

# ウェブ情報アーカイブの利活用方法に関する 調査研究報告の概要

平成17年6月15日

株式会社 文化環境研究所

## 目的

---

- ウェブ情報のアーカイブ化促進に関する平成 16 年度実証実験と併行して、ウェブ情報アーカイブの利活用モデルについて検討するための調査研究を実施し、平成 17 年度以降の取組みに資する。

# 構成

---

## 1章：海外におけるウェブ情報アーカイブの利活用動向

- 既存文献、E-mail や直接訪問による取材

## 2章：ウェブ情報アーカイブの利活用に関する情報収集

- 関連機関・有識者へのヒアリング
  - 目的：ウェブ情報アーカイブに対する様々な分野での利活用に関するヒアリング
- 生活者に対するアンケート調査
  - 目的：ウェブ情報アーカイブに対する認知・関心の確認等
- 企業等、民間組織に対するアンケート調査
  - 目的：企業等、民間組織におけるウェブサイト上の情報の利活用状況の把握等

## 3章：ウェブ情報アーカイブの用途とそれに応じた特性の整理

## 終章：ウェブ情報アーカイブの確立に向けた今後の課題

- ウェブ情報アーカイブの利活用等、今後の課題について整理

# 海外におけるウェブ情報アーカイブの利活用動向調査の主な対象

## 1章 海外におけるウェブ情報アーカイブの利活用動向

国名	主体	アーカイブもしくはプロジェクト名称	アーカイブへのアクセス方法	方法
オーストリア	オーストリア国立図書館	AOLA	実験段階で公開されていない	E-mail
イギリス	英国図書館他、5つの機関が参加	UK ウェブ・アーカイビング・コンソーシアム	収集段階で公開されていない	E-mail
スウェーデン	スウェーデン王立図書館	Kulturarw <sup>3</sup>	図書館内のみでアクセス可能(インターネット上からアクセス不可)	E-mail
チェコ	チェコ国立図書館	WebArchiv	権利関係処理が済んでいる情報のみ図書館内でアクセス可能	E-mail
デンマーク	デンマーク王立図書館	netarchive.dk	インターネット検索まで可能。詳細内容を読むには王立図書館、オーフス州立大学図書館の定められたPCからアクセス	E-mail
ノルウェー	ノルウェー国立図書館	PARADIGMA	インターネットも含めて未公開	ヒアリング
フィンランド	フィンランド国立図書館	なし	インターネットも含めて未公開	E-mail
アメリカ	アメリカ議会図書館	MINERVA	インターネットからアクセス可能	ヒアリング
アメリカ	インターネットアーカイブ	インターネットアーカイブ	インターネットからユーザーインタフェース「ウェイバックマシン」を使用してアクセスする	E-mail
オーストラリア	オーストラリア国立図書館	PANDORA	インターネットからアクセス可能	ヒアリング

## 海外におけるウェブ情報アーカイブの利活用動向結果

---

### 1章 海外におけるウェブ情報アーカイブの利活用動向

- 各国ともウェブ情報アーカイブの目的を「収集・保存」に据えている。
- どのようにしたら正確に、大量かつすばやくウェブサイト上の情報を収集・保存できるかが議論の中心に置かれている。
- アクセスを含めた利活用に関して議論は意外と進んでいない。
- 現場レベルでは収集・保存をウェブ情報アーカイブの主目的としているため、利活用については別の次元の問題と認識している。

