

国立国会図書館における “データ”の収集・保存・提供の取組

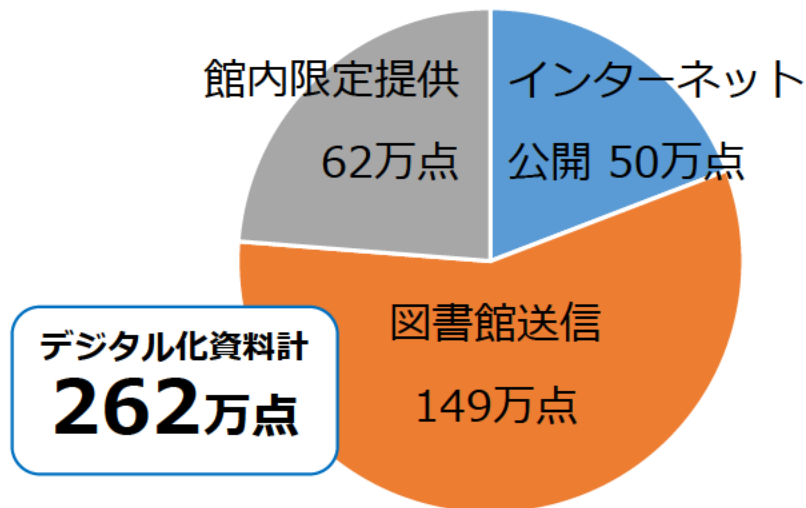
平成29年2月28日

国立国会図書館電子情報部

主な取組

- ① デジタル化資料（画像データ）の提供
- ② インターネット資料・オンライン資料の収集・保存・提供
- ③ 国会会議録（テキスト・画像データ）の提供
- ④ 書誌データ・典拠データの提供

① デジタル化資料（画像データ）の提供



現状

- **262万点**のデジタル化資料を提供（平成29年1月現在）
※当館所蔵資料のデジタル化は、著作権法第31条第2項の規定による
- 絶版等で入手困難な**149万点**は、全国の公共・大学図書館等770館に送信
※著作権法第31条第3項の規定による
- 著作権保護期間満了を確認できたもの（PD）、文化庁長官裁定を受けたもの、著作権者の許諾を得たもの、**計50万点**をインターネット公開
- うち、パブリックドメイン（PD）資料の35万点は、当館への申込み不要で、営利・非営利問わず自由に利用することが可能
- 残りの62万点は、国立国会図書館内でのみ閲覧可能
- 図書館内での複写サービスはプリントアウト提供のみ



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

課題

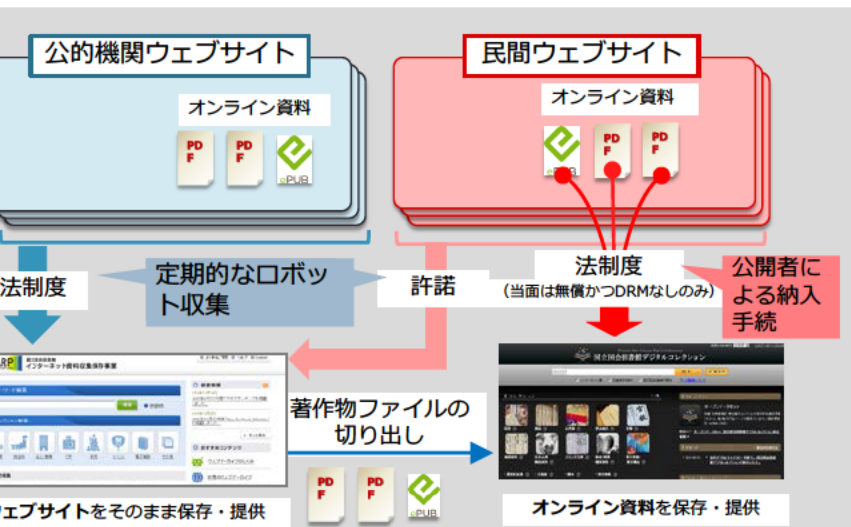
テキストデータは、目次のみ作成し提供している
 現行法上、NDLが画像データをOCR処理して本文のテキストデータを作成し、スニペット表示等で利用提供するには、著作権者の許諾が必要であるため、本文のテキスト化は進んでいない

