



国立国会図書館が取り組む 資料デジタル化及びテキスト化の全体像

Overall picture of digitization and text
conversion at the National Diet Library

国立国会図書館 電子情報部 主任司書

電子情報企画課資料デジタル化推進室長 井上 佐知子

Sachiko Inoue, Digitization Promotion Office Director

Digital Information Department, National Diet Library





国立国会図書館

1948年開館

約4700万点の資料を所蔵

National Diet Library (NDL)

Opened 1948

Holding approximately
47 million items

国立国会図書館ビジョン2021-2025

NDL Vision 2021-2025

重点事業 / Seven initiatives

1. 国会サービスの充実
2. インターネット提供資料の拡充
3. 読書バリアフリーの推進
4. 「知りたい」を支援する情報発信

5. 資料デジタル化の加速

6. デジタル資料の収集と長期保存
7. デジタルアーカイブの推進と利活用

https://www.ndl.go.jp/jp/aboutus/vision_ndl.html
https://www.ndl.go.jp/en/aboutus/vision_ndl.html

5. 資料デジタル化の加速

デジタルで全ての国内出版物が読める未来を目指し、この5年間で100万冊以上の所蔵資料をデジタル化します。テキスト化も行い、検索や機械学習に活かせる基盤データとします。

5. Accelerating digitization of materials

Over the next five years, the NDL will digitize more than a million library materials to achieve our vision of making all domestic publications available in digital form. We will convert them to full-text data, making them searchable and suitable for processing with machine learning.

デジタル化とテキスト化の最新状況

Latest status of digitization and text conversion

資料群 Material type	インターネット公開 Available Online	送信サービス For Registered Users	館内限定公開 Only at the NDL	合計 Total
図書 / Books	370,000	1,040,000	650,000	2,060,000
雑誌 / Periodicals	20,000	830,000	530,000	1,370,000
古典籍 / Rare Books and Old Materials	80,000	20,000	3,000	100,000
博士論文 / Doctoral Dissertations	10,000	140,000	20,000	180,000
新聞 / Newspapers	-	-	170,000	170,000
その他 / Others	150,000	30,000	120,000	300,000
合計 / Total	630,000	2,050,000	1,500,000	4,180,000

上記のうち、全文検索が可能な資料

Of the above, full-text searchable materials

2,710,000

(Round numbers as of July 2024)



国立国会図書館デジタルコレクション

N D L D I G I T A L C O L L E C T I O N S

📄 ログインなしでも閲覧可能

📄 送付サービスでも閲覧可能

📄 国立国会図書館内限定



キーワード:

検索

たとえば... [三輪由富治](#) [デザイン](#) [かごさとし](#) [書文一致体](#) [カット集](#)

[画像検索](#)

[詳細検索](#)

すべてのスポットライト

デジタル化資料

オンライン資料

便利ツール

お知らせ

お知らせ



Cookieの同意画面を設置しました
Cookieにつきましては、「NDLデジタル
コレクションプライバシーポリシー」の
ページや「よくあるご質問 17・18」
をご参照ください。

