インターネット資料収集保存事業（WARP）
～ウェブアーカイブの活用と課題～

NDLデジタルライブラリーカフェ　2020/12/10

国立国会図書館電子図書館課
高峯 康世
warp@ndl.go.jp

1. インターネット資料収集保存事業（WARP）とは

2. WARPの現状

3. 収集データの活用

4. WARPの今後・課題

1. WARPとは：概要

Q： インターネット資料収集保存事業（WARP）って何？

Ａ： 国立国会図書館が行っている、国内のウェブサイトを収集・保存して後世へ伝えていく事業です。

Q： どんなウェブサイトを保存しているの？

Ａ： 政府や自治体、独立行政法人、国公立大学など公的機関のウェブサイトを、国立国会図書館法に基づいて網羅的に収集しています。また、民間のウェブサイトについては、許諾に基づいて、政党や私立大学、公益法人など様々なウェブサイトを収集しています。

Q： どのように役に立つの？

Ａ： インターネット上の情報は日々更新されるため、紙の資料に比べ消失しやすいという特徴があります。消失してしまった過去のページや報告書なども、WARPが保存していれば将来にわたって見ることができます。

1. WARPとは：沿革

2002年：　実験事業として開始

2006年：　本格事業化

2010年：　国立国会図書館法改正（施行）
→公的機関サイトの網羅的な収集が可能に

2015年：　収集タイトルが10,000件を突破

1. WARPとは：トップページ

- キーワード検索
- コレクション検索
- 今月の特集
- アクセスランキング

など

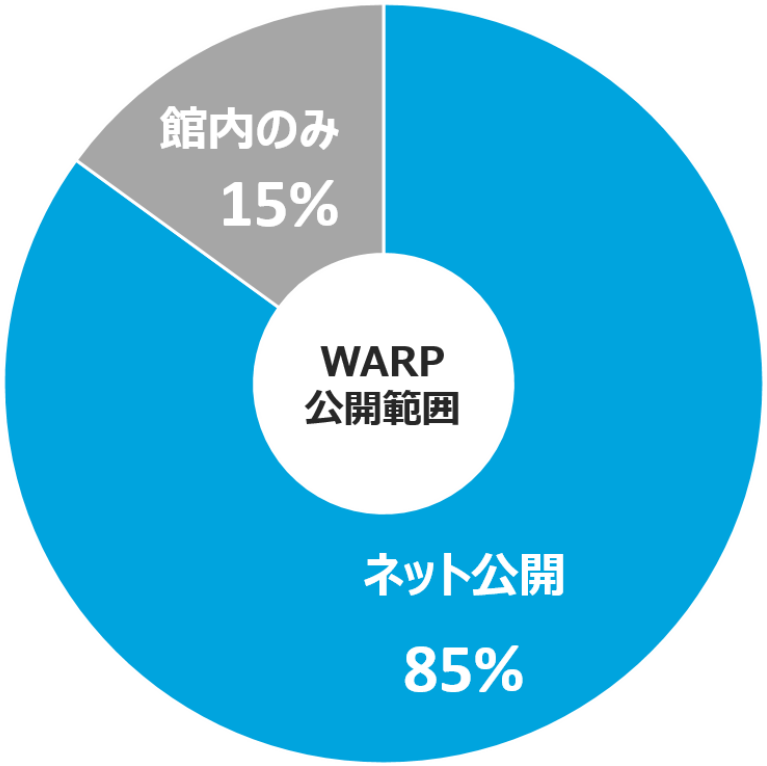
https://warp.da.ndl.go.jp/

1. WARPとは：収集概況

区分	根拠	対象	収集頻度	タイトル数	容量	ファイル数
公的機関	法律	国の機関	月1回	5,800	1.9 PB	95億 ファイル
		地方自治体 独立行政法人 国公立大学	年4回			
民間	契約	私立大学、公益法人、学協会、第三セクター、業界団体、文化施設、政党、インフラ、イベント、震災など	年1～4回	6,900		