② インターネット資料・オンライン資料の収集・保存・提供



インターネット資料収集保存事業 (WARP)

<http://warp.da.ndl.go.jp/>



現状

- **インターネット資料収集保存事業 (WARP)** は、平成14年から収集を開始、平成29年1月現在で累計約11.7万件のウェブサイト (1.1万タイトル, 855TB) を収集・保存
- 国の機関、地方公共団体、国立大学等の公的機関のウェブサイトは、制度的に収集
 - ※著作権法第42条の4及び国立国会図書館法第25条の3の規定による
- 民間のウェブサイトは、公益法人、私立大学、政党、国際的イベント等を中心に、発信者の許諾を得て収集
- **オンライン資料** (ウェブに掲載されている文献相当の資料) は、無償かつDRM無しものを制度的に収集し、**40万点**を保存
 - ※著作権法第42条の4及び国立国会図書館法第25条の4の規定による
- インターネット資料やオンライン資料は、国立国会図書館内でのみ閲覧が可能。**発信者の許諾が得られたものは、インターネット公開**

課題

現行法上、NDLが収集・保存しているインターネット資料・オンライン資料のインターネット公開には発信者の許諾が必要。これらの資料をAIの学習用データ等に利用可能な形式でのデータ複製の提供には、法制度の整備が必要