お知らせ



資料追加のお知らせ(2024-06-21)
図書約89,000点を「国立国会図書館デジ
タルコレクション」に追加しました。

お知らせ



資料追加のお知らせ(2024-06-18)
新聞約54,000点を「国立国会図書館デジ
タルコレクション」に追加しました。



翻車考 1冊 寄別11-29 00-008

国立国会図書館



コンテンツ	目次	画像調整	全文検索
4	5	6	
7	8	9	
10	11	12	
13	14	15	
16	17	18	

?
🔖 コマ番号 **9** /18
 📄 右から左に進む
 📄 1コマずつ表示
 ➖ 17% ➕

🔖 翻車考

📖 書誌情報 : 著者 栗丹洲<栗本丹州> // [著] 出版者 写

📖 書誌情報

📄 永続的識別子 info:ndljp/pid/1286944

📄 タイトル 翻車考

📄 国立国会図書館サーチの書誌へ

🖨️ 印刷

共有



📄 この資料のURL

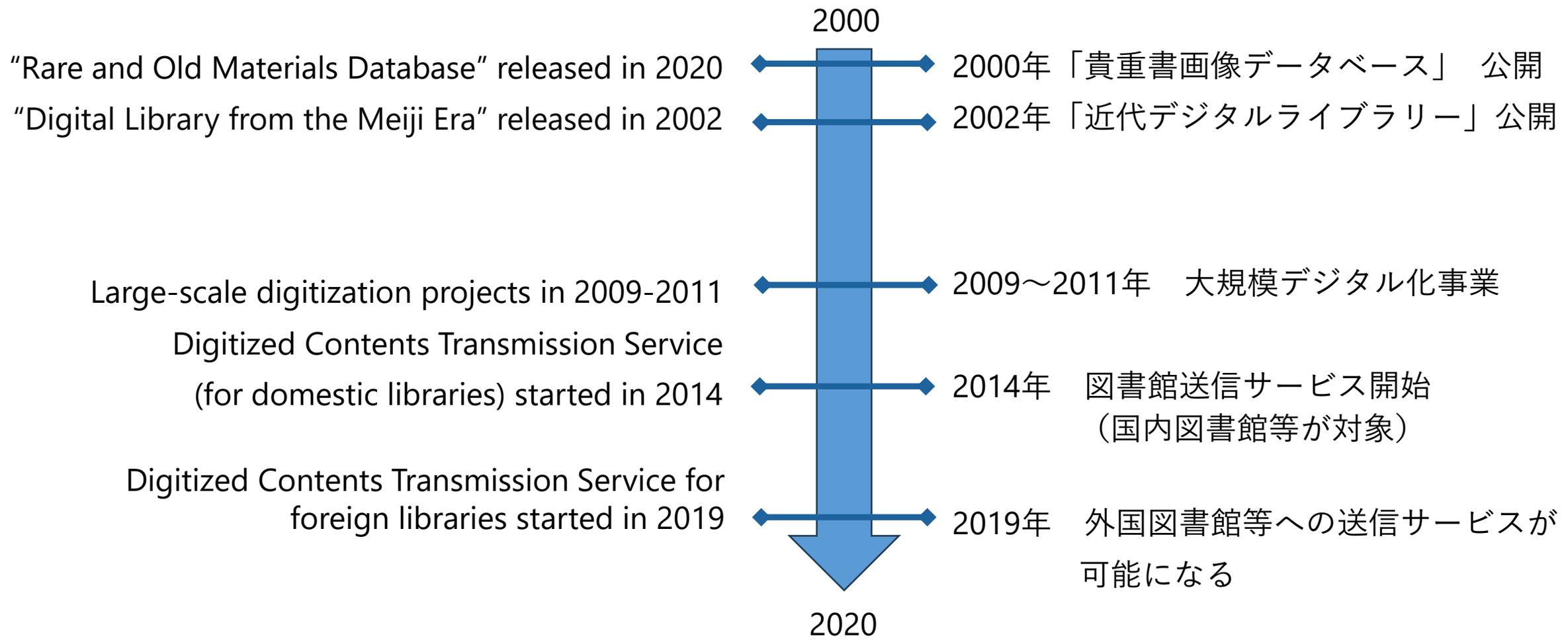
目次 Content

1. 2020年までの実績
2. 2021年以降の取り組み
3. 課題と今後の展望

1. Achievements through 2020
2. Initiatives since 2021
3. Challenges and future prospects

2020までの実績①

What the NDL has done up until 2020 ①

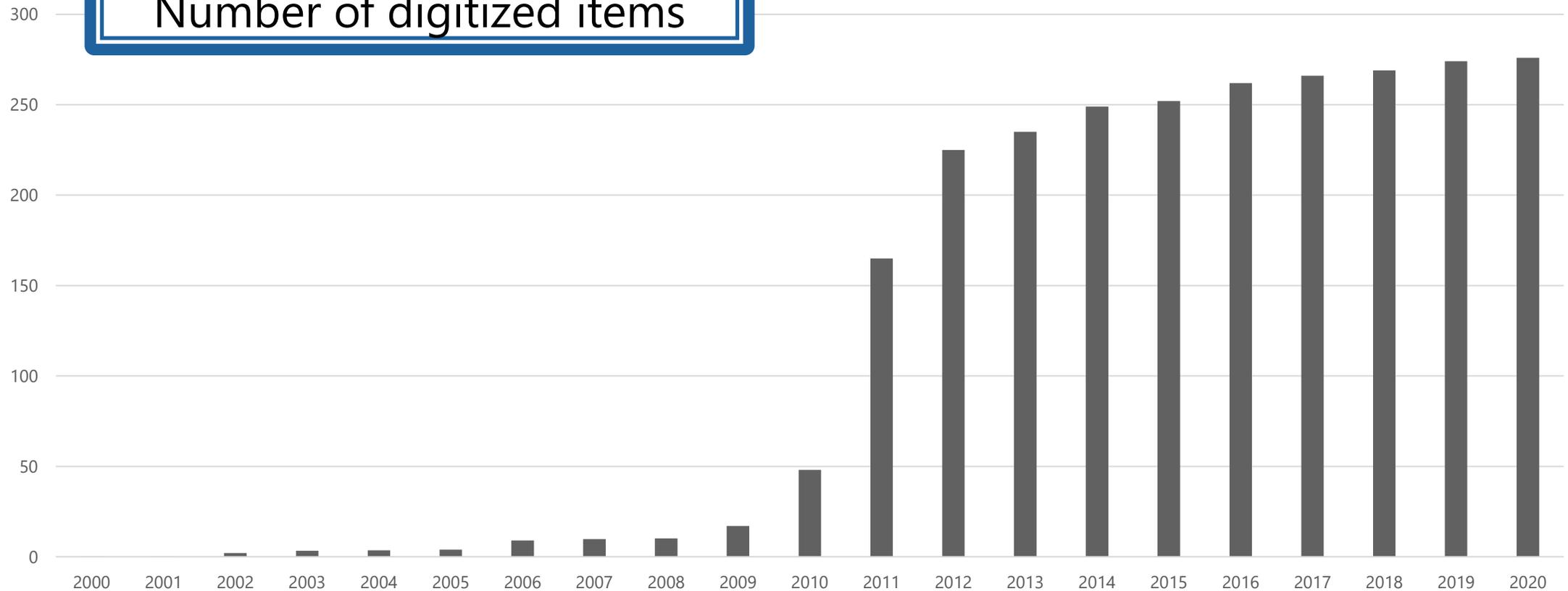


2020までの実績②

What the NDL has done up until 2020 ②

万件
10,000s of items

デジタル化資料提供点数
Number of digitized items



2020年に何があったか？ What happened in 2020?

新型コロナウイルス流行
COVID-19 pandemic



図書館は閉館
Libraries were closed



資料にアクセスできない
Unable to access materials

インターネット経由での資料閲覧が求められる
Need to read materials via the internet



著作権法改正（2021年）

Amendments to the Copyright Act (2021)

- 国立国会図書館から個人への絶版等資料のインターネット送信
Internet transmission of out-of-print and other materials from the NDL to individuals