# 関連機関・有識者へのヒアリング結果

---

## 2章 ウェブ情報アーカイブの利活用に関する情報収集

- 日本は明確にウェブ情報アーカイブの分野において立ち遅れている事実を国民に広く知らしめる必要がある。
- 収集用のロボット開発には膨大な費用がかかる、そういったものをそれぞれの地方独自で開発するのではなく、汎用性のあるシステム構築が必要。
- 独自のシステムを構築、あるいは独自のサービスのみを提供するのではなく、海外の機関を含め連携できるようにする。先行してオープンスタンダードな環境を構築している組織・機関・団体に資金を提供して技術を取得し、それによって国際連携をはかれるシステムを構築するという方法もある。
- ウェブサイト上のページが改竄されていないことを誰が保証するのか。日本では公的機関による保証が必要ではないか。
- 権利関係がわからない素材や情報が多いため、現場での利活用の妨げになっている。
- ウェブ情報アーカイブがインターネット検索エンジンと区別が付かないというのは問題。構築されるウェブ情報アーカイブには専用のインタフェースが必要。自分が今どこにいるのか、確実にわかるインタフェースは必須。
- デジタル化情報時代に対応したパブリックユースできる素材の著作権管理について検討する必要がある。
- 重要なウェブサイト上の情報を預かるというビジネスモデルを取るのであれば、どういう責務を負い、また責務の解除について定める必要がある。
- 情報系でない研究者が研究するにはハードルが高い。データを多く扱うのに慣れている研究者でないといけない。
- 国との役割分担や守備範囲について今後、詳細な調整が必要に思う。ウェブ情報アーカイブの構築について、国としてどのように考えているのか、またそれぞれの役割分担についても、検討の余地がある。
- システムで収集できていない情報がたくさんあるので、他国も含め、同様のテーマでウェブ情報アーカイブ組織機関があれば連携したい。

# 生活者に対するアンケート調査結果

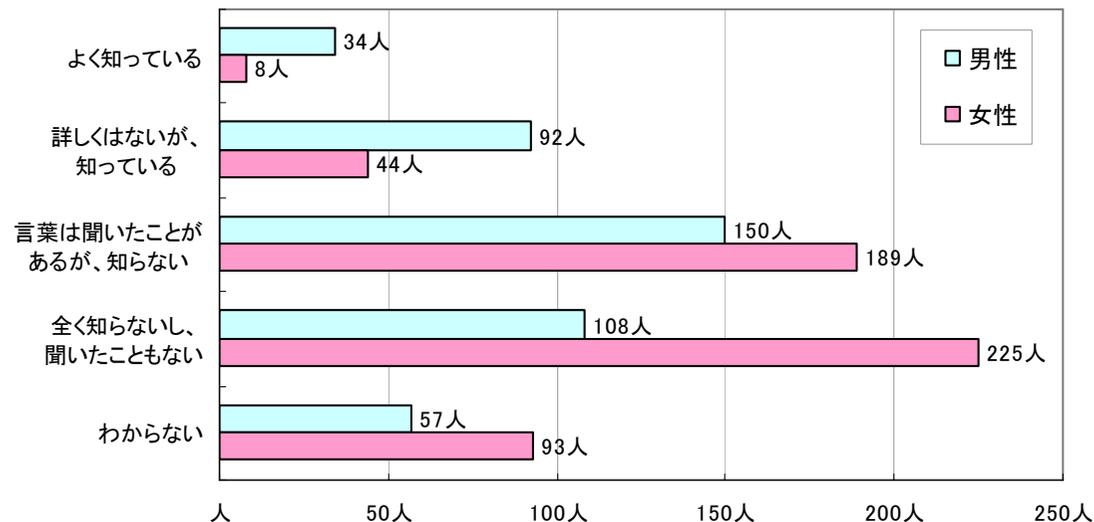
## 2章 ウェブ情報アーカイブの利活用に関する情報収集

### 「ウェブアーカイブ」という言葉の認知状況

全 体：「知っている（よく知っている＋詳しくはないが知っている）」は17.8%、「知らない（言葉は聞いたことがあるが、知らない＋全く知らないし、聞いたこともない）」は67.2%となった。

性 別：男性は「言葉は聞いたことがあるが、知らない（150人）」が、女性は「全く知らないし、聞いたこともない（225人）」が最も多い。「知っている」層は、男性が126人（男性回答者の28.6%）、女性は52人（女性回答者の9.3%）となる。

利用歴：「知らない」層の46.1%（310人）が、インターネット利用歴「3年以上～6年未満」の層となっている。インターネットの利用歴が長い程、認知度が高い層が多い傾向が見られる。



- 調査実施方法：インターネットアンケート方式
- 実施期間：2005年3月16日～17日（約20時間）
- 収集サンプル数：1,000サンプル

# 企業・民間組織に対するアンケート調査結果（1）

## 2章 ウェブ情報アーカイブの利活用に関する情報収集

自社の企業・組織活動から見て、ウェブ上で提供して欲しいサービスや情報（データ）

- 45社中14社（37.8%）が具体的に回答。
- 「体系化した情報」と「データ・資料情報」に分類できる回答が多く、統合・整理された情報を求める傾向にある。
- 14社中12社が、「提供されるなら有償でも利用する」と回答。

