性、新たなネットワーク、社会のDX化