2020年10月末現在

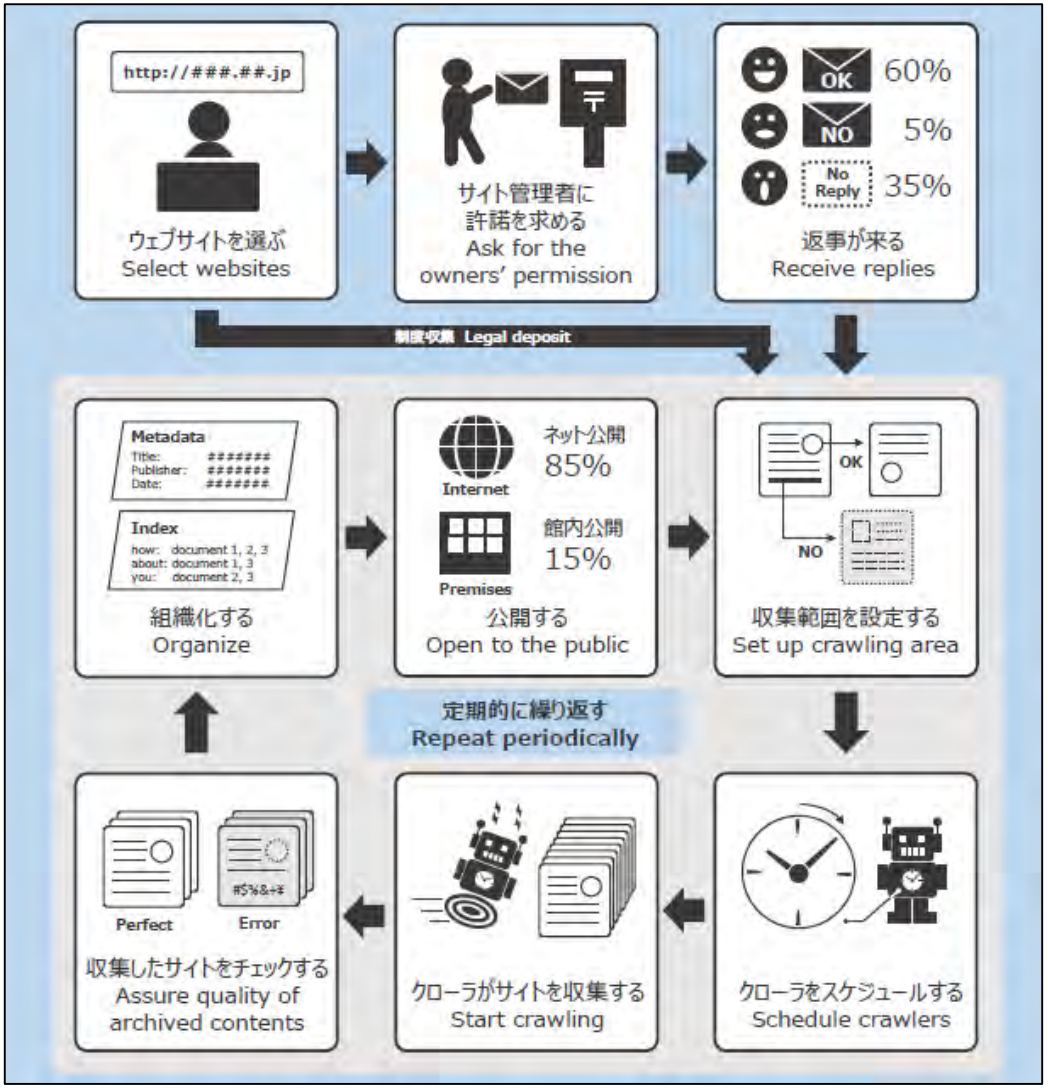
1. WARPとは：公開範囲



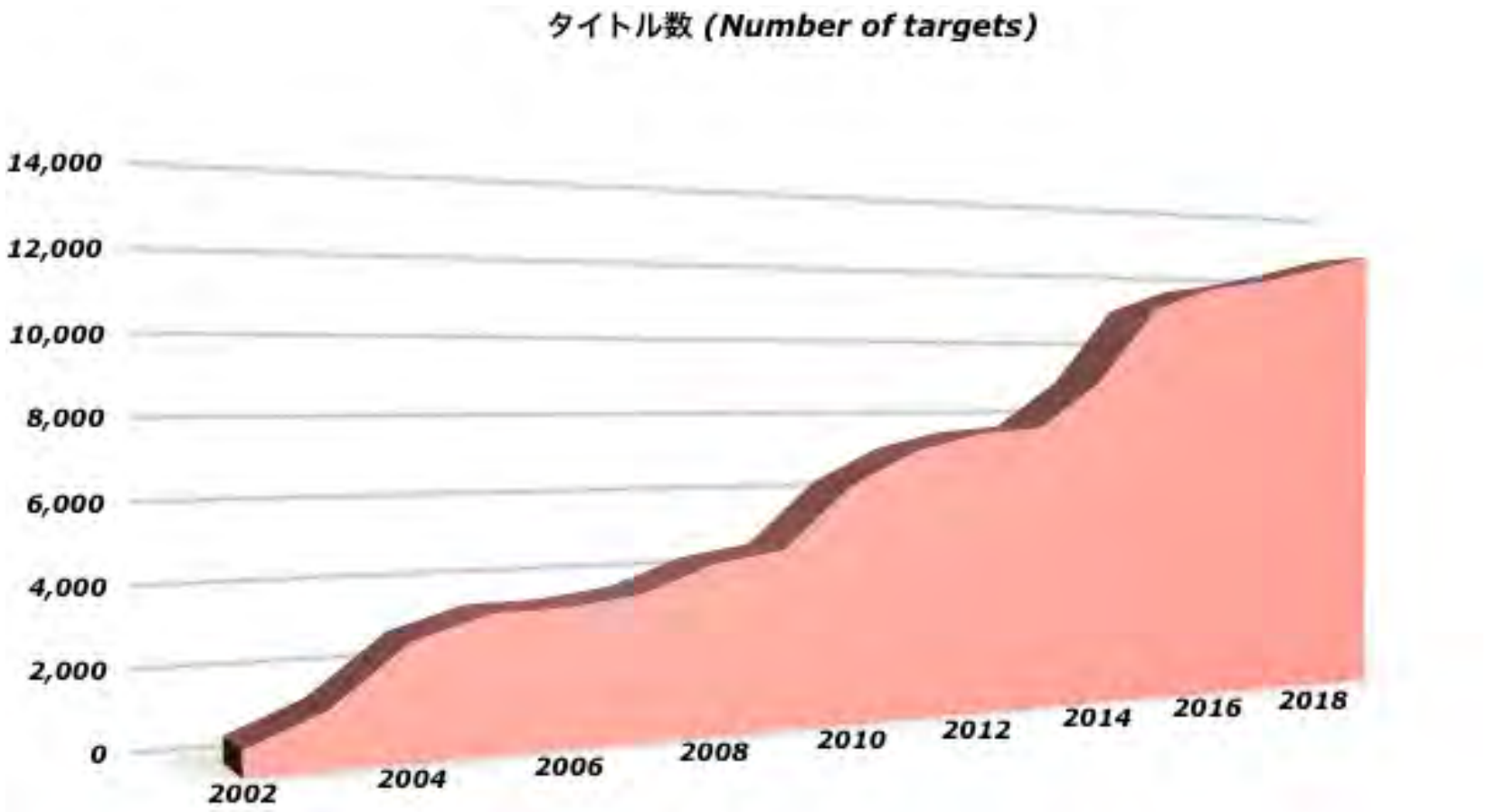
公的機関・民間共に
