資料6－1

ウェブ情報のアーカイブ化促進に資する実証実験　概要（案）

平成16年7月30日

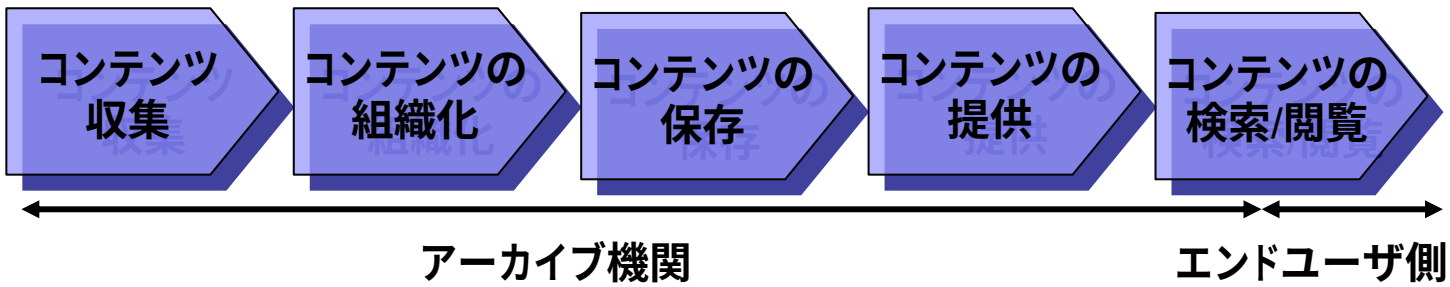
総務省

本実証実験の目的

- ウェブ情報にはデジタル時代の知識･文化が結集
- しかしながら、ウェブ情報は日々の更新、消去により散逸

⇒本実証実験により、デジタル時代の貴重な文化遺産であるウェブ情報のアーカイブ化及びその利用を促進するための技術･仕組みを構築･実証

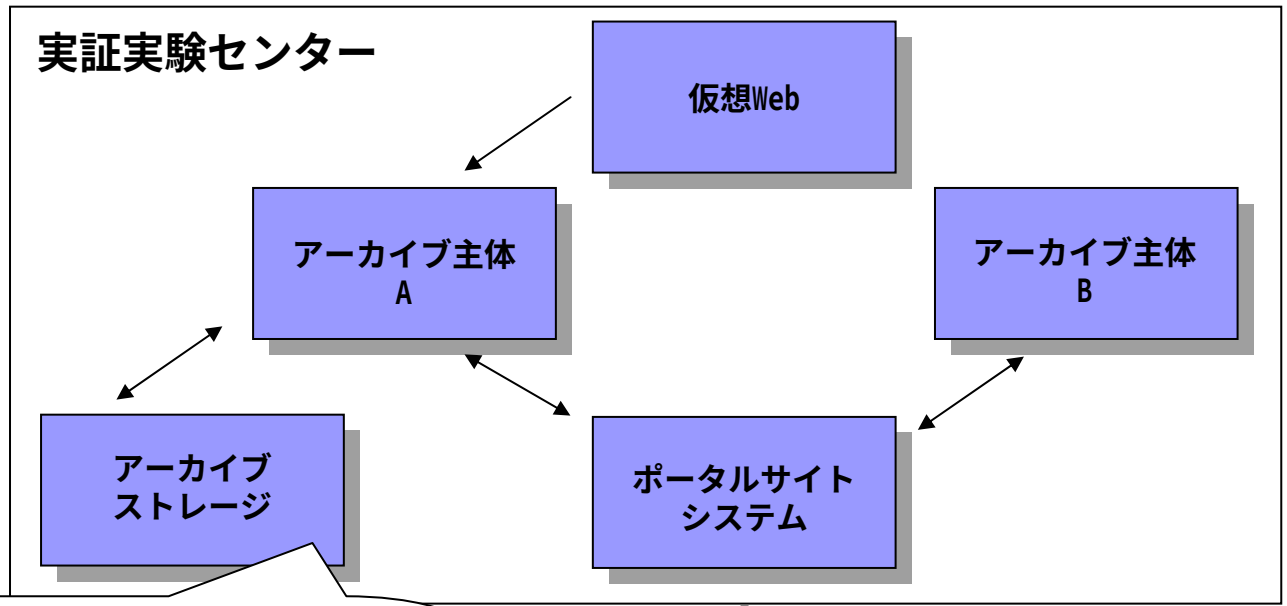
平成16年度実証実験の概要



16年度実施項目（案）

(1) 異なるアーカイブ間の横断検索を実現する汎用的なメタデータ体系の確立
(2) メタデータに基づく時系列別検索、特定コンテンツに係る時系列別表示の実現
(3) コストの縮減、長期的な保存を可能とするアーカイブ技術の開発等
(4) 同一のウェブ情報を重複して保存することを回避する技術に関する調査研究
(5) 収集範囲拡大に関する調査研究

