

国立国会図書館の ウェブ・アーカイビング

国立国会図書館
総務部 企画・協力課
電子情報企画室長 植月献二

ウェブ情報利活用ワーキンググループ 2004.07.30



ウェブ・アーカイブの必要性

社会的基盤としてのデジタル情報

- 人々のさまざまな活動がコンピュータと情報通信ネットワークに深く依存
- あらゆる領域で情報が電子的に生産・流通し、利用されており、デジタル情報が社会的基盤として重要となっている。

インターネット情報資源の量

インターネット総体では . . .

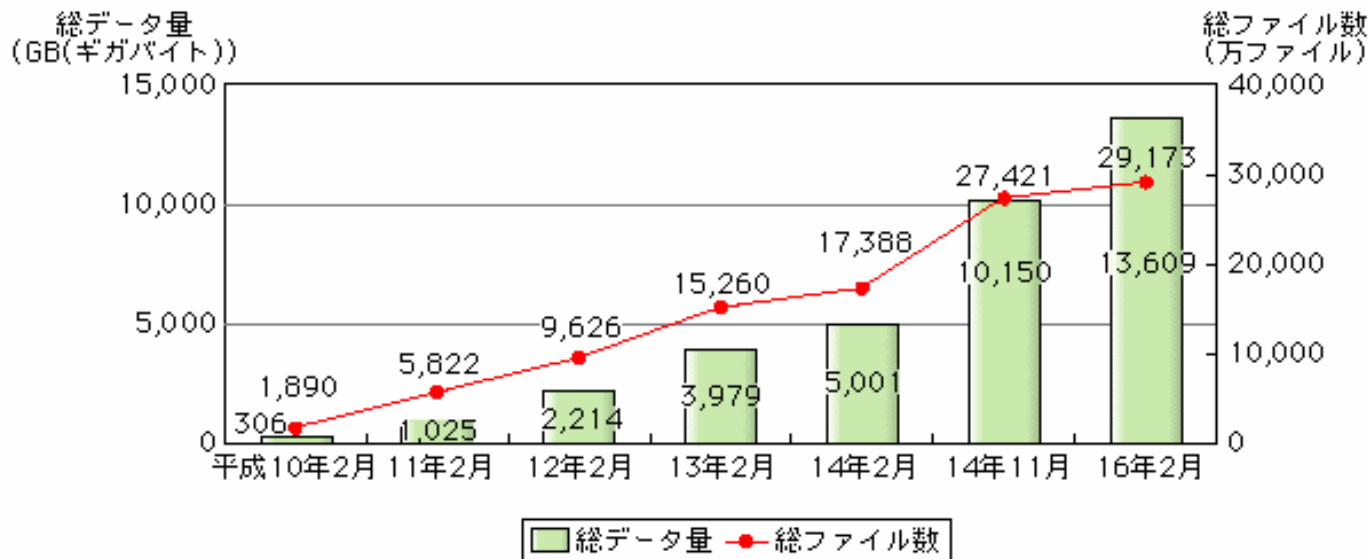
表層ウェブ 167TB 機械的自動収集可能な範囲
約500倍

深層ウェブ 67 ~ 90PB (PB = 1000TB)

参考 : How much information 2003 (<http://www.sims.berkeley.edu/research/projects/how-much-info-2003/internet.htm#wbpub>)

インターネット情報資源の量

JPドメインでは



【参考】

	平成10年2月	11年2月	12年2月	13年2月	14年2月	14年11月	16年2月
サーバー数(台)	36,000	75,000	95,000	152,000	197,000	308,000	425,000
ページ数(万ページ)	1,023	2,953	4,255	6,107	6,558	7,438	8,590

平成16年版情報通信白書

デジタル情報に関する国際情勢

ユネスコ (第32回総会(2003年))

- 「デジタル遺産の保存に関する憲章」
- 「多言語主義の促進・使用及びサイバースペースへの普遍的アクセスに関する勧告」

各国での取り組み

国際的事業 (IIPCなど)