インターネット公開のためには
許諾が必要

タイトル単位では8割以上
かなりのオープンアーカイブ

1. WARPとは：ワークフロー

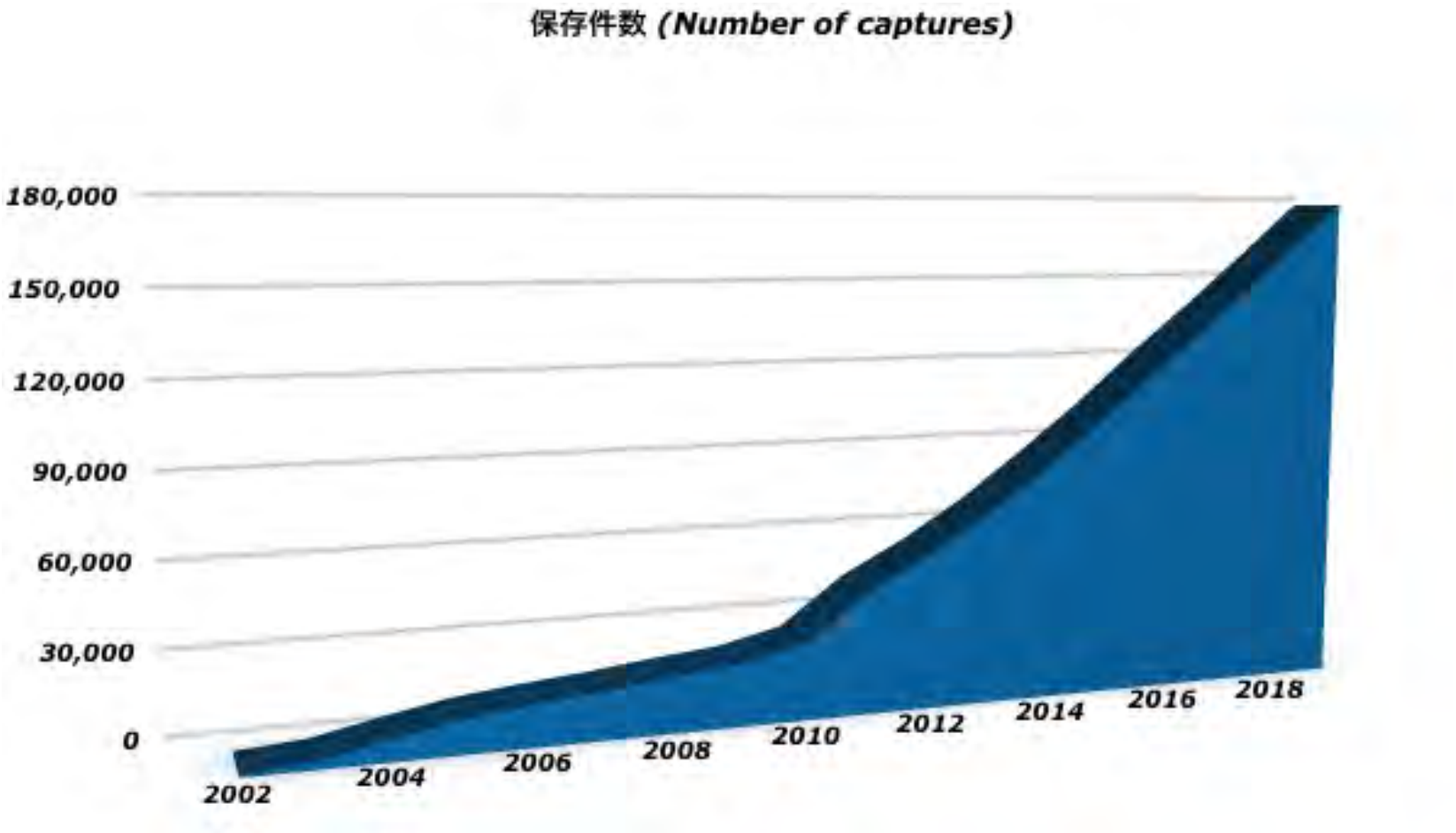