- 図書館等による図書館資料のメール送信等
Transmission of library materials by libraries

資料デジタル化基本計画2021-2025①

Basic Plan for Digitization of NDL Collections 2021-2025 ①

デジタル化の優先順位

- 唯一性・希少性
- 資料の利用機会の拡大
- 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

Priority for digitization

- Uniqueness and rarity
- Expanding access to materials
- Deterioration status of materials, urgency of preservation
- Social and academic needs for digitization
- Contribution to the construction of national and global systematic digital collections

資料デジタル化基本計画2021-2025②

Basic Plan for Digitization of NDL Collections 2021-2025 ②

デジタル化の対象範囲

国内で刊行された以下の資料

- 2000年までに刊行された図書
約170万点
- 刊行後5年以上経過した雑誌
- その他の資料

古典籍資料、録音・映像資料、博士論文、憲政資料、日本占領関係資料、日本移民関係資料、地図資料、新聞資料

Scope of digitization

The following materials published in Japan

- Books published up to 2000
Approximately 1.7 million items
- Journal volumes published more than 5 years ago
- Other materials

Classical materials, audio/visual materials, doctoral dissertations, constitutional documents, materials related to the occupation of Japan, materials related to Japanese immigration, maps, newspapers

国内刊行図書デジタル化の進捗状況

(2024年7月時点の推計値)