WARPとは

- **国立国会図書館インターネット資源選択的蓄積実験事業**(WARP: Web Archiving Project)
 - 表層ウェブを著作権者との許諾契約に基づき選択的に収集、蓄積
- **経緯**
 - 納本制度調査会答申(平成11年2月22日)
 - 「ネットワーク系電子出版物を契約により収集」
 - 納本制度審議会: ネットワーク系電子出版物審議(2002年3月～)
 - ネットワーク系電子出版物の納本制度への組み入れの可否、組み入れられない場合の国立国会図書館としての収集すべき範囲等を検討を諮問。
 - 最終答申は2004年12月

コレクションの種類と量

- 3つのコレクション
 - 電子雑誌 (約1,100タイトル)
 - 政府機関 (8機関)
 - 協力機関 (約600サイト)
- 「蔵書」量
 - 10,910,000 files
 - 622GB

参議院、参議院法制局、国土地理院、環境省、
国立国会図書館、国際子ども図書館、防衛庁、
会計検査院

WARPの所蔵データ(2004年6月末現在)

	タイトル数	個体数	ファイル数(万)	容量(MB)
電子雑誌	1,108	2,306	111.5	130,401
政府機関	8	98	131.3	76,607
協力機関	598	1,096	848.4	415,956
合計 / 平均	1,716	3,500	1,091.2	622,964

提供サービス

NDL-ホームページからのサービス



The screenshot shows the NDL homepage with the following elements:

- Header:** Address bar shows <http://www.ndl.go.jp/>. Navigation icons for home, search, and move are visible.
- Search:** A search box labeled "サイト内検索" with a "検索!" button and a "サイトマップ" link below it.
- Language:** "Contents" and "English" links.
- Main Navigation:**
 - 資料の検索 (資料の検索) - NDL-OPAC
 - 電子図書館の蔵書
 - 日本の国会・世界の議会 (日本の国会・世界の議会) - 国会会議録
 - テーマ別調べ案内
- Visual Elements:** A large gold paperclip graphic and the text "National Diet Library 国立国会図書館".
- Service Points:** "サービスポイント" section with links to: 東京本館, 関西館, 国際子ども図書館, お近くの図書館から, 海外から, お問合せ.
- Publications:** "ギャラリー" and "刊行物" buttons.
- System Maintenance:** "システム・メンテナンス" section with a notice: "NDL-OPACは、システムメンテナンスのため、平成16年5月2日から5日、6月20日から21日、6月25日から27日、9月19から23日は、検索および申込みサービスがご利用いただけません。また、6月22日から24日は、検索はできませんが、申込みはご利用いただけます。"
- Footer:** "国立国会図書館について" (お知らせ), "図書館員のページ" (採用情報), and "利用上のご注意".



ウェブ・アーカイビング・プロジェクト

WARP

国立国会図書館
インターネット資源選択的蓄積実験事業

本事業の紹介

推薦のお願い

お知らせ

- 電子雑誌が1,000タイトルを超えました。
- Internet Explorer 5.5SP2, Netscape 6.2以上推奨
- パンフレットはこちら[\[日本語\]](#)[\[English\]](#)

電子雑誌コレクション

政府ウェブコレクション

協力機関ウェブコレクション

国立国会図書館トップへ 電子図書館の蔵書へ

ヘルプ

All Rights Reserved Copyright © National Diet Library Japan 1999-2003

電子雑誌コレクション



JEA NEWS
(日本環境協会)