実験システム全体構成図



(3) コストの縮減、長期的な保存を可能とするアーカイブ技術の開発等

(1) 異なるアーカイブ間の横断検索を実現する汎用的なメタデータ体系の確立

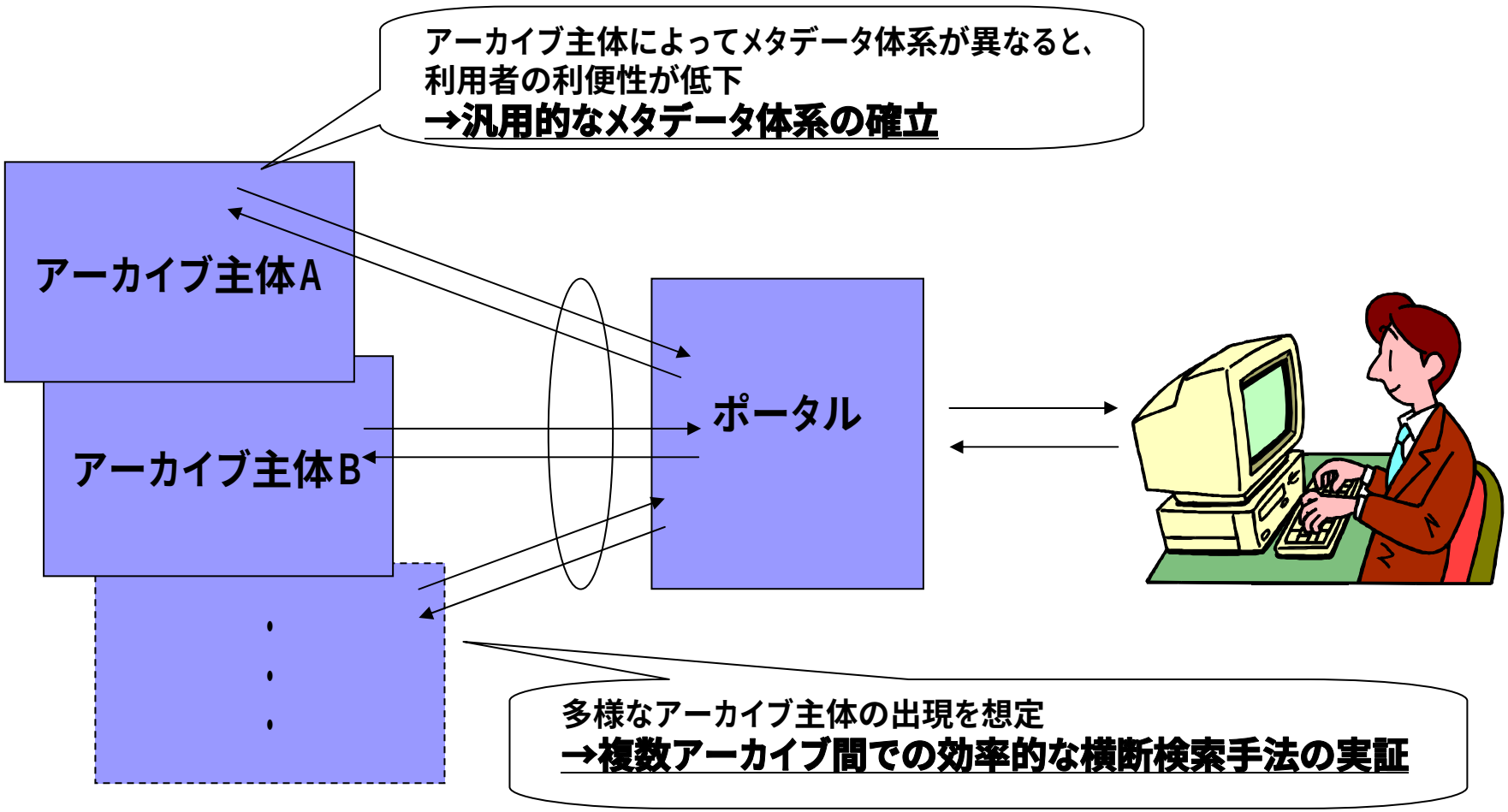
検証用端末
利用者

(2) メタデータに基づく時系列別検索、特定コンテンツに係る時系列別表示の実現

(4) 同一のウェブ情報を重複して保存することを回避する技術に関する調査研究
(5) 収集範囲拡大に関する調査研究

①異なるアーカイブ間の横断検索を実現する汎用的なメタデータ体系の確立

・ウェブ情報に関する汎用的メタデータ体系の確立
・複数アーカイブ間での効率的な横断検索手法の実証



①異なるアーカイブ間の横断検索を実現する汎用的なメタデータ体系の確立（補足）

- 本実証実験においては、ウェブ情報に関する汎用的なメタデータ体系を確立し、異なるアーカイブ間を横断的に検索できるようにする。具体的には、アーカイブA及びアーカイブBのウェブ情報にメタデータを付与し、両アーカイブ間の横断検索について実証を行う。なお、実証を行うにあたっては、以下に示す二通りの方式について横断検索の実証を行う。
  - メタデータをそれぞれのアーカイブ主体に分散したまま横断検索する方式
  - メタデータをポータルサイトサーバに集中させて横断検索する方式

- メタデータ体系の確立にあたっては、以下の状況を勘案することとする。
  - ①現在、国立国会図書館がインターネット情報資源選択的蓄積実験事業（WARP）において用いているメタデータ体系
  - ②J/Meta（＊1）等のデジタルコンテンツ一般を対象とする国内の汎用的なメタデータ体系
  - ③Dublin Core（＊2）、RDF（＊3）等、ウェブコンテンツに関する海外の汎用的なメタデータ体系
  - ④その他、国立国会図書館以外の多様なアーカイブ主体が用いることが想定されるメタデータ体系