Progress in digitization of books published in Japan

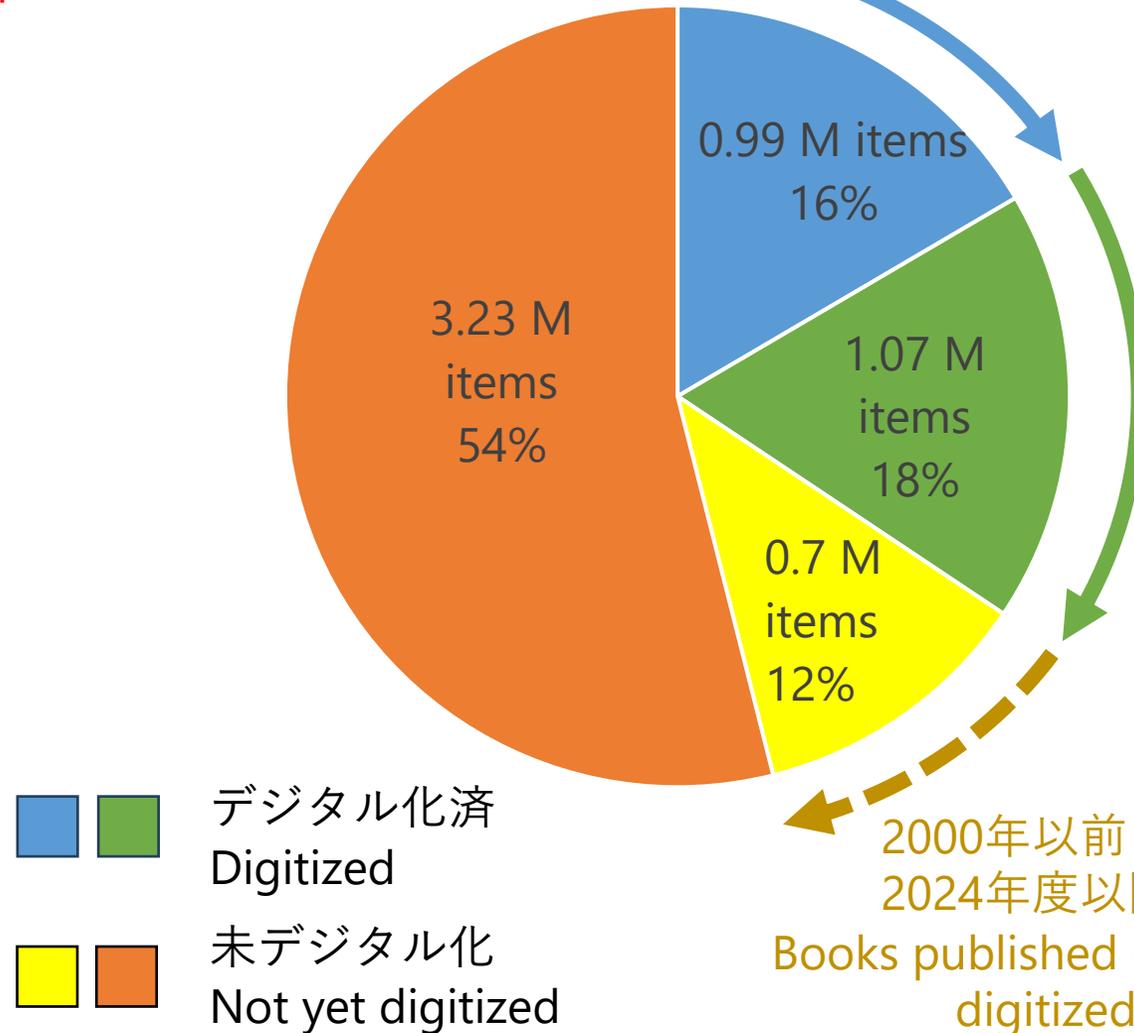
(Estimated as of July 2024)

