

国立国会図書館における デジタルアーカイブと研究データ ~保存と活用の可能性

国立国会図書館 電子情報部副部長 川鍋 道子

内容

1. 国立国会図書館の様々なデジタルアーカイブ

2. デジタルデータ ≒ 研究データ!?

3. 国立国会図書館による「研究データアーカイブ」への積極的関与の可能性

4. デジタルアーカイブと研究データの保存と活用における共通課題

1. 国立国会図書館 様々なデジタルアーカイブ

① デジタル化資料(画像データ)の提供



国立国会図書館デジタルコレクション

http://dl.ndl.go.jp/

インターネット NDL館内限定提供 64万点 公開 53万点

図書館送信 デジタル化資料計 266万点

149万点

概要

- 266万点のデジタル化資料(画像) と目次のテキストデータを提供
- 一部音声や映像データもあり
- 絶版等資料の149万点を、全国の公 共・大学図書館等約850館に送信
- 著作権処理をした計53万点をイン ターネットで公開 (うち、パブリックドメインの**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64万点は、国立国会図書館内 でのみ利用可能

◆(参考) デジタル化資料の内訳

(2017年6月時点)

| 資料種別 | インターネット 公開 | 図書館送信 | NDL館内 限定提供 | 合計 | 概要 |
|---------------|---------------|-------|----------------------|-------|---|
| 図書 | 35万点 | 55万点 | 7万点 | _ | 明治期以降、1968年までに受け入 れた図書、震災・災害関係資料の 一部(1968年以降受入含む) |
| 雑誌 | 1万点 | 80万点 | 47万点 | 128万点 | 明治期以降に刊行された雑誌(刊 行後5年以上経過したもの) |
| 古典籍 | 7万点 | 2万点 | 1 | 9万点 | 貴重書・準貴重書、江戸期以前の 和漢書等 |
| 博士論文 | 1万点 | 12万点 | 1万点 | 14万点 | 1991~2000年度に送付を受けた論 文 |
| 録音·映像 関係資料 | - | - | 0.4万点 | 0.4万点 | カセットテープ、ソノシートなどの録 音資料、脚本、手稿譜等 |
| その他 | 8万点 | - | 9万点 | 17万点 | 官報、憲政資料、日本占領期関係 資料、歴史的音源等 |
| 合計 | 53万点 | 149万点 | 64万点 | 266万点 | |

※概数なので計が合わない場合があります。

② インターネット資料の収集・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/

概要

- 国内サイトを対象としたウェブアー カイブ
- 2002年から収集開始し、累計約13 万件のウェブサイト(1.1万タイト ル,1036TB)を収集・保存・提供
- 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、法 律に基づき網羅的に収集
- 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
- ○国の機関は毎月、その他は4半 期に1回の収集
- 発信者の許諾が得られたものは、イ _ ンターネット公開

③ 震災に関する記録・データの収集・保存・提供



NDL東日本大震災アーカイブ(ひなぎく)

http://kn.ndl.go.jp/

【記録の形式は、紙媒体、デジタルを問わない】

- ・図書、雑誌、報告書、調査レポート
- ・ウェブサイト
- ・写真、映像・動画
- ・音声
- ・ファクトデータ

概要

- ○東日本大震災に関する記録・ データを一元的に検索・活用で きるポータルサイト
- ○東日本大震災に関するあらゆる 記録・教訓を次の世代へ伝え、 被災地の復旧・復興事業、今後 の防災・減災対策に役立てるた めに、関連する音声・動画、写 真、ウェブ情報等を包括的に検 索可能
- **39機関45データベース**と連携
- 詳細検索画面では、一部の資料 に対して本文検索が可能
- ○メタデータを各種APIで提供

4 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/



概要

- 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

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

2. デジタルデータ ≒ 研究データ!?

①古典籍資料もインターネットで研究の時代へ

山の幸 [石寿観秀国] [編] ; [勝間龍水] [画]. 明和2 [1765] 序

- ・国立国会図書館は、下巻を2冊所蔵。 - 請求記号 WB1-22とWB1-23
- ・NDL以外に、国内では東北大学、東洋文庫などがデジタル画像を公開
- ・国外では、ボストン美術館、ゲティ財団、などでデジタル画像を公開