2. 現状：収集タイトル数の推移



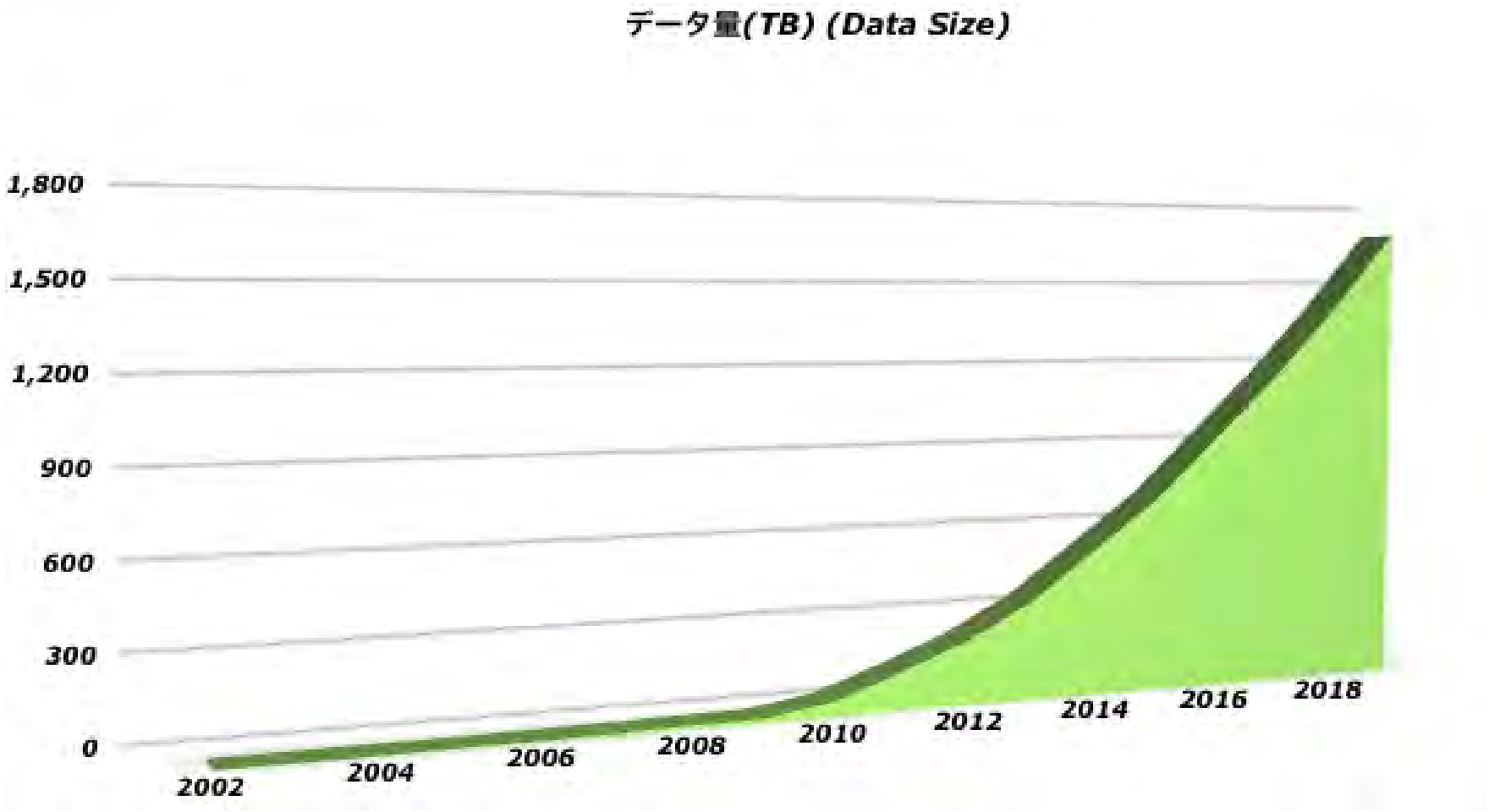
WARPについて：統計
https://warp.da.ndl.go.jp/info/WARP_statistic.html