＊1　（財）マルチメディア振興センターが策定した、様々なコンテンツのネットワーク流通・促進を目的としたメタデータ共通フレームワーク。

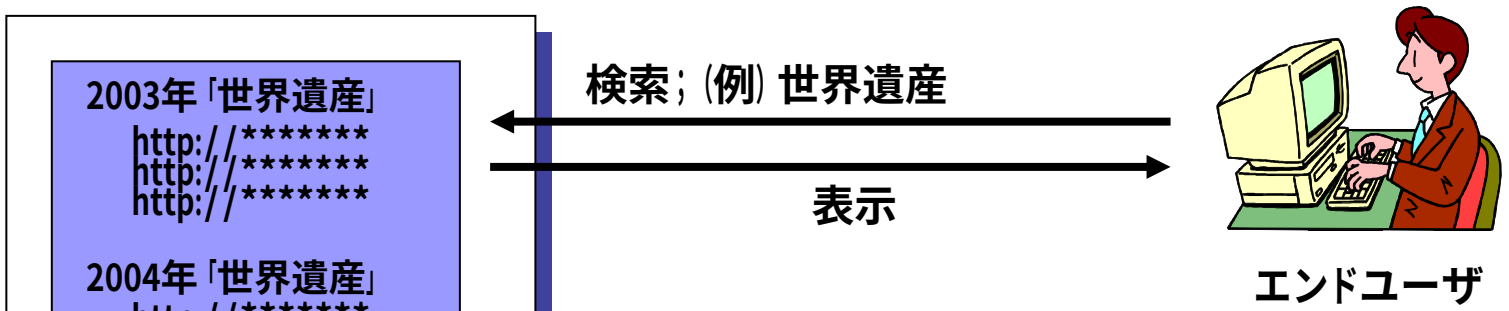
＊2　Dublin Core Metadata Initiative が提唱している、インターネット上の情報資源についての共通メタデータ項目定義であり、15項目の要素を基本とする。

＊3　W3Cで規格化されている、インターネット上の情報資源を記述するためのメタデータ共通フレームワーク。(RDF ; Resource Description Framework)