提供して欲しい情報の内容

種類	コンテンツ内容
体系化した情報	各種補助金等の情報を一元管理したデータベース、世界各国のマーケット情報新製品情報を体系化した情報（リアルタイムな情報）
データ・資料情報	標準化ドキュメント、国際会議資料、学会論文、各種調査統計資料（※）、マーケティング活動に利用できる幅広いデータ（※）、インターネット利用等（特に広告）についての統計データ（できれば第三者からのもの）、歴史書百科事典的なもの
行政情報	各行政のページ（弊社防災ページでリンクしているため） 政府自治体議会関連（※）
評価指標	企業情報（評価）、株価情報
その他	郵便番号検索のウェブサービス、地方での低料金高速回線の接続サービス、ウェブ上では信頼性のある情報（官庁・企業等）と不確かな情報（個人・任意の団体のHP）が混在しているため、信頼性のある情報のみを検索できるポータルサイトの充実

（※）が付いている情報は「有償であれば利用しない」と回答したもの。付いていない情報は「有償でも利用したい」と回答。

- 調査実施方法：郵送回答方式
- 調査実施主体：株式会社文化環境研究所
- 実施期間：2005年3月10日～25日（約2週間）
- 調査対象企業数：300
- 収集サンプル数：45サンプル ※有効回答率：15%（45/300）

## 企業・民間組織に対するアンケート調査結果（2）

### 2章 ウェブ情報アーカイブの利活用に関する情報収集

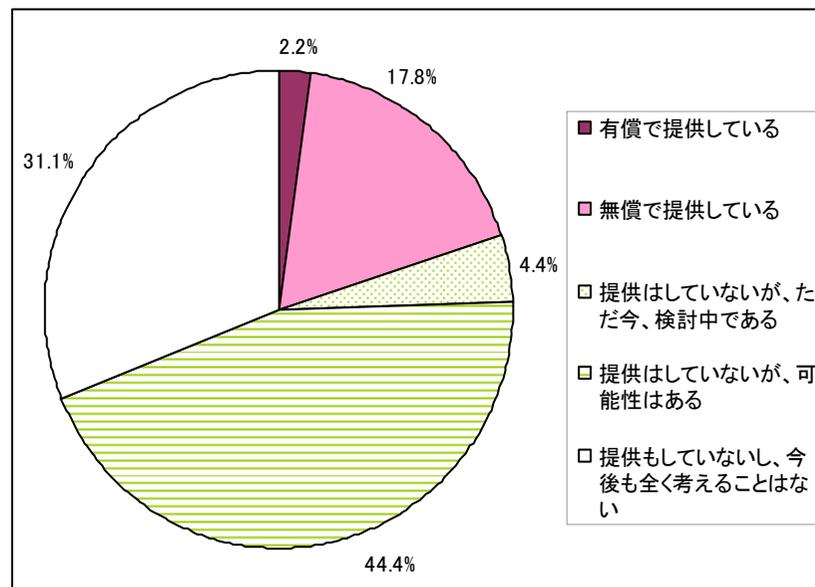
#### 企業・組織活動に深く関連しているウェブサイト上の過去の情報を含めた“コレクション（収集・整理・ウェブサイト上での提供）”の有無

- 45社中9社（20.0%）が提供していると回答。
- 検討中、可能性はあると回答した企業は22社（48.8%）。
- 提供している9社のうち8社が無償で提供している。
- 提供コンテンツとしては、「（第3者機関による）報告書」「オリジナルコンテンツ」「リンク集」があげられる。

#### ウェブ情報“コレクション”の内容

種類	提供コンテンツ
報告書	有価証券報告書、世論調査内容、マニュアルレポート
オリジナルコンテンツ	未来への扉（イノベーションストーリー）、ブランドコンテンツ ハチエモン（関西テレビのキャラクター）PR版動画配達
リンク集	関係機関・学会関連サイトのリンク集
その他	過去の全ニュースリリース、デパート情報