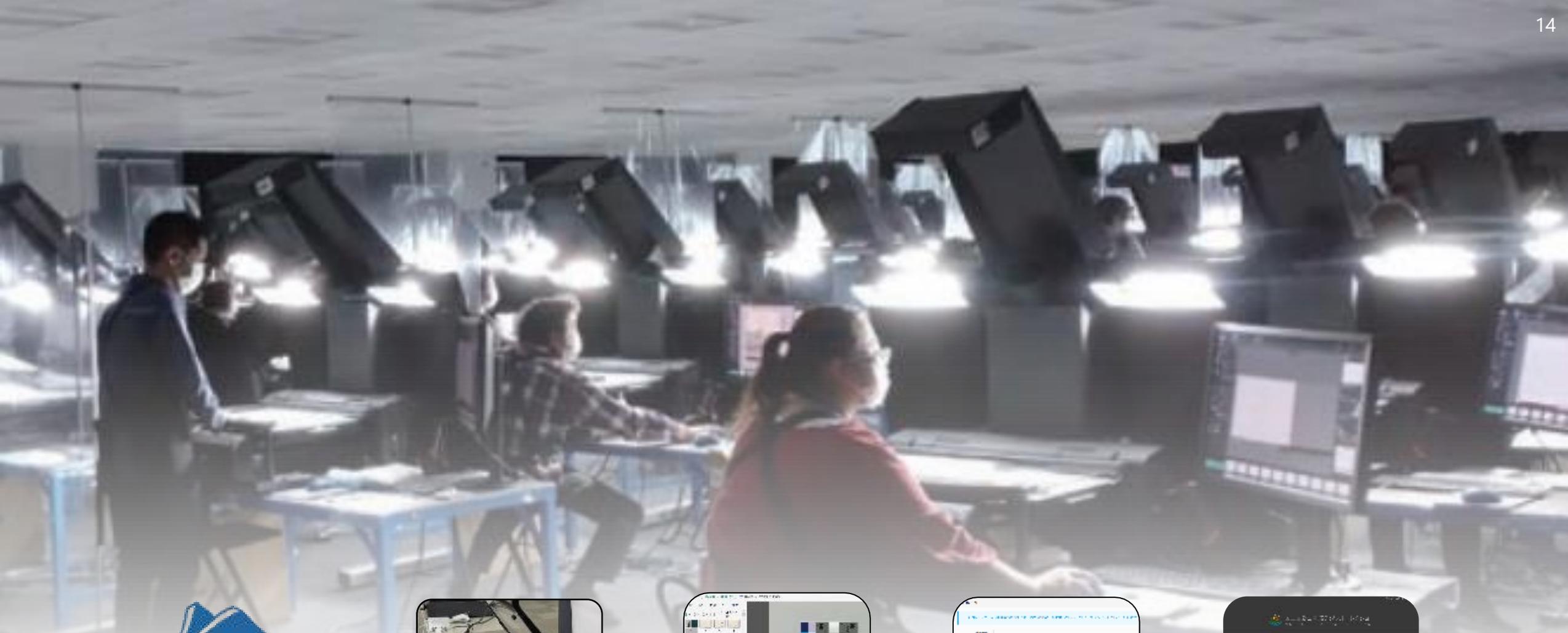
残りは主に2001年以降に
刊行された図書
The rest are mainly books
published after 2001

1968年までに刊行された図書は2020年度
までにデジタル化済
Books published up until 1968 were
digitized by FY2020

1995年以前に刊行
された図書を
2023年度までに
デジタル化
Books published
up until 1995 were
digitized by
FY2023

2000年以前に刊行された図書は
2024年度以降にデジタル化予定
Books published up until 2000 will be
digitized in and after FY2024

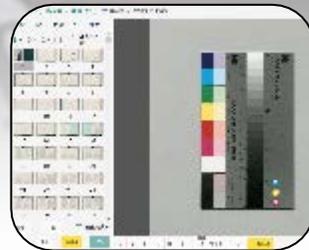




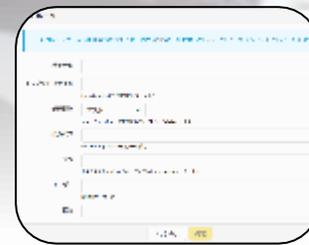
①資料選定
Select a material



②スキャン
Scanning



③品質検査
Quality check



④データ登録
Data entry



⑤公開
Release

資料デジタル化のコスト

Cost of digitization

本一冊を（一般的な形態の原本から）デジタル化する場合...
When digitizing a book (a standard book, from the original) ...

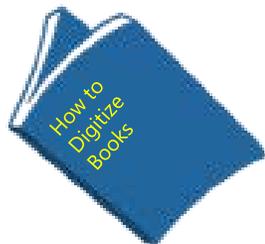


見開き（2ページ）を一コマとして撮影する

A spread (2 pages) is taken as one frame

（2023年度実績） 1コマ撮影単価は 60～80 円

（In FY2023） Cost of 1 frame 0.43 US\$ - 0.57 US\$



本一冊のデジタル化費用はページ数に応じて変わる

The cost of digitizing a book depends on the number of pages

100 pages	50 frames	3,500円 / 25\$	
200 pages	100 frames	7,000円 / 50\$	
400 pages	200 frames	14,000円 / 100\$	(140円 ≒ 1US\$ in 2023)

デジタル化資料の公開範囲

Scope of public access to digitized materials

出版後の時間経過
Time lapse from publishing



出版
Published

品切れ
Out-of-stock

絶版
Out-of-print

著作権保護期間満了
Expiration of copyright protection period

公開範囲
Scope of public access

入手可能
Available in the market

入手困難
Difficult-to-obtain

パブリックドメイン
Public domain

NDL館内での閲覧 In the NDL building only	○	○	
送信サービス For transmission services	×	↔	○
インターネット公開 Available online	△ 許諾があれば可能 Possible with permission	△ 許諾があれば可能 Possible with permission	○

「除外手続」で決定
Determined by
"Exclusion Procedure."