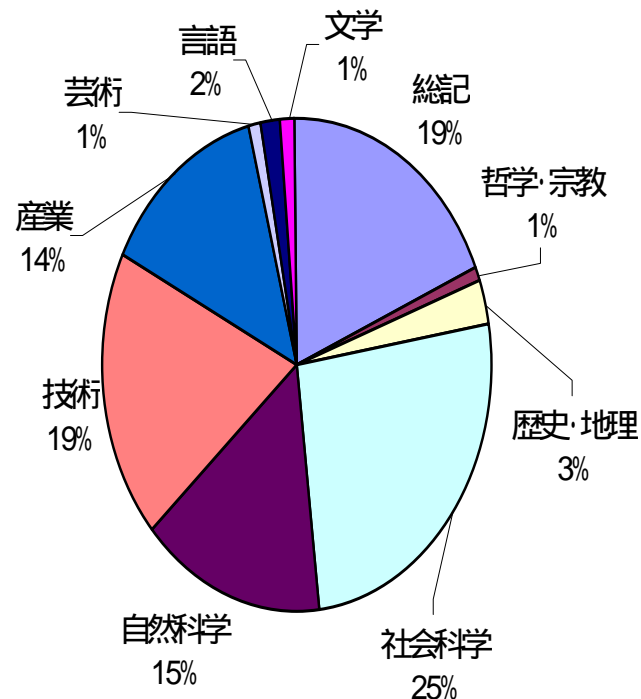


燦
(ひょうご科学技術協会)



DBJournal
(日本政策投資銀行)

電子雑誌の分野内訳(2004年7月13日現在)



- **定義**

- 同一のタイトルのもとに終期を予定せず、
巻次・年月次等の表示を伴って、
継続的に発行されるネットワーク系電子情報

- **約1,100タイトル**

- **分野**

- 経済学、教育学、
工学、医学、
農林水産関係、
図書館学などが充実

- **ネット上から「廃刊」した雑誌も**

- 『高齢者のケアと行動科学』
- 『月刊社会民主』 など

電子雑誌コレクション

[国立国会図書館トップへ](#) [電子図書館の蔵書へ](#) [WARPトップへ](#) [ヘルプ](#)**電子雑誌一覧**

コレクションのすべてを一覧できます。

表示**かんたん検索**

キーワードを入力してください。

検索**詳しく検索**

いろいろな項目から検索できます。

詳細検索**テーマ検索**

探しているテーマを選んでください。

総記	哲学	歴史	社会科学	自然科学
技術	産業	芸術	言語	文学



電子雑誌コレクション 検索結果一覧

[国立国会図書館トップへ](#) [電子図書館の蔵書へ](#) [WARPトップへ](#) [電子雑誌コレクションへ](#)

[ヘルプ](#)

検索結果: 400件

検索条件: 電子雑誌コレクション

1 2 3 4 5 6 7 8 [次へ→](#)

1 - 50 / 400件中

1. [ADALTIORASEMPER](#) / 神戸市外国語大学図書館
2. [Annual Report](#) / 放射線医学総合研究所
3. [Biblia](#) / 東京純心女子大学図書館
4. [CAST NEWS:北海道大学先端科学技術共同センターニュース](#) / 北海道大学先端科学技術共同研究センター
5. [CCIJ News](#) / 生協総合研究所
6. [CDNLAO Newsletter](#) / CDNLAO / National Diet Library
7. [CLAIR REPORT](#) / 自治体国際化協会
8. [Crec](#) / 中部開発センター
9. [DBJournal](#) / 日本政策投資銀行総務部 / 日本政策投資銀行
10. [DORE\(参考調査課発ニュースレター\)](#) / 北海道立図書館奉仕部参考調査課 / 北海道立図書館
11. [DPRI newsletter](#) / 防災研究所ニュースレター編集委員会 / 京都大学防災研究所
12. [Economic Review](#) / 宣十通総研経済研究所

電子雑誌コレクション 本文一覧

国立国会図書館トップへ 電子図書館の蔵書へ WARPトップへ 電子雑誌コレクションへ ヘルプ

1件目/1件中 [↑検索結果一覧へ](#)