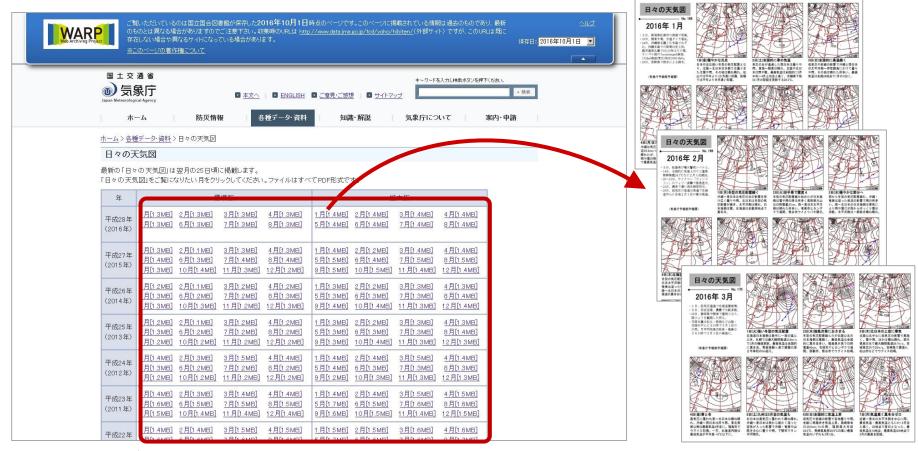
玉虫の下に葉のある版とない版がある。葉のある版は、 国立国会図書館とボストン美術館のみ

(出典) 国立国会図書館月報651号 (2015.7) DOI: 10.11501/9422688

http://dl.ndl.go.jp/view/download/digidepo_9422688_po_geppo1507.pdf?contentNo=1

②WARPにおける研究データの収集例

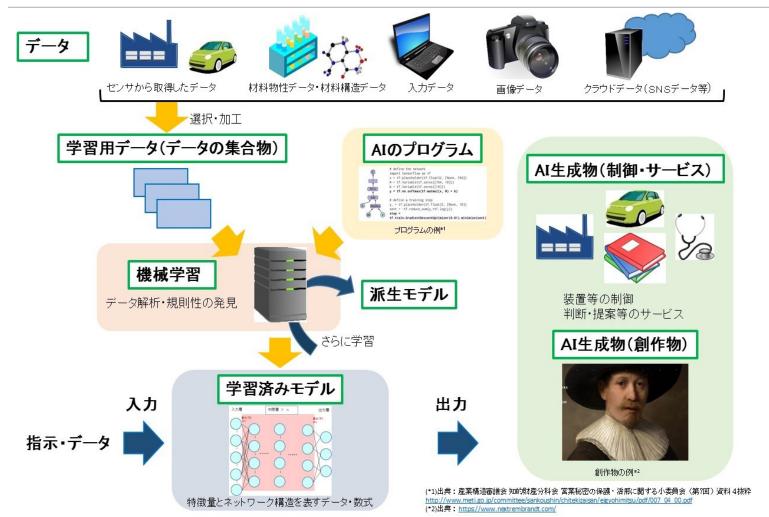
収集ロボットでクロールできる範囲で収集・保存が可能



気象庁のウェブサイト(2016年10月1日)

http://warp.da.ndl.go.jp/info:ndljp/pid/10202790/www.data.jma.go.jp/fcd/yoho/hibiten/

③人工知能(AI)の学習用データとして



(出典) 内閣府「新たな情報財検討委員会報告書 – データ・人工知能(AI)の利活用促進による産業競争力強化の. 基盤となる知財システムの構築に向けて – 」(平成29年3月)図6(機械学習を用いた A I の生成過程のイメージ)抜粋

http://www.kantei.go.jp/jp/singi/titeki2/tyousakai/kensho hyoka kikaku/2017/johozai/houkokusho.pdf

3. 国立国会図書館による「研究データアーカイブ」への積極的関与の可能性

①DOIを使ったデジタル画像の利活用の促進

国立国会図書館

原資料のデジタル化





デジタル画像へのDOI 付与



デジタル化事業の評価の フィードバックも

デジタル画像の利用

インターネッ トで利用 発表論文に参考文 献をDOIで記載







DOIによる評価・分析

- ・論文とデータの被引用度等で デジタル画像の評価が可能に
- ・デジタル化の成果が研究動 向にどう反映しているのか、 などの分析も可能に

デジタル画像の活用

◆他の研究者がDOI検索し、目的のデジタル画像を探す



簡単にたどり着けるうえ、 メタデータに記された情 報も入手できる

◆発表論文にもDOIを付与し、 論文と参考文献のデジタル 画像を結びつける



アクセスポイントが増加 し、検索しやすくなる

研究活動の活性化・高度化、コンテンツの利活用の促進へ

②他機関が作成した研究データの長期保存

インターネット資料収集保存事業(WARP)を用いたデータの長期保存



研究データへのDOI付与

(データの概要を説明するウェブページ の用意、メタデータの整備等)