除外手続

“Exclusion Procedures”

データベース等との突合による入手可能性調査
Availability check using databases of publications available in Japan

送信候補資料リストの公開と除外申出の受付
Publicize the list of candidate materials for transmission and accept requests for exclusion

送信開始後も除外申出の受付を継続
Exclusion requests are accepted even after transmission begins

除外基準

- ❑ 当該資料又は同内容の著作物が市場で流通している、または流通予定がある。
- ❑ 当該資料又は同内容の著作物の著作権が著作権等管理事業者により管理されている。
- ❑ 当該資料の著作者から送信利用の停止の要請がある。

Exclusion Criteria

- ❑ The material in question or other material with identical content is currently commercially available or is scheduled to be published.
- ❑ The material in question or other material with identical content is copyrighted and managed by a copyright management organization.
- ❑ The author has requested that the material not be made available via the Service.

関係者協議

Consultation with stakeholders

- ◆資料のデジタル化とそれによるサービスは、著作者や出版者等の利益を不当に害することがないように留意して実施
- ◆2008年から著作権者・出版者団体、図書館等の関係団体や機関の協議を継続的に実施。
- ◆デジタル化対象資料の選定とそれを用いたサービス実施に関する出版社等との調整手続を合意事項として定める。
- ◆Digitization of materials and related services shall be conducted with care so as not to unreasonably harm the interests of authors and publishers, etc.
- ◆Since 2008, the NDL has conducted ongoing consultations with copyright holders, publishers' organizations, libraries, and other related groups and institutions.
- ◆Agreements define the procedures for coordination with publishers, etc. regarding the selection of materials to be digitized and the implementation of services using such materials.

資料のテキストデータを用いたサービス Services using text data of materials

OCRによるテキスト化
Text extraction by OCR

デジタル化画像 テキストデータ
Digitized images Text data



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的確につなげることがために、今後

2024年からNDLが開発した
NDLOCRを利用
Since 2024, using NDLOCR
developed by the NDL.



NDL Digital Collection

デジタル化資料の全文検索
Full-text search of digitized materials



NDL Material Search
for Persons with Disabilities
(Mina Search)

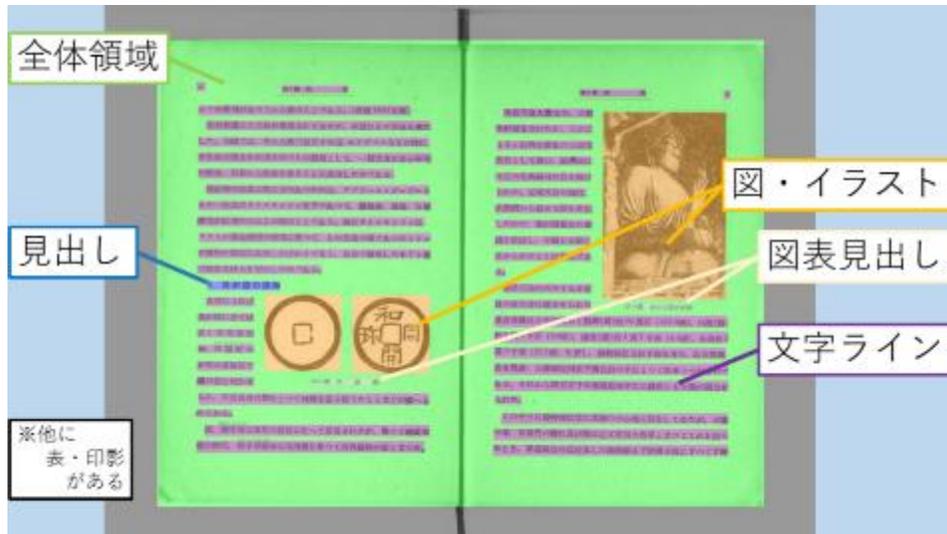
視覚障害者等用データ送信サービス
Text download service for those who are print disabled