※ 全て無償提供の回答内容。有償提供と回答した企業は記入なし



ウェブ情報“コレクション”の有無

# ウェブ情報アーカイブ類型

---

## 3章 ウェブ情報アーカイブの用途とそれに応じた特性の整理

### デジタルアーカイブをウェブサイト上で提供

- 岐阜女子大学デジタルミュージアム
- デジタル岡山大百科（岡山県立図書館）
- 文化遺産オンライン

### 時系列で収集される情報をアーカイブ化し、分析等のサービスを提供

- 超大規模ウェブソリューション（東京大学生産技術研究所）
- 言語天文台（長岡技術科学大学）

### インターネット空間の一部、あるいは全体を時系列に保存・再現

- インターネット資源選択的蓄積実験事業（国立国会図書館）
- 農学情報資源システム（農林水産研究情報センター）

※ちなみに海外では、インターネットアーカイブがこの類型にあてはまる

# 想定されるウェブアーカイブの用途モデル

## 3章 ウェブ情報アーカイブの用途とそれに応じた特性の整理

モデル	想定される主体	想定される利用者	特徴	主体が抱えるであろうアーカイブ構築に関わる課題
文化保存・記録	<ul style="list-style-type: none"> <li>公的機関</li> <li>学術機関</li> <li>一般</li> </ul> 等	<ul style="list-style-type: none"> <li>公的機関</li> <li>学術機関</li> <li>一般</li> </ul> 等	文化の記録という側面よりなるべくオリジナルに忠実なウェブの収集・保存が必要。公的収集・保存の役割。	<ul style="list-style-type: none"> <li>商利用を含めて利用範囲をどのように決定するのか。またその利用方法(施設内、インターネットなど)</li> <li>収集対象のプライバシーおよび権利をどのように保護するのか。</li> <li>収集データの永続的保存体制の確立。</li> </ul>
クリッピング	<ul style="list-style-type: none"> <li>企業</li> </ul> 等	<ul style="list-style-type: none"> <li>一般</li> <li>調査会社</li> <li>公的機関</li> <li>学術機関</li> </ul> 等	ニュース、新聞、テレビ等。インターネット、企業のプレスリリース等。特にウェブでは先端技術に相当する情報がリアルタイムに近く発表される。	<ul style="list-style-type: none"> <li>意識的に削除される情報の扱い。</li> <li>報道の記録という側面と内容の信頼性のバランスの評価。</li> <li>再利用は困難。</li> <li>プライバシー、個人情報保護の確立と権利処理と利用環境の整備。</li> </ul>
マーケティング	<ul style="list-style-type: none"> <li>企業</li> <li>調査会社</li> </ul> 等	<ul style="list-style-type: none"> <li>一般</li> <li>企業</li> <li>調査会社</li> <li>公的機関</li> <li>学術機関</li> </ul> 等	社会、市場の動向の分析。分析しやすいように収集時に情報を加工して保存する。	<ul style="list-style-type: none"> <li>目的に応じた分析手法の確立。</li> <li>分析手法に応じた収集・保存手法の確立。</li> <li>収集データの信頼性の評価。</li> <li>収集データの再利用環境の確立。</li> </ul>
モニタリング	<ul style="list-style-type: none"> <li>公的機関</li> <li>学術機関</li> </ul> 等	<ul style="list-style-type: none"> <li>公的機関</li> <li>学術機関</li> </ul> 等	環境指標、気候変動等の事実としての情報および報道、一般社会での受け止められ方等の収集。	<ul style="list-style-type: none"> <li>定性的データと一般生活者のデータとの組み合わせ手法。</li> <li>定性的データ以外の収集データの信頼性の評価。</li> <li>モニタリング対象地域・対象からの継続的データ収集</li> </ul>
教育	<ul style="list-style-type: none"> <li>教育機関</li> <li>公的機関</li> <li>学術機関</li> </ul> 等	<ul style="list-style-type: none"> <li>教育機関</li> <li>公的機関</li> <li>学術機関</li> </ul> 等	教育、教材作成の素材として加工して再利用する。	<ul style="list-style-type: none"> <li>教材として再利用するための権利処理。</li> <li>二次利用して作られた成果物の利用環境の整備。</li> <li>指導要領との整合性の確保。</li> <li>素材提供環境としてのアーカイブの確立</li> </ul>