【利用上のご注意】
WARPではデータの収集、保存、提供に関する許諾契約を結んでいます。収集データの著作権は、オリジナルのデータの著作権者が保有しています。著作権に十分ご留意の上、ご利用下さい。

書誌のまとめり

タイトル : にっぽんNOW
URL : <http://www.kantei.go.jp/jp/nippon-now/>

- 2004年4月6日 収集
- 2002年12月5日 収集
- 2002年9月5日 収集
- 2002年8月21日 収集

個体

書誌情報

編者 今週の日本
出版者 今週の日本
資源識別子 [URI] <http://www.kantei.go.jp/jp/nippon-now/>



にっぽんNOW

[\[政府広報\]](#)

にっぽんNOWは、内閣府政府広報室編集協力のもとに、政府の施策や暮らしに役立つ行政情報を、写真、イラスト、図表等を使って、分かりやすく解説した新聞です。

毎月2回月曜日・年23回発行
 株式会社 今週の日本 編集・発行
 電話03-3265-6031

※ここに掲載した文書を読むためには [Adobe Acrobat Reader](#)が必要です

◎平成14年8月5日号 [\[PDF\]](#)

- [1面](#) 住民基本台帳ネットワークシステム
- [2面](#) 住宅火災から生命を守る！！
知ってる？ 青い羽根募金
- [3面](#) 増えています！ 家庭での不慮の事故

◎平成14年7月15日号 [\[PDF\]](#)

- [1面](#) 何を基準に食品を選んでいきますか？ 品質表示よく確かめて
- [2面](#) 夏のレジャーを安全に 子ども霞が関見学デー
- [3面](#) 海の日・海の恵みにありがとう！
「お札づくりの日」記念特別展 秋季雅楽演奏会
- [4面](#) 子どもや高齢者の「低アルコール飲料」誤飲
「読者から」証券会社・保険会社のペイオフは？

◎平成14年7月1日号 [\[PDF\]](#)

- [1面](#) 青少年を地域でまぐもぐ
- [2面](#) 海外旅行－安全を第一に楽しく
愛の献血にご協力を！
- [3面](#) 知っておきたい税の知識(保険、株式)
創刊一周年「小泉内閣メールマガジン」 ふるさとの味

◎平成14年6月17日号 [\[PDF\]](#)

- [1面](#) 市町村の合併 全国の自治体数1000を目標
- [2面](#) 男女共同参画社会の実現を目指して
韓国へ旅行される皆様へ

電子雑誌コレクション 詳細検索

[国立国会図書館トップへ](#)
[電子図書館の蔵書へ](#)
[WARPトップへ](#)
[電子雑誌コレクションへ](#)

[ヘルプ](#)

詳細検索

タイトル

タイトル中の単語で検索

部分一致 ▾

AND ▾

[用例](#)

編者

部分一致 ▾

AND ▾

[用例](#)

出版者

部分一致 ▾

AND ▾

[用例](#)

URL

部分一致 ▾

OR ▾

[用例](#)

ISSN

前方一致

[用例](#)

NDC

NDC

完全一致 ▾

OR ▾

[用例](#)

項目間: **AND** ▾ で検索し **タイトル** ▾ を **昇順** ▾ で **50** ▾ 件ずつ表示する

検索

クリア



政府機関コレクション

【インターネット公開】

- **参議院**
 - 2002年10月より月1回の高密度な収集を実施
 - 本会議・委員会の審議経過、議案審議情報等
- **国土地理院**
 - 4.3GBに及ぶ全国2万5千分1地形図等
- **環境省**
 - 審議会答申・議事録、環境統計、国際会議、環境調査、年次報告書等々
- **国立国会図書館**

【館内提供に限定】

- **防衛庁** 条約、防衛白書等
- **会計検査院**



協力機関コレクション



協力機関の内訳 (2004年7月13日現在)