NDLOCRの研究開発

Research and development of NDLOCR

資料のOCRテキスト化に使用するAI-OCR処理プログラムを開発

Developed AI-OCR processing program used to convert materials to OCR text



レイアウト認識の例

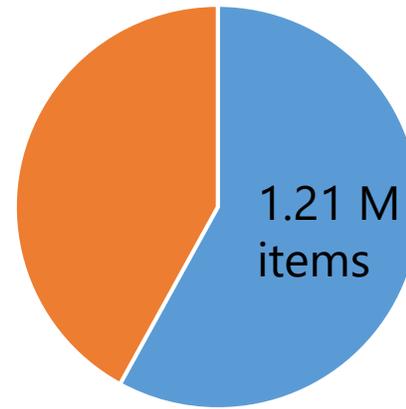
Example of layout recognition

- ◆NDLの資料を用いてレイアウト認識プログラムとそのデータセット（NDL-DocL）を作成
- ◆OCR処理を行う前段階でレイアウト認識による画像処理を行うことで、読み取り精度向上
- ◆ソースコードと、学習用データセットはGitHubで公開
- ◆Set up a layout recognition program and data set (NDL-DocL) using NDL materials
- ◆Improved reading accuracy by image processing with layout recognition as a preliminary step to OCR processing
- ◆Source code and training datasets are available on GitHub

NDLデジタルコレクション
で全文検索可能な資料
の割合 (2024.7)

Percentage of full-text
searchable materials on
the NDL Digital
Collection (July 2024)

図書 Books
2.06 M items



■ 全文検索可能
Full-text search available

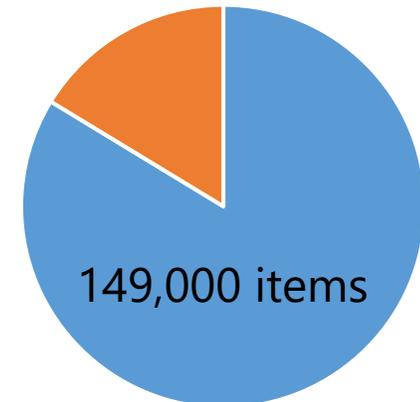
■ 全文検索未対応
Full-text search NOT
available

雑誌 Periodicals
1.37 M items



博士論文

Doctoral Dissertations
178,000 items



NDL障害者用資料検索：みなサーチ

NDL Material Search for Persons with Disabilities: Mina Search

みなサーチ
国立国会図書館障害者用資料検索

本の検索 すべてを絞り込む検索します

校正済のデータに絞る ダウンロードできるデータに絞る

いろいろな検索

詳細検索
細かい条件で絞り込んで検索

全文検索
本の内容でキーワードで検索

ジャンル検索
題になるジャンルで検索

最近1か月の新着(みなサーチでダウンロードできる資料)

電子の新着 > DAISYの新着 > テキストデータの新着 >

資料形態ごとの一覧(みなサーチでダウンロードできる資料)

音声DAISYの一覧 > マルチメディアDAISYの一覧 > テキストDAISYの一覧 >

電子書籍 (EPUB) の一覧 > 音声化された電子書籍 (EPUB) の一覧 > 透明テキストのPDFの一覧 >

Word (DOCX) の一覧 > フレーシアテキストの一覧 > 音声データの一覧 >

みなサーチとは

みなサーチは、目の見えにくい方・見えにくい方、活字の認識を助けるのが難しい方など、さまざまな障害のある方が、利用しやすい形式の資料を探し出すことができるサービスです。

みなサーチで検索できるさまざまな資料

点字、DAISY、テキストデータ、大活字本、リブレット、電子書籍、パブリックドメイン資料など、いろいろな種類の資料が検索できます。

- ◆ 「国立国会図書館ビジョン2021-2025」の重点事業3「読書バリアフリー」の一環
- ◆ 視覚障害その他の理由で通常の活字の印刷物の読書が困難な方が対象
- ◆ DAISYや他機関から収集した資料も合わせて258万点（2024.7）をダウンロード可能
- ◆ Facilitating access to materials for those with print disabilities (Initiative No. 3, NDL Vision 2021–2025)
- ◆ For those who are blind, visually impaired, dyslexic, or otherwise print disabled.
- ◆ 2.58 million items available for download, including DAISY and materials collected from other institutions

