



国立国会図書館における データ等のオープン化の取組

国立国会図書館 電子情報部 電子情報流通課



国立国会図書館
National Diet Library, Japan

国立国会図書館(NDL)のオープン化の取組

オープンなコンテンツ・データ等の提供

- オープンに利用※できるコンテンツ(パブリックドメイン)

所蔵資料のデジタル化 → 著作権処理・調査

→ 著作権保護期間満了資料のデジタル化画像・テキストデータのオープン化

※著作者の人格的利益やプライバシーその他の人権等への配慮は必要

- オープンなライセンスで提供しているメタデータ

- オープンソースプログラム、学習用データセット 等



国立国会図書館デジタルコレクション

- NDLが収集・保存しているデジタル資料を検索・閲覧できる
- 著作権保護期間の満了した(パブリックドメイン)資料は許諾不要で自由に利用可能
- 個々の資料閲覧画面上には、書誌情報とあわせて公開範囲、転載に関する情報を表示



国立国会図書館デジタルコレクション <https://dl.ndl.go.jp/>



国立国会図書館デジタルコレクションの公開範囲

公開範囲	資料種別
インターネット公開(保護期間満了)	著作権保護期間満了が確認できた資料、法令や通達等の著作権のない著作物、著作物でないもの →許諾なしで自由に利用可能<オープンなもの>
インターネット公開(許諾)	著作権者の許諾によりインターネット上で公開している資料
インターネット公開(裁定) ※ 著作権法第67条1項による	著作権の状況が不明等の理由により文化庁長官の裁定を受けてインターネット上で公開している資料
国立国会図書館内/図書館・個人送信限定	著作権保護期間内であるが、絶版等の理由により入手困難であり、図書館等への送信が可能な資料
国立国会図書館内限定	上記にあてはまらない資料



参考:国立国会図書館ウェブサイトからのコンテンツの転載 <https://www.ndl.go.jp/use/reproduction/>

書誌データの提供

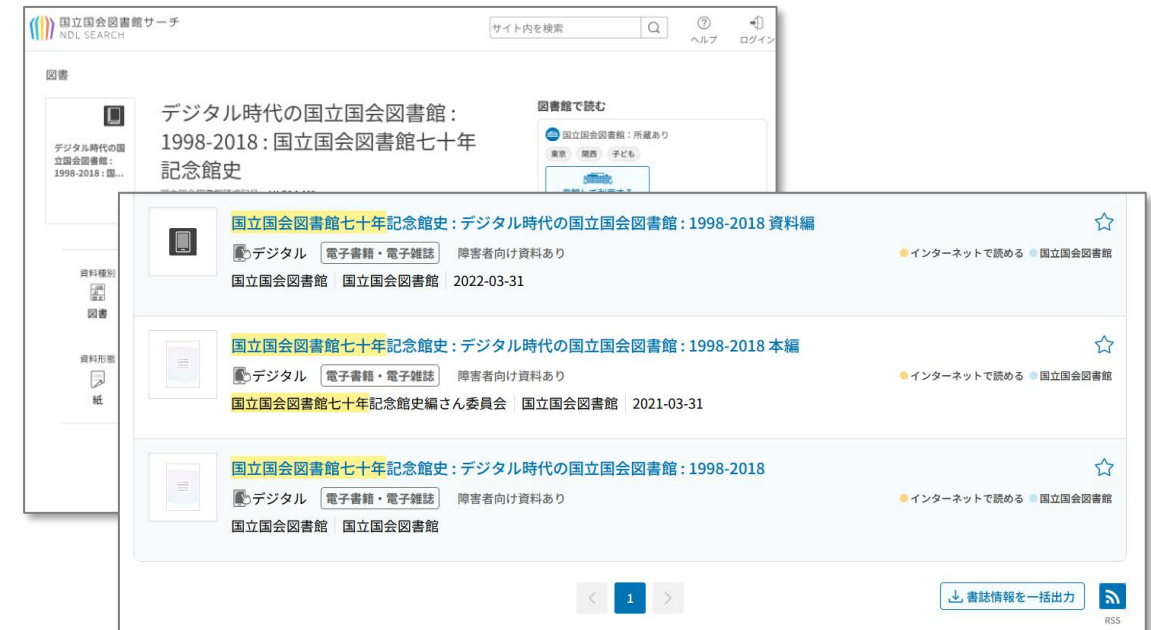
- NDLが作成した書誌データ等(書誌データ・典拠データ・雑誌記事索引データ)は、営利・非営利の目的を問わず無償で利用可能(CC BY相当)
- 検索結果一覧、書誌詳細、「お気に入り」(要ログイン)からダウンロード可能
- NDLの書誌データを用いて出版傾向を分析した研究もあり