方針

- 公共性・学術性の高いウェブ情報
- 失われやすいウェブ情報を優先

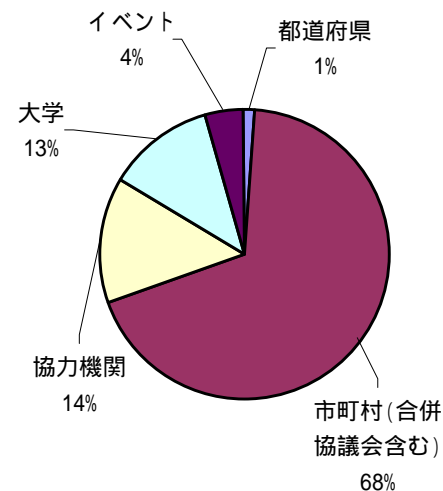
イベント、合併自治体、大学、特殊法人等 約600タイトル、416GB

最も古いコレクション

- 2000年9月13日収集「大地の芸術祭」

イベントコレクションの特徴

- 文化的、国民的価値の高いサイトを多数収録
 - FIFAワールドカップ日本組織委員会
 - 2002年「日本年」「中国年」実行委員会 等
- イベント終了に伴い、ネット上から消失したサイトも
 - FIFAワールドカップ日本組織委員会
 - やまがた花咲かフェア'02 等



協力機関コレクション(続き)



- **大学コレクションの特徴**

- 博士論文(九州芸術工科大学)、研究報告等、学術文献を収蔵

- **自治体コレクションの特徴**

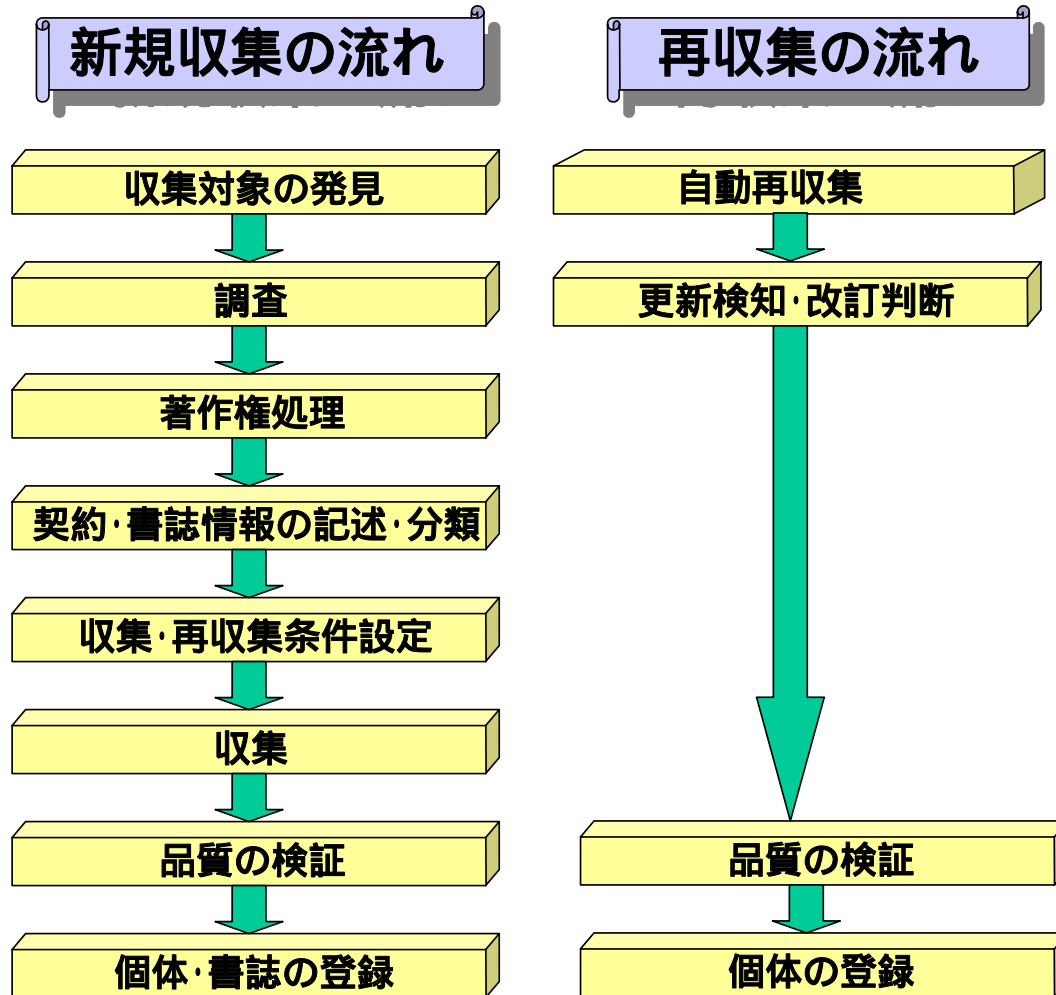
- 歴史的に価値の高いサイトを収録
 - 上九一色村 等
- 合併に伴いネット上から消失したサイトも
 - 万場町、若草町、穂積町 等々