2. 現状：収集件数の推移



WARPについて：統計
https://warp.da.ndl.go.jp/info/WARP_statistic.html

2. 現状：データ量の推移

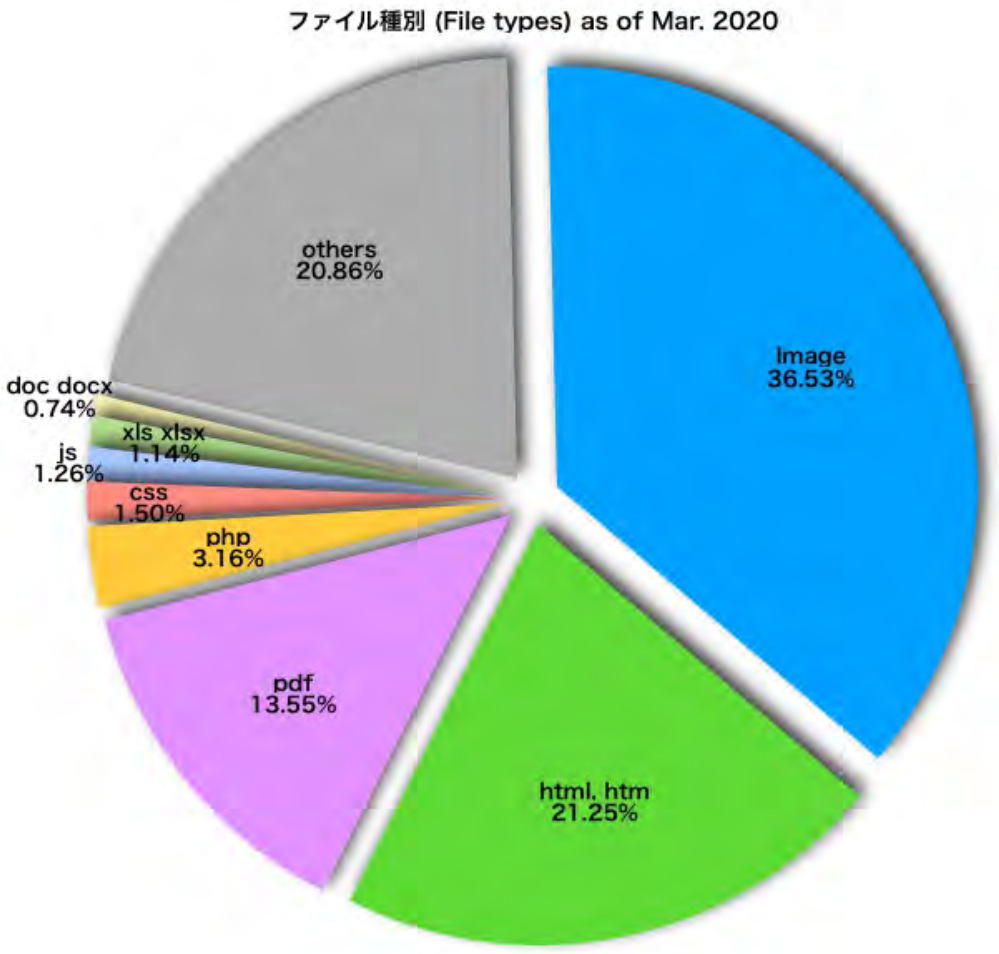


WARPについて：統計
https://warp.da.ndl.go.jp/info/WARP_statistic.html

2. 現状：収集ファイルの種別割合

ファイル種別	ファイル数	割合
画像 (jpg, png, tif等)	3,121,183,815	36.53%
html, htm	1,815,579,824	21.25%
pdf	1,157,233,204	13.55%
php	269,920,761	3.16%
css	128,562,887	1.50%
js	107,786,843	1.26%
xls, xlsx	97,223,641	1.14%
doc, docx	63,490,449	0.74%
その他	1,782,432,446	20.86%