③ 国会会議録（テキスト・画像データ）の提供

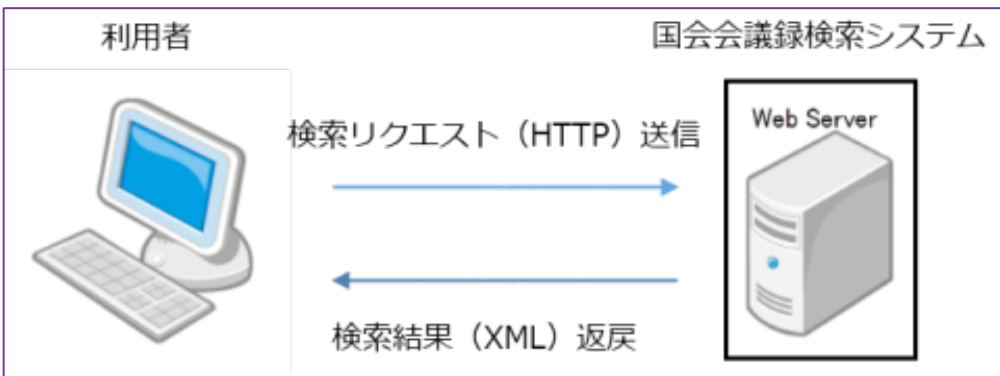


国会会議録検索システム <http://kokkai.ndl.go.jp/>

現状

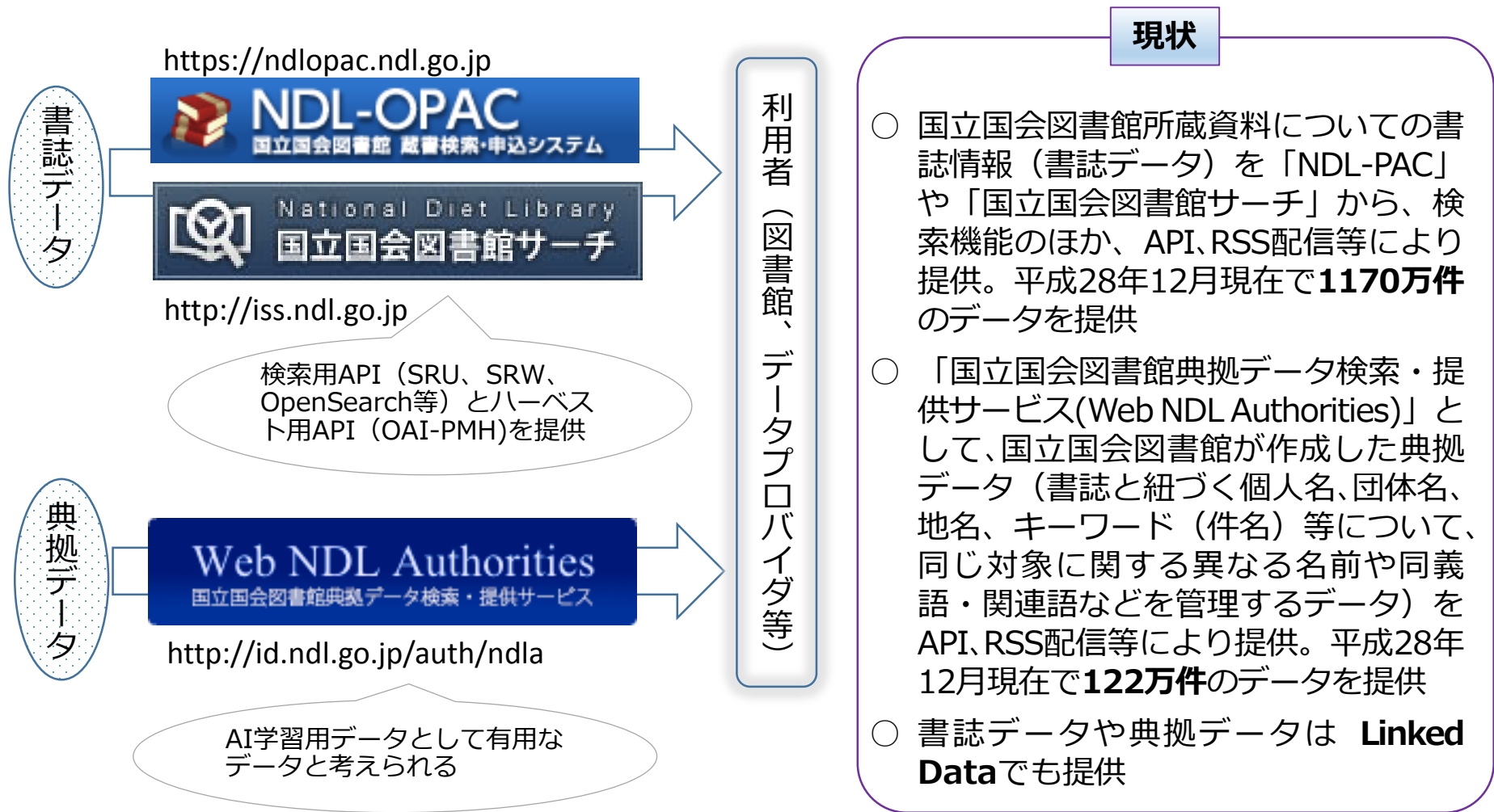
- 衆議院・参議院各事務局と共同で「国会会議録フルテキスト・データベース」を構築
- 第1回国会以降の本会議・委員会等の会議録のテキストデータ（会議録情報と議事部分）及び画像データ（173万ページ、平成29年1月現在）を提供
- 「国会会議録検索システム」による検索・閲覧に加えて、検索結果の複製（著作権者の許諾が必要な場合を除く）、「国会会議録フルテキスト・データベース」との検索用APIによる機械的な連携（営利目的を除く）が可能

※ 著作権法第40条（政治上の演説等の利用）の規定により、原則として、著作権者の許諾を得ずに利用することが可能



検索用APIにより、テキストデータの機械的な入手・利用が可能

④ 書誌データ、典拠データの提供



課題

現在、NDLの書誌データ及び典拠データは、非営利目的の場合に無償で提供しているが、更なる利用の拡大に向けた取組が必要