- **特殊法人等コレクションの特徴**

- 年報、事業報告書、財務情報等
- 学術的に価値の高い資料も
 - (例)宇宙開発事業団・・・宇宙開発関連画像・動画、衛星軌道、気象観測データ、宇宙法、報告書、資料 等々
- 解散に伴いネット上から消失したサイトも
 - 新世代通信網実験協議会(BBCC)等



WARP収集業務の流れ



WARPの課題

1. **著作権**：複製権、公衆送信権等の許諾
2. **情報発見**：収集対象発見は手作業
3. **粒度**：選択するからには単位概念が必要(サイトとは?)
4. **セレクション**：何が重要か? 内容で区別すべきなのか?
5. **収集性能**：HTMLのバージョン、リンクの深さ...、深層ウェブは?
6. **品質管理**：完全な見え方というのは何か? 確認方法は?
7. **再収集**：あらゆる版の収集は不可能。妥当な頻度?
8. **メタデータ**：どのように生成するのか。必要なデータ項目は?
9. **識別子**：情報に対する安定的なナビゲーションのために
10. **アーカイブ内ナビゲーション**：タイトル、URL、全文検索...
11. **原本性保証**：同一性、真正性 再現性との関係は?

深層ウェブへはナビゲーション

外部資源へのナビゲーション： Dnavi（約9000件）
機械的に収集不可能なインターネット情報資源へは案内だけ



