

古い雑誌を誰でも閲覧——NDLデジタルコレクションをユーザ目線で(プレゼン資料)

短縮: <https://bit.ly/3lrApfO>

映像: <http://bit.ly/3JE3zZZ>

2023/03/14 イースト 下川和男 [shimokawa@est.co.jp](mailto:shimokawa@est.co.jp)

国立国会図書館(NDL)「[デジタルコレクション](#)」の「[個人送信](#)」に様々な雑誌を加えるセミナーの資料です。(個人的な見解であり、この文面は追記・修正することがあります)

デジタルコレクションには「ネット公開」「個人送信」「館内限定」の3レベルがあり、このメモでは、「館内限定」を「個人送信」に変更する方法を紹介しています。

「個人送信」は、日本在住の個人が[NDLに登録\(無償\)](#)すると、本文の閲覧が可能になる制度です。

『送信される資料は、絶版等で入手困難な資料に限られます。商業出版に係る雑誌は、入手困難であっても送信を留保されていますが、発行元の出版社等から送信希望があった場合には、一定の手続きを経て送信することができます』([雑誌のデジタル化に伴う調整](#))

NDLからの公開を希望される出版社は、[除外手続Q&AのQ24](#)を参考に「雑誌名」「送信希望号数(例:1巻1号から25巻12号まで)」などを連絡してください。

雑誌が所蔵されているか? デジタル化されているか?については、「[国会図書館オンライン](#)」で、出版社名、雑誌名などを入れて検索してください。

・検索結果で右に「[デジタル](#)」と表示されるものは、デジタル化が完了しており、2000年までのバックナンバーは、比較的早い時期に個人送信できると思われます。

・デジタル化されていない雑誌も、「[ビジョン2021-2025](#)」及び「[資料デジタル化基本計画2021-2025\(PDF\)](#)」の次の5か年計画などで、デジタル化される可能性があります。

・雑誌が見つからない(≒納本されていない)、初期の号が所蔵されていない、欠号があるものは、「納本(出版社から)」または「寄贈(個人から)」が可能です。その場合、[お問い合わせフォーム](#)から連絡してください。

※[著作権法](#)(31条)に基づく制度ですので、出版社は送信希望を出すのみ。

※デジタルコレクションの概要は、[2023/02/28](#)の講演映像をご覧ください。

※ビジョン2021-2025の概要は、[2022/02/02](#)、[2021/04/21](#)の講演映像をご覧ください。

---

#### 【参考】

・「明治・大正・昭和のGoogle」あっと驚く情報がいっぱい。誰からも見える!

平成元年:1989年≒日本のインターネット元年、ファミリー・ヒストリー

2022年12月21日に公開 OCR化 [247万点](#)をフルテキスト検索できる。

電子化済みは333万点、内180万点を個人送信、ネット公開は57万点 蔵書の1/10?

・2000年までの書籍・雑誌などのデジタル化:日本のDTPは21世紀から

1960~90年代に良書がたくさん、雑誌も厚い 出版の隆盛期:1996年2.7兆円

Windows 95/98 インターネットの普及と共に出版売上が減少

・個人送信が行われている雑誌 **11,547誌**(2022/07現在)

[情報管理](#) [コンピュータピア](#) [スウィングジャーナル](#) など

・デジタル化が完了している雑誌 約**14,000誌**

[月刊アスキー](#) [しにか](#) [Yahooインターネットガイド](#) [商店界](#) [蛍雪時代](#)など

- ・[オープンデータセット](#)で具体的な雑誌名、書籍名なども公開されています。
- ・著者の立場では、自分が書いた文章が[再び世に出るのは嬉しい方](#)が多い。  
出版は編集者と著者が車の両輪 編集者も歓迎
- ・読者の立場では、若い頃に刺激を受けた雑誌や記事が「入手困難」ではなくなり、インターネットで簡単に閲覧できるのは、この上ない幸せ。  
公共図書館は20年以内での廃棄が多く、20世紀の雑誌は読めない。  
読みたい雑誌があれば、出版社にリクエストしてみてください。

- ・松田弁護士:この事業から様々なビジネスが生まれる  
『[できるNDL](#)』『[NDL navi](#)』『[NDL fan](#)』  
BtoBtoCモデル NDL ⇒ 大学・研究機関 ⇒ 学生・研究者 (NDL登録は個人)  
サポート&サービス [Linux](#)と[RedHat](#)、[GIGAスクール](#)18歳以下 [MeL](#)、[KinoDen](#) SSO
- ・NDLだと「全文検索」「アクセシビリティ」「紙の劣化(非破壊スキャン)対策」「AI投入(後述)」などのメリットあり。