WARPで収集・保存されているデータに、研究者自らがDOI付与へ

【WARPによる収集】

研究機関のウェブサイト上の

研究データ

ウェブサイト収集

WARP上の

研究データ

URL: http://warp.da.ndl.go.jp/

+永続的識別子

+オリジナルサイトのURL

【研究データのDOI登録】

DOIのメタデータ

タイトル情報:xxxxx

出版情報:xxxxx

DOI: 10.xxxxx/xxxxxx

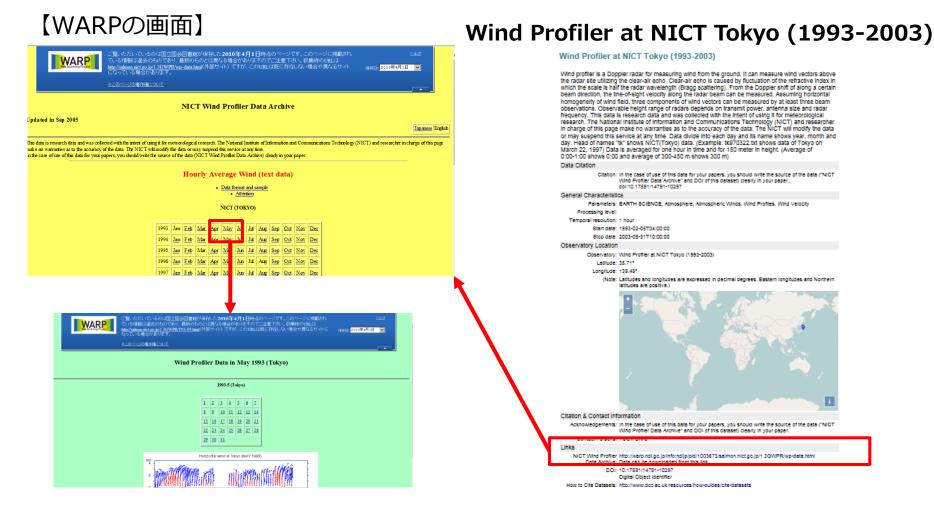
URL: (WARPのURL)

永続的アクセス

論文(研究データの 引用)等

◆(参考) WARPを活用したDOI付与第一号

http://doi.org/10.17591/14791-10297



③研究データのメタデータ連携の促進

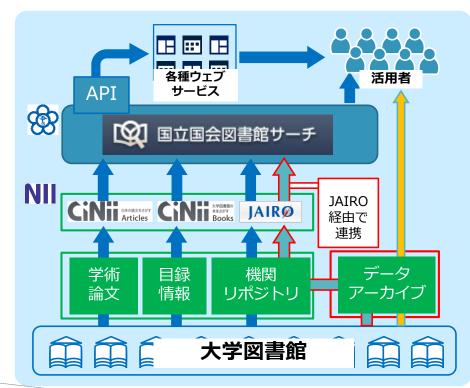
例えば

保証

記を含む)の流通

- ・大学図書館に関しては、機関リポジトリにメタデータが登載されるこ とで、NIIのJAIRO経由で今の枠組みでもNDLサーチと連携が可能に
- ・国の機関等の公共データや、学 協会・企業等の研究データのメ タデータ流通のためには、**DOI** (デジタルオブジェクト識別 子) **のメタデータの活用**へ





ただし、今はまだ研究データの、標準メタデータ フォーマットの確立や、APIによる提供は未実現

NDL関与 の余地?

NDLが共同運営するジャパンリンクセンター(JaLC)を通じた研究データのDOI付与・普及へ

4. デジタルアーカイブと研究データ の保存と活用における共通課題

デジタルアーカイブと研究データの共通課題?!

【メタデータの整備、標準化】

- 分野ごとに必要な情報が異なるため、分野 ごとのメタデータフォーマットの標準化が 必要
- メタデータ整備に係る人的コストの確保が 課題(中身がある程度理解できる人が関わ る必要がある)

【データの長期アクセス保証】

- データを長期保存する意識が低い
- 長期保存のためには技術的支援が必要

【データの拡充・オープン化】

- 利活用可能なデジタルデータを増やす には、ラインセンスの整備が必要
- 権利処理をするための法務的知識や人的コストの確保が課題
- プライバシー保護や肖像権などの著作 権以外の問題を抱えている

【データの利活用の促進】

データの利活用が求められるがその具体策が見えない

課題解決に向けて

コミュニティ基盤構築の必要性 意識啓発/インセンティブ付与の必要性



素敵な出会いがありますように…

ご清聴ありがとうございました