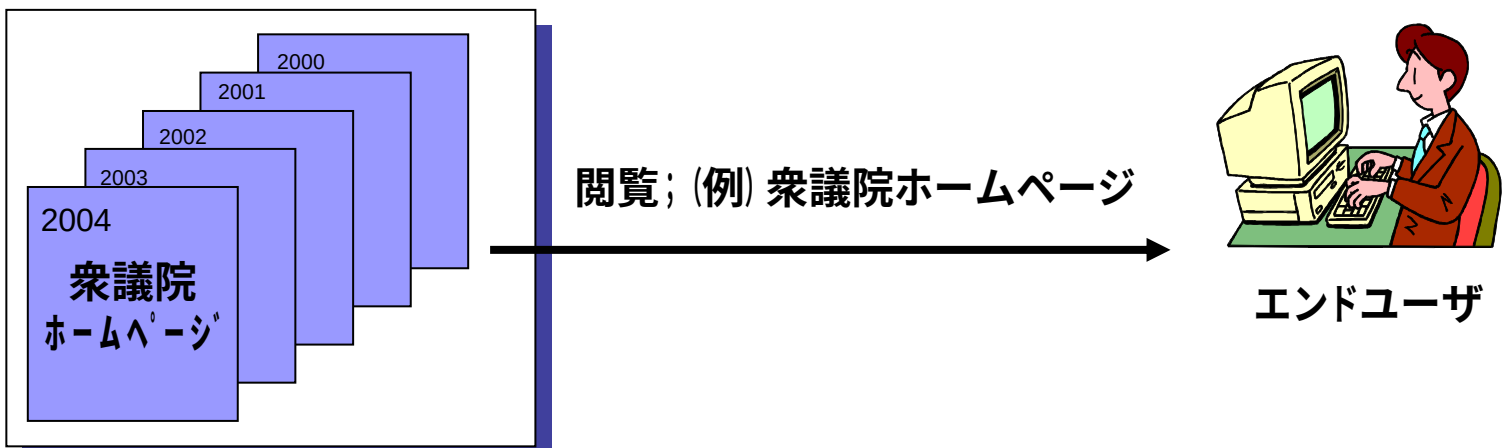
②メタデータに基づく時系列別検索、特定コンテンツに係る時系列別表示の実現

・時系列データを有効に検索・閲覧する機能（(1)、(2)）の実証

(1) 検索結果を年代別に検索・表示する機能