当館ウェブ・アーカイブの今後

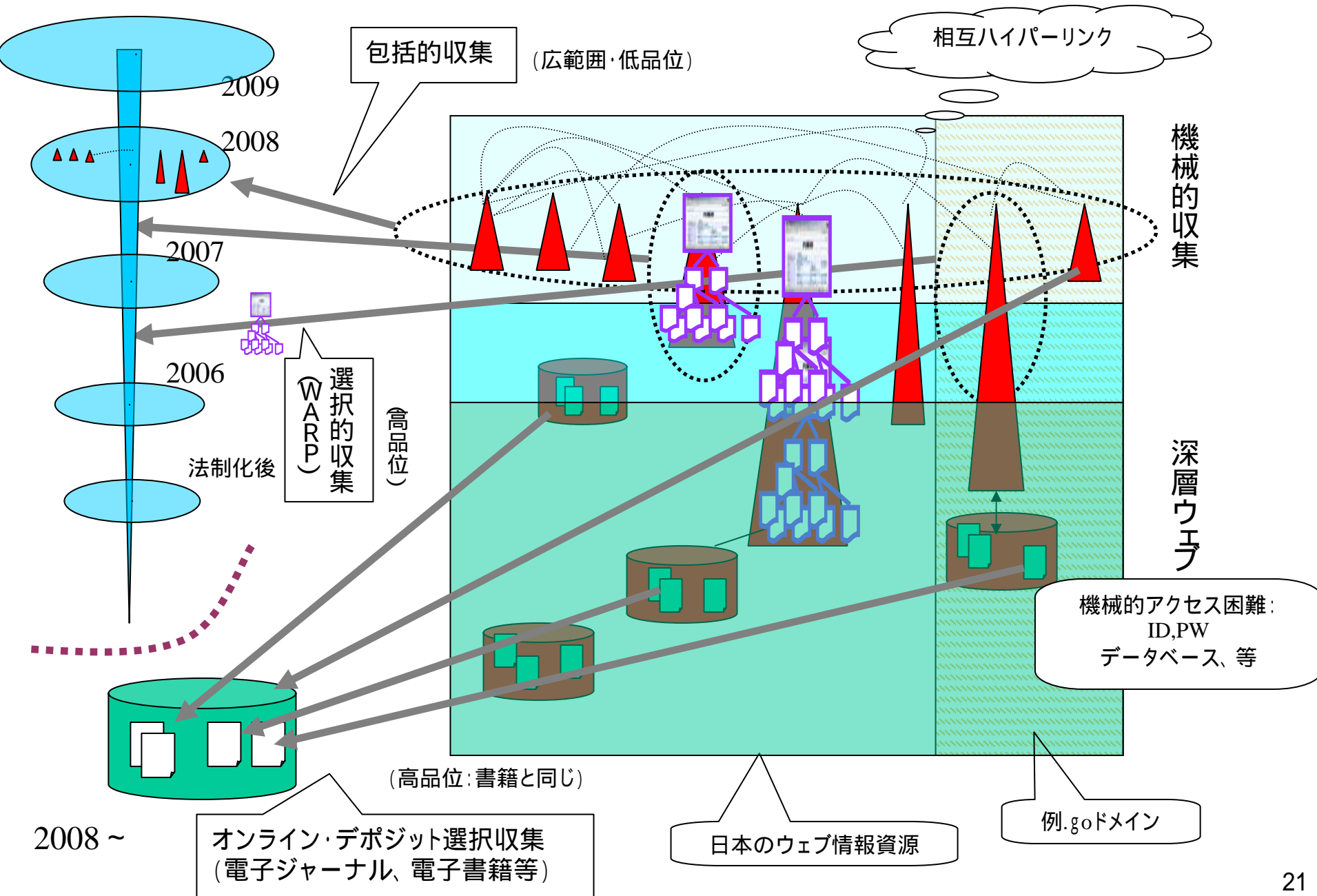
• WARPの継続拡張

- 電子雑誌コレクション
 - さらなる拡充へ
- 政府機関コレクション
 - 政府のe-Japan 戦略との連携を強化、拡大へ
- 協力機関コレクション
 - 統廃合の激しい自治体、大学、特殊法人等に注力、文化資産としてのWeb情報の消失を未然に防止

• WARPで不可能なことに対して

- **機械的一括収集**(ハイパーリンクで構成されたウェブ文化の全体保存: 包括的収集)
- **電子納本的収集**(論文や電子書籍など、著作単位での収集・保管・提供: 選択的収集)

ネットワーク系電子出版物の収集(概念図)



新しい収集方法における課題

- 機械的一括収集

権利制限 納本制度審議会で検討中

アーカイブ内ナビゲーション

- メタデータの付与

- 格納仕様、再生機能仕様、大容量検索

大容量データの運用

- 大容量ストレージ構築運用、高速圧縮・解凍

- 電子納本（知的著作単位ごとの収集）

長期保存 = 再現性の確保

（マイグレーション、エミュレーション等）

今年度の実施予定項目

ウェブ・アーカイビングに関するH16年度の国立国会図書館の実施予定項目

- ネットワーク系電子情報収集に関する法制化準備
- W A R P システム環境のリプレース
- 中央府省等のアーカイブ拡充