

国立国会図書館の図書等のデジタル化についての提言

令和2年9月1日
自由民主党 政務調査会

新型コロナウイルス感染拡大の影響により、全国の図書館が休館するなど研究者や教育者、学生などに大きな影響を及ぼした。同時に、ウィズコロナ時代のニューノーマルに対応した新たな研究・教育活動を下支えする学術環境の早期構築の必要性が明らかになった。

国立国会図書館の図書等のデジタル化の状況を振り返ると、著作権法（第31条第2項）により所蔵資料の網羅的なデジタル化が可能とされているが、実際には、年間約43万点の図書等の収集に対し、わずか2万点（約5%）のデジタル化処理能力（予算）しか備えていない。また、国内刊行の図書等のうち、概ね1968年以前の発刊の図書等244万点（約20%）のみしかデジタル化されておらず、デジタルアーカイブを前提とした図書等の活用に至っていないのが現状である。

また、①デジタル化された図書等についてOCR（光学文字認識）処理がなされておらず本文検索ができない、②有償の電子書籍等や新聞社のWEBのみに掲載されたニュースについては収集もデジタルアーカイブもされていない、③もともとデジタルで作成された図書等についても紙資料での納本後改めてデジタル化せざるを得ないなどの課題が指摘されている。

政府の「骨太の方針2020」においても「教育・研究環境のデジタル化・リモート化」「誰もがデジタル化の恩恵を受けられる体制整備」などと指摘されており、また、関係府省庁では、国立国会図書館でデジタル化された資料へのアクセスを容易化するための著作権法改正が検討されていることから、これと軌を一にして、日本の知の中心である国立国会図書館の図書等のデジタル化を推進し、収集した資料の保存と知の活用を強かに押し進める必要がある。

具体的には、まず、デジタル化されていない2000年以前に出版された図書等（165万点）について、5年以内に紙資料を電子化する。その際、データにはOCR処理を行い、画像だけでなくコンピュータ文字として認識できる状態で保存し、すでにデジタル化している図書等についても併せてOCR処理を行う。2001年以降に出版された図書等については、2026年までに広くデジタル化の実現手段を探る。

活用面では、国立国会図書館の関係者協議や関係府省庁において、上記課題やウィズコロナ時代に対応した図書館送信制度、絶版等資料へのアクセス容易化のほか、デジタルアーカイブされた図書等の全文検索とスニペット表示（検索キーワードの前後文章の表示）の公開促進などの実現に向けた検討がなされることが望ましい。

いずれにしても、専門的知見に基づく調査や豊富な情報資源の提供によって国会の活動をサポートするという重要な役割を担っている国立国会図書館におけるデジタル化の推進は、我々立法府に課せられた責務である。また、政府に対しても予備費の活用を含めた予算面での手当などその推進を強く求めるものである。

以上

所蔵資料デジタル化事業の促進について

合計 207億円

- ① 資料デジタル化 190億円
※38億円/年×5年間
⇒1969～2000年刊行の図書約165万点のデジタル化を外部委託により行う。デジタル化された資料は③によるOCRで全文テキスト化し、日本語ビッグデータとしても整備し、当該期間の資料の分析から様々な再発見を行うための基盤とする。
- ② 恒常的なデジタル化設備の整備 3億円
⇒年間10万点程度のデジタル化を行う設備を東京本館内に整備する。デジタル化するものは、③で開発されるOCRによって全文テキストによる検索を可能とする。
- ③ 資料利活用のための高精度分析OCRシステムの開発 4億円
⇒デジタル化資料の画像データから、全文テキスト化（OCR）と、挿絵や写真等の抽出・検索を可能とするためのAIシステムを開発する。利用者の利便性が向上するだけでなく、テキストと画像というAI等のコンピュータで再利用可能な形に資料を転化させることを可能とする。
- ④ 電子書庫（ストレージ）の増強及びシステムの改修 10億円
⇒大量のデータを処理するために電子書庫（ストレージ）の増強及び国立国会図書館デジタルコレクションの改修を行うとともに、大規模なデジタル化作業の進捗・品質を効率的に管理するための情報システムを構築する。

期待される効果：

- (1) 大学・図書館等が閉鎖された状況下での研究・教育活動を下支えする学術環境整備（オンラインによる情報アクセスの拡大）
- (2) ジャパンサーチとの連携により、我が国のコンテンツ活用に貢献
- (3) 全文テキストデータの提供により、データ駆動型社会・研究に貢献