検索キーワードを入力

☒ 国立国会図書館 ☒ 全国の図書館 インターネットで見ることができるものに絞る ☒ 絞り込み条件

国立国会図書館サーチとは



詳しくは…

書誌データの利用方法：https://www.ndl.go.jp/data/data_service/quickguide

NDLサーチのAPI等を通じたメタデータの提供

- 通常の画面表示に加え、API・全件ファイルでメタデータを提供
- NDLと全国の公共・大学・専門図書館、学術研究機関等のデータベース110DBと連携、約2億件のメタデータを提供(令和7年度末時点)
- NDLが作成したメタデータは無償で自由に利用可能(CC BY相当)
- 他機関が作成したメタデータは、各機関の利用規約に従い利用

リポジトリ 番号	データプロバイダ	データ提供機関	データプロバイ ダID	検索用 API	ハーベス ト用API	メタデータ利用条件 非営利 営利	
R000000004	国立国会図書館雑誌記事索引	国立国会図書館	zassaku	○	○	CC BY 🔗 ⁽¹⁾	
R000000004	国立国会図書館雑誌記事索引オンライン資料編	国立国会図書館	zassaku- online	○	○	CC BY 🔗 ⁽¹⁾	
R000000005	新書マップ	NPO法人連想出版	shinsho	○	×	○	△

NDLOCR

- デジタル化資料のテキスト化を目的としてNDLが開発した光学的文字認識(OCR)処理プログラム
- 明治期の資料から現代の雑誌まで幅広い年代の刊行物を対象にテキスト化可能
- CC BYライセンスでGitHubからソースコードを公開
→ オープンソースのため自由に利用可能
- 動作には一定スペック以上のGPU環境が必要



参考: 「NDLOCR ver.2の公開について: <https://lab.ndl.go.jp/news/2023/2023-07-12/>

NDLOCR-Lite

- GPU不要の軽量版
(令和8年2月公開)
- ノートPC等の一般的な
家庭用PCやOS環境で
動作

参考:

NDLOCR-Liteの公開について

<https://lab.ndl.go.jp/news/2025/2026-02-24/>



資料画像の出典: 芥川竜之介 著『傀儡師』,新潮社,1919. 国立国会図書館デジタル
コレクション <https://dl.ndl.go.jp/pid/1904835/1/53>

テキストデータの利活用

- 次世代デジタルライブラリー <https://lab.ndl.go.jp/dl/>
 - 著作権保護期間満了(パブリックドメイン)資料(図書・古典籍約35万点)のテキストデータの検索・ダウンロードが可能な実験サービス
 - APIから構造化データ(XML)としてテキストをダウンロード可能
 - 表のレイアウトを構造化(TSV/HTML)してダウンロード可能
 - 活用事例(JOSS2025「AI×文学研究の可能性を探る」セッションより)
 - 近藤泰弘先生「国立国会図書館デジタルコレクションのOCRデータからの復元」
 - 日比嘉高先生「コンピュータに小説のジャンルが判定できるか」
 - <https://lab.ndl.go.jp/event/joss2025/>



NDLのオープン化の取組と課題

オープンなコンテンツ・データ等の提供

- パブリックドメインのコンテンツ、メタデータ、オープンソースプログラム 等（スライド1参照）

オープンデータの利用実態やニーズの把握の困難さ

- ハッカソン等のイベントによるコミュニティとの接触の試み

パブリックドメインとして提供するための著作権調査

- デジタル化・テキスト化成果物をオープンに活用できるようにするためにはパブリックドメインであることの確認が必要

- 膨大な著作者の存在→公開調査の実施 <https://dl.ndl.go.jp/opening>