実験的なサービス Experimental search service

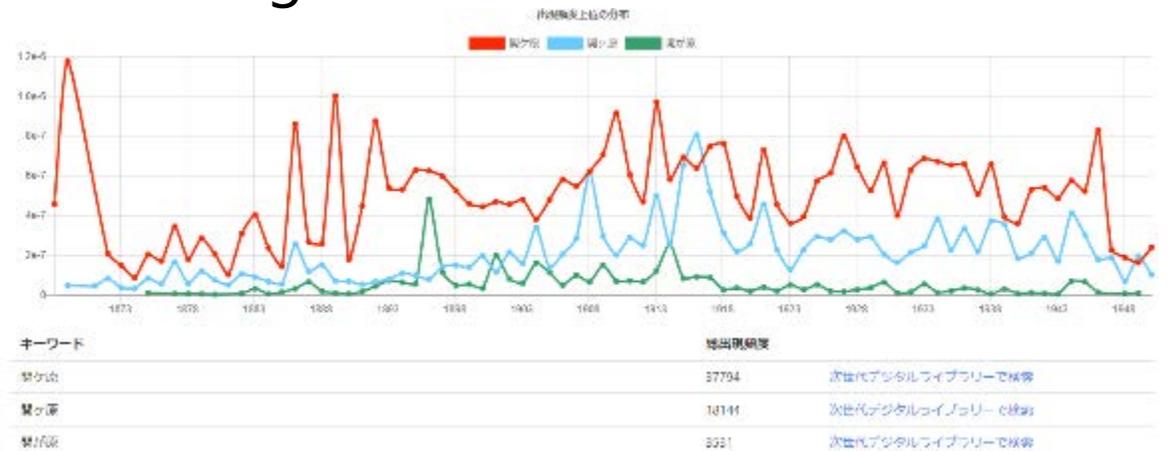
次世代デジタルライブラリー Next Generation Digital Library



著作権保護期間満了の図書28万点、古典籍資料8万点の全文検索とテキストデータのダウンロードが可能

Full-text search and text data download of 280,000 books and 80,000 classical materials whose copyright protection period has expired

NDL Ngram Viewer



デジタル化資料の本文中に特定の検索語が表れる頻度を列挙し、刊行年の時系列で可視化できるツール

A tool that lists the frequency with which specific search terms appear in the text of digitized materials and visualizes them in chronological order by year of publication.

今後に向けて考えなくてはいけないこと

Things to consider for the future

- 今後も大規模なデジタル化を継続するのか？
- デジタルで全ての国内出版物が読める未来を目指す過程で、NDLが果たすべき役割は？
- Will we continue with such large-scale digitization in the future?
- What role should the NDL play in the process of creating a future where all domestic publications are available digitally?

今後の展望－期待されること Future outlook – Expectations

- ◆オンラインで出版されNDLに電子的に納入される資料が増え、NDLがデジタル化する必要がなくなる
- ◆原本の出版者が自らデジタル化・テキスト化し、インターネット上で（商業的な）サービスとして提供する
- ◆大学図書館や地域の公共図書館、博物館、民間企業等様々な組織が、それぞれ特色あるコンテンツをデジタル化するようになる
- ◆More materials will be published online and delivered electronically, eliminating the need for the NDL to digitize them.
- ◆Original publishers will digitize and convert text themselves and offer it as a (commercial) service on the internet.
- ◆A variety of organizations, including university libraries, local public libraries, museums and private companies will digitize their own unique content.

今後の展望－これからできること

Future outlook－ What we can do going forward

- ◆資料デジタル化・テキスト化を継続し、送信サービスによる入手困難資料の提供を着実に実施する。
- ◆NDLのデジタル化資料だけでなく他組織が提供するデジタル資源に利用者を適切に誘導する。
- ◆他組織が提供するデジタル資源についても、永続的なアクセスを確保するための取り組みに協力する。
- ◆Continue digitization and text conversion of materials and steadily provide hard-to-obtain materials through the transmission service.
- ◆Guide users not only to the NDL's digitized materials, but also to digital resources provided by other organizations.
- ◆Cooperate with efforts to ensure permanent access to digital resources provided by other organizations.

ご清聴ありがとうございました。
Thank you for your attention.



この資料（国立国会図書館のロゴを除く。）はクリエイティブ・コモンズ 表示 4.0 国際 ライセンスの下に提供されています。