(2020年3月末)



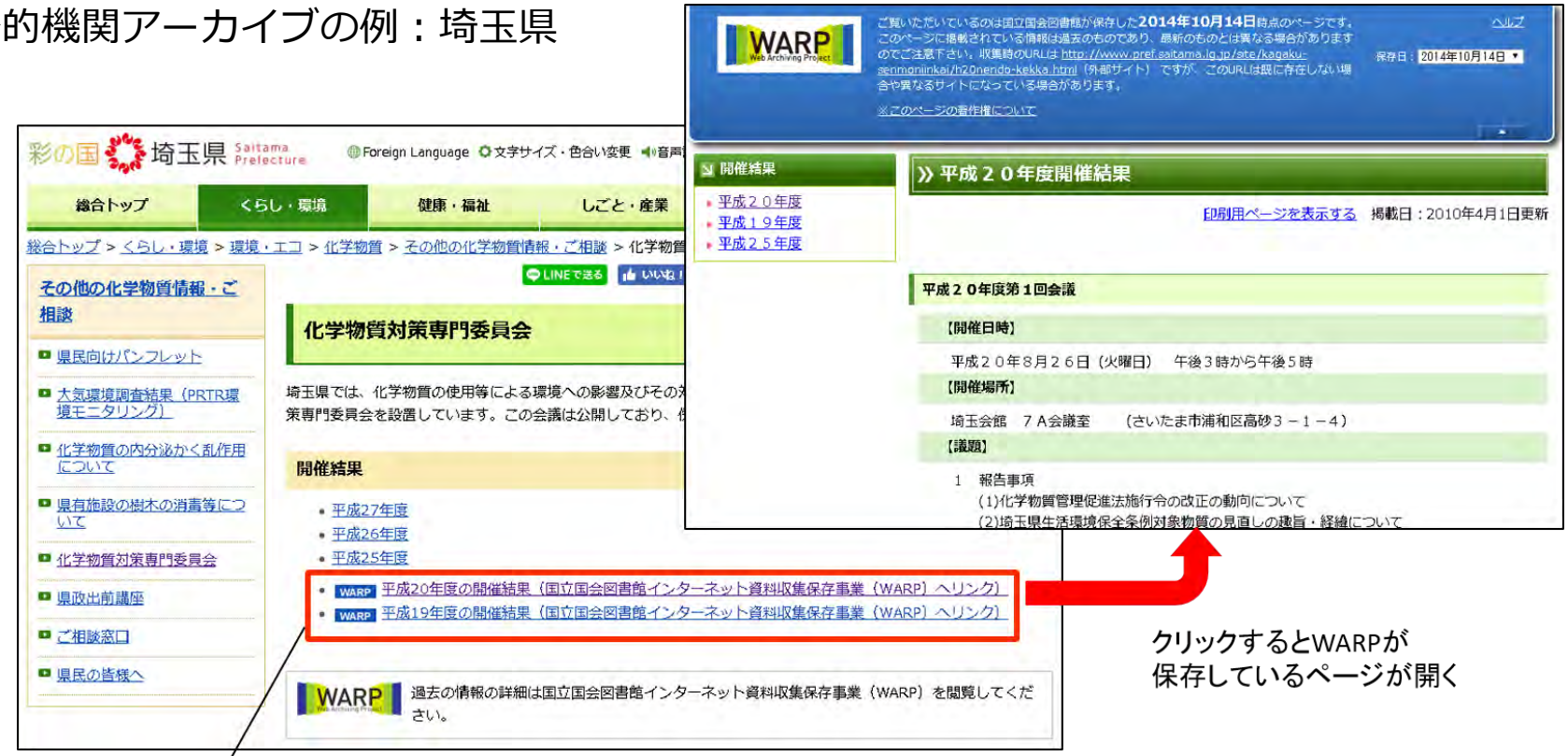
WARPについて：統計
https://warp.da.ndl.go.jp/info/WARP_statistic.html