# ウェブ情報アーカイブ用途モデル構築に共通した課題

## 3章 ウェブ情報アーカイブの用途とそれに応じた特性の整理

課題の区分	課題	備考
目的に対する課題	<ul style="list-style-type: none"><li>・ 収集・保存の範囲をどのように決定するのか。</li></ul>	<ul style="list-style-type: none"><li>・ 目的を明確化し目的に応じた収集・保存範囲を決定する</li></ul>
技術的課題	<ul style="list-style-type: none"><li>・ 深層ウェブをどのように取り込むのか。</li><li>・ 大規模な情報をどのように検索するのか。</li><li>・ 収集時の先端技術による表示の再現性をどのように確保するのか。</li><li>・ 収集データの信頼性の保証</li></ul>	<ul style="list-style-type: none"><li>・ 目的に応じ技術的課題が達成されるべき範囲を決定</li></ul>
権利保護・権利利用	<ul style="list-style-type: none"><li>・ 収集時における対象の権利の保護とそれに応じた収集</li><li>・ 収集された対象の利用時における権利の保護と利用に応じた権利の運用</li></ul>	<ul style="list-style-type: none"><li>・ 個人の権利をいかに保護するか</li><li>・ 公的価値をもつデータをいかに公的に利用できるようにするか</li></ul>
社会的課題	<ul style="list-style-type: none"><li>・ 国家的見地・文化的見地より収集するものと社会的・市場的価値により収集・利用するものの区別</li></ul>	<ul style="list-style-type: none"><li>・ 公的利活用と事業的(企業的)利活用の区別</li><li>・ 公的利活用の推進および事業的利活用の推進</li></ul>

# ユーザーインタフェースのありかたについて（1）

---

## 3章 ウェブ情報アーカイブの用途とそれに応じた特性の整理

### 検索結果の表示

- 検索の結果、テキストや関連するリンク、時間の前後関係等の表示方法の検討
- 目的にあった情報を「時間・インターネット空間・事項」と関連付けた表示方法の検討
- 色使いを始め、フォントや画面の動きなど「デザイン」面の充実

### 時間・期間の指定

- 時間や期間の指定方法の検討
- 時間概念の利用方法について、人間工学、認知工学等の研究からの検討

## ユーザーインタフェースのありかたについて（2）

---

### 3章 ウェブ情報アーカイブの用途とそれに応じた特性の整理

#### 全体と細部

- 他のウェブサイトやページへのリンク等を同時に表示する場合、情報をクラス分けできるようなツールの検討

#### ディレクトリ、分類

- ユーザーインタフェースの有効性を高めるため、ディレクトリの体系付け
- 保存のためのディレクトリと利用のためのディレクトリの違いの認識
- ウェブ情報アーカイブの実情に即したディレクトリ構成

#### ユーザーインタフェースの研究開発

- 認知工学、人間工学、情報デザイン等、人の行動や活動を調査研究している専門家による検討

# ウェブ情報アーカイブ利活用促進に向けた対応の方向性

## 終章 ウェブ情報アーカイブの確立に向けた今後の課題

利活用促進のための課題	対応の方向性
ウェブ情報アーカイブの普及啓発	一般生活者にウェブ情報アーカイブの社会的必要性、利用法等を広く伝える。
ウェブ情報アーカイブの基盤技術の確立	収集・保存、検索等基盤技術の確立。
広範囲での利用体制の確立	図書館、自治体、学校、博物館等社会的に広範囲にわたる機関での利用を想定したシステム開発体制の確立。オープンソースによる開発の推進等。共通システムを利用できる体制および組織のネットワーク確立。
信頼性の保証	内容に対する信頼性、改竄されていないことを示す信頼性を付与する仕組みの研究・開発。日本では公的機関の役割が重要になることが予想される。
著作権・プライバシー等の権利の保護と利活用の両立	著作権・プライバシー等の権利の保護を確保しつつ、利活用の促進を図る方策を検討していくことが重要。
巨大で複雑な情報空間におけるユーザーインタフェースの研究開発	ウェブ情報アーカイブのような巨大で複雑な情報空間において、人はどのように行動するかといった側面からの研究およびその成果に基づくユーザーインタフェースの研究開発。
共同研究の推進	ウェブ情報アーカイブはきわめて多岐の分野に関わるため、これらの研究者が共同で研究開発を進める必要がある。たとえばエンジニアリングとデザインのコラボレーションは重要。