#### ※[NDL雑誌リスト\(私的なメモ\)](#)

- ・国の仕組みを活用、続きは個別Zoomか居酒屋で

---

#### 付録:【米国インターネット・アーカイブの活動】

- ・[Internet Archive](#)『[Byte](#)』『[Creative Computing](#)』『[工学社 I/O](#)』『[アスキー](#)』
- ・日本の書籍、雑誌も少し入っている:利用者からの投稿(裁断スキャン)非営利団体
- ・米国は[フェアユース](#)とオプトアウト。[松田政行弁護士「Google Books事件」](#)(2016)

---

#### 付録:【世界の中の日本 EPUB事例】

- ・世界にNDLテキストを公開? 漫画・アニメで日本を知りNDLで更に深く
- ・日本は島国、[日本語はマイナーな言語](#) [2010年EPUB日本語組版](#) [ネットで鎖国は無理](#)  
アクセシビリティ:a11yと国際化:i18n、三瓶「世界中に日本語の本が売れる」
- ・[fantasai](#)+[村田](#)・[小林](#)・村上・石井 ⇒ 映像[木田](#)+[Dave Hyatt\(WebKit\)](#) ⇒ [新名](#)
- ・[EPUB 3.3](#) 映像[高見](#)・[Florian](#)

---

#### 付録:【シンギュラリティ前夜】

- ・検索の代替ではなく、「インターネット+スマホ」と同等の影響力
- ・[大量の失業者](#) ⇒ リスキリング [1兆円/5年](#)
- ・160~200時間/月 ⇒ 100時間で同じ成果
- ・人工知能に正確な情報を大量に与える ⇒ 国会図書館
- ・NDL [OCRの精度向上](#)、97 ⇒ 99.8% これにもAIが使える。API公開

---

#### 付録:【NDL検索とChatGPT】

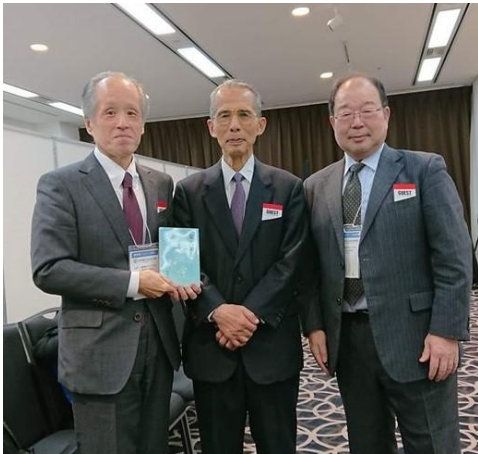
- ・[大規模言語モデル](#)([ChatGPT](#)など)への期待:校閲済みのテキスト(学習データ)
- ・イーロン・マスク(Tesla, SpaceX, StarLink, Twitter)とサティア・ナデラ(MS)

- ・[OpenAI](#)+Microsoft、GAFAIにMがない理由、W3Cが推進する分散型ID [DID](#)も提供
- ・AmazonもYahoo!も「検索」が「AIチャット」に

---

付録:【国産 大規模言語モデル】

- ・[数百億円](#): 日本ディープラーニング協会理事長 松尾豊先生、[PFN/Rapidus](#)
  - ・NDLにとっては、21世紀の書籍デジタル化の機会 時間軸 [TimeMap](#)
  - ・出版社からPDFを購入する手も [Bingチャット](#)風に引用表示⇒書籍販売に繋がる
  - ・これが本当の「長尾構想」2007-12 日本の知識、叡智の海、集合知 オープンアクセス
- ※学術論文 [DOI](#)と[CrossRef](#)、論文のオープン化、映像[オープンサイエンス](#)、[J-STAGE](#)



---

付録:【出版社が商用で雑誌をデジタル化する場合】

- イーストの「[電子復刻](#)」「[技術雑誌 電子復刻](#)」『[bit](#)』『[谷根千](#)』などの制作実績。
- ・権利処理などで数百万円のコストがかかる場合も。
  - ・[文化庁長官裁定制度](#): デジタルの時代 ⇒ [新制度\(2023年～\)](#) 数年後
  - ・事例: 『[bit](#)』『[週刊東洋経済](#)』

---

付録:【NDL OCRを活用しPDFをアクセシブルに】

- ・[NDL Lab](#): GitHub上にオープンソースで公開
- ・これと、Adobe PDFライブラリを併用し、[PDFest.jp](#)を開発中
- ・PDFからHTML、PDFからEPUBなど構造化されたテキストを出力
- ・WebサイトにあふれるPDFをアクセシブルなHTMLやEPUBに、書籍もEPUB化

# OCRテキスト化実験

目的

- ①デジタル化資料の利活用促進 → 資料の本文や図版を検索できるようにする。
- ②データセットの提供 → 大量の学習用データとしてデータセットを公開する。
- ③視覚障害者等用データへの活用 → 利用可能なテキストデータを増大させる。