2. 現状：コレクション

- 公的機関
  - 国の機関、地方自治体、独立行政法人、国公立大学など
- 民間
  - 私立大学
  - 国政政党
  - 公益法人
  - 学協会
  - 第三セクター
  - 業界団体
  - 科研費（科学研究費助成事業）
  - 文化施設：博物館、美術館、動物園など
  - インフラ：道路、空港など
  - イベント：国際会議、スポーツ、お祭り、芸術祭など
  - 東日本大震災関連：電力会社、被災者支援団体など
  - 東京2020五輪関連：スポーツ団体など
  - 新型コロナウイルス感染症関連：医療関連学協会、業界団体

3. 収集データの活用：公的機関アーカイブ、パーマネントリンク

公的機関アーカイブの例：埼玉県



クリックするとWARPが
保存しているページが開く

遷移先のページを自機関サイトから
消してWARPにリンクする

3. 収集データの活用：公的機関アーカイブ、パーマネントリンク

公的機関によるアーカイブ活用事例

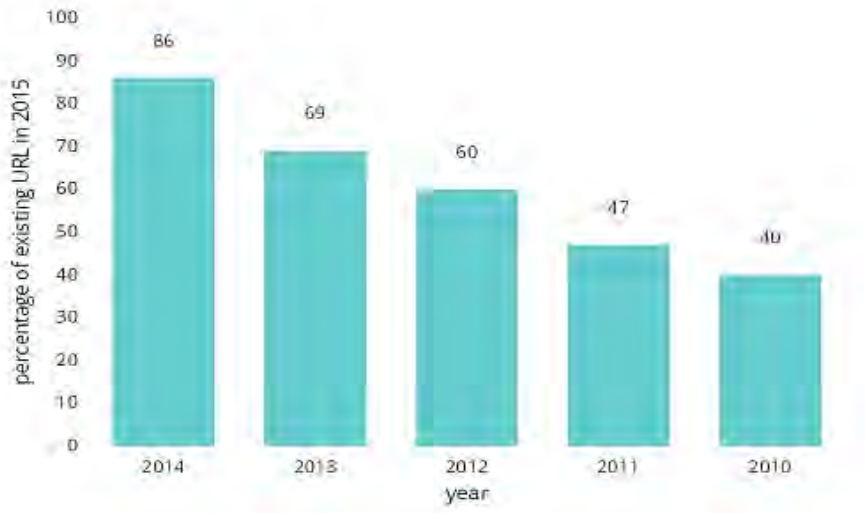
- 首相官邸「過去の官邸ホームページ」（第3次安倍内閣以前）
- 内閣府「行政刷新 過去の取り組み」（「事業仕分け」等）
- 財務省「予算」（平成28年度以前の予算関連資料）
- 群馬県「群馬県報」（平成26年以前の県報）

パーマネントリンクとしての活用

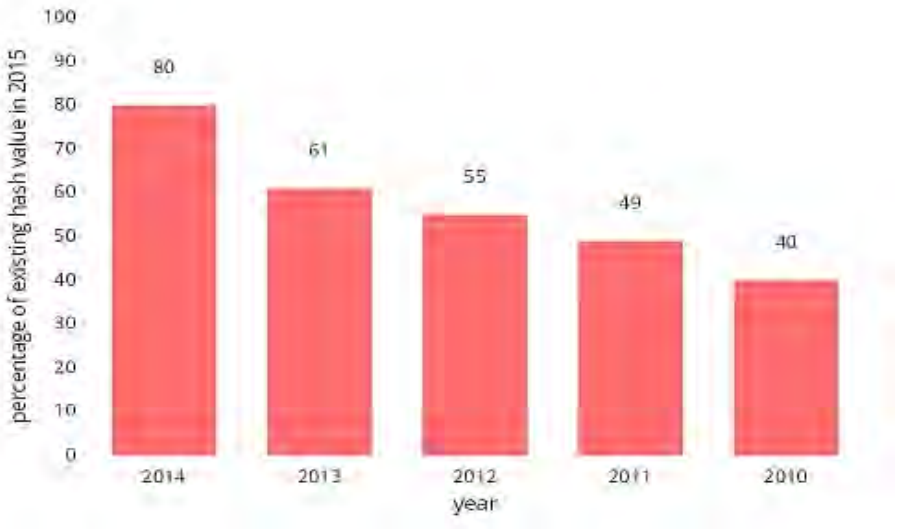
- 国立研究開発法人日本原子力研究開発機構（JAEA）
  「福島原子力事故関連情報アーカイブ（FNAA）」

今月の特集（2020年10月）「こんな所にWARPへのリンク 」
https://warp.da.ndl.go.jp/contents/special/special202010.html

3. 収集データの活用：ウェブコンテンツ残存率調査



URLの残存率

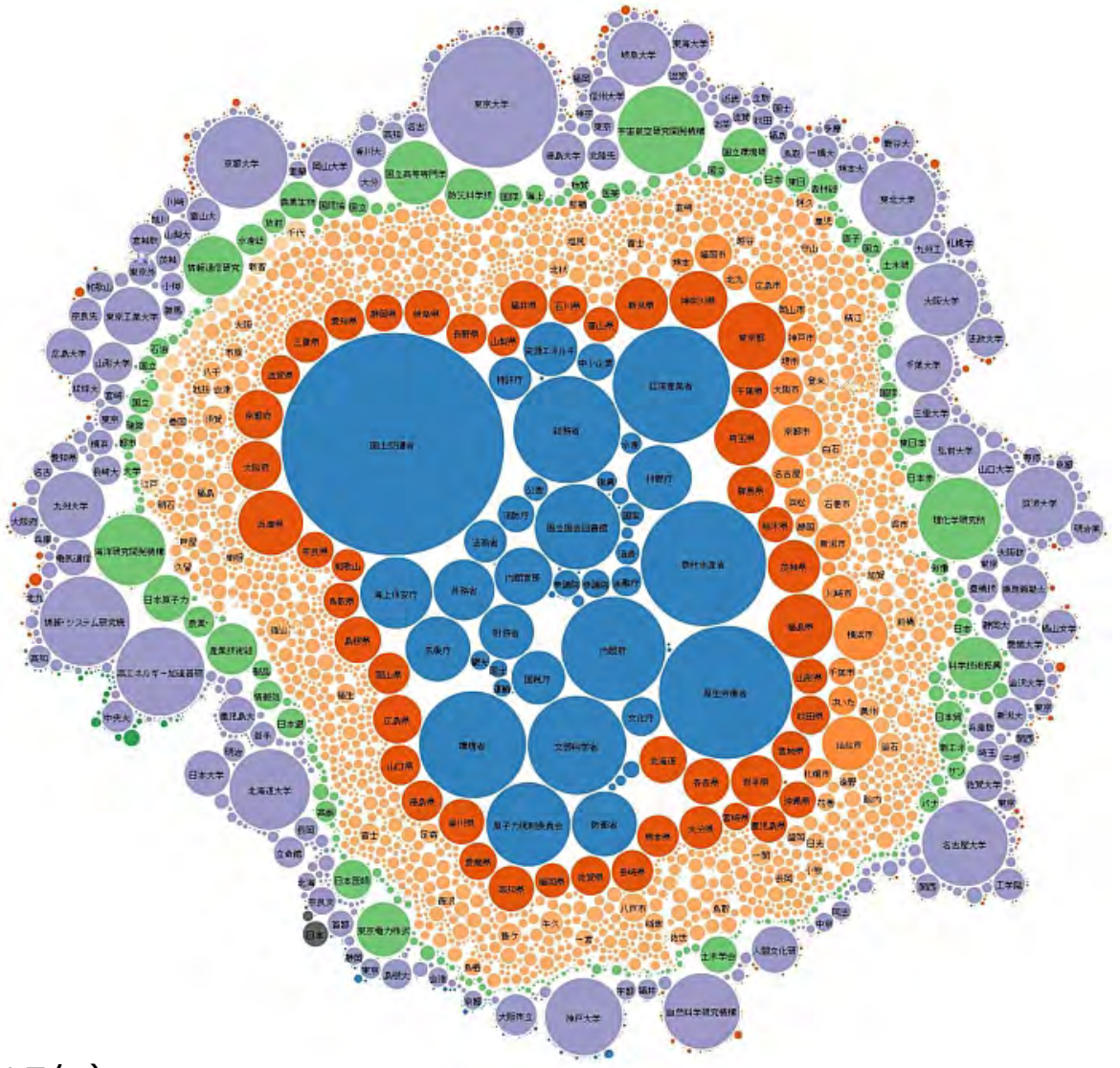


内容の残存率

国の機関サイトの残存率調査（2015年）
https://warp.da.ndl.go.jp/contents/reccommend/collection/linkrot.html
https://current.ndl.go.jp/e1757

3. 収集データの活用：保存サイトの可視化

収集量をタイトル・種別に
円の大きさで表して可視化



保存した1万サイトの可視化（2015年）
https://warp.da.ndl.go.jp/contents/reccommend/collection/bubble.html

4. WARPの今後・課題

- 収集コンテンツの偏り　→公的機関サイトが中心
- 民間サイトの収集拡大　→2020年度：新型コロナウイルス感染症流行関連業界団体、学協会
- 公的機関のSNS・動画収集
- ネット公開タイトルの拡大
- ストレージ：収集データ蓄積量拡大への対応
- 検索システムの改善
- WARPに関するデータの利活用

是非WARPを一度、ご利用ください！

https://warp.da.ndl.go.jp/