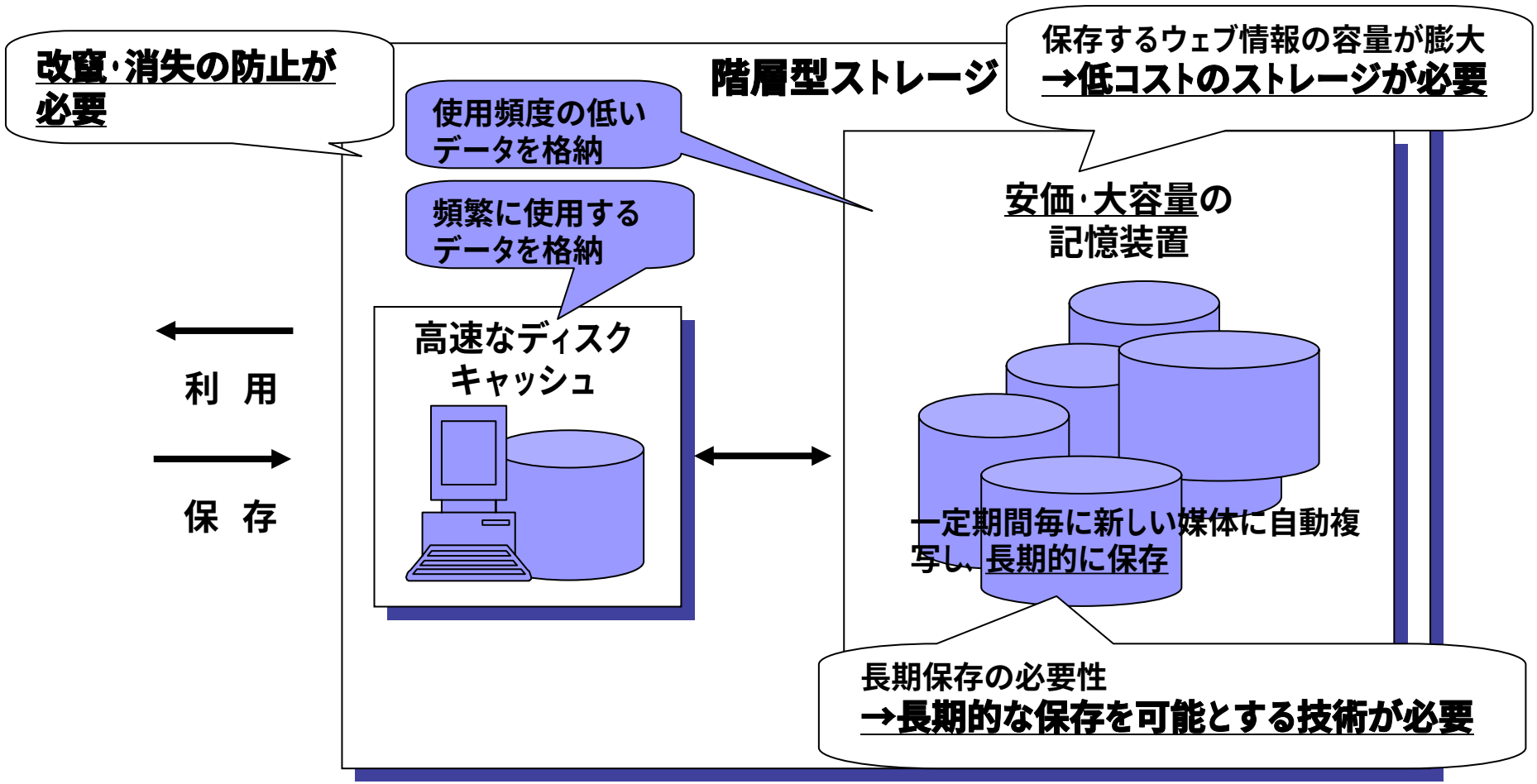


(2) 特定のコンテンツを時系列に並べて表示する機能

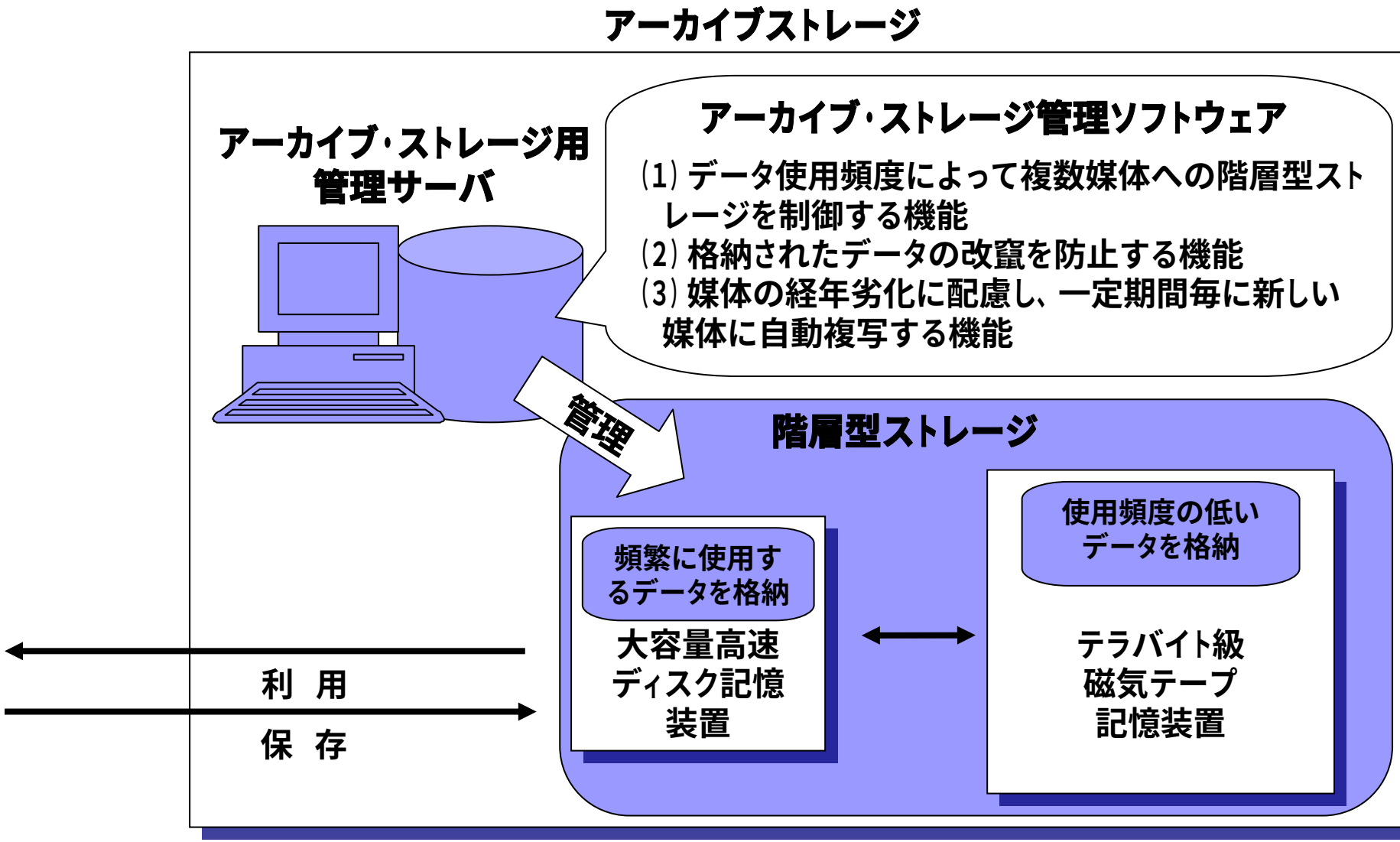


③コストの縮減、長期的な保存を可能とするアーカイブ技術の開発等

・低コストのストレージを可能とする技術の実証
・長期的な保存を可能とする技術の実証
} 階層型ストレージの実証
・保存したウェブ情報の改竄・消失防止技術の実証



③コストの縮減、長期的な保存を可能とするアーカイブ技術の開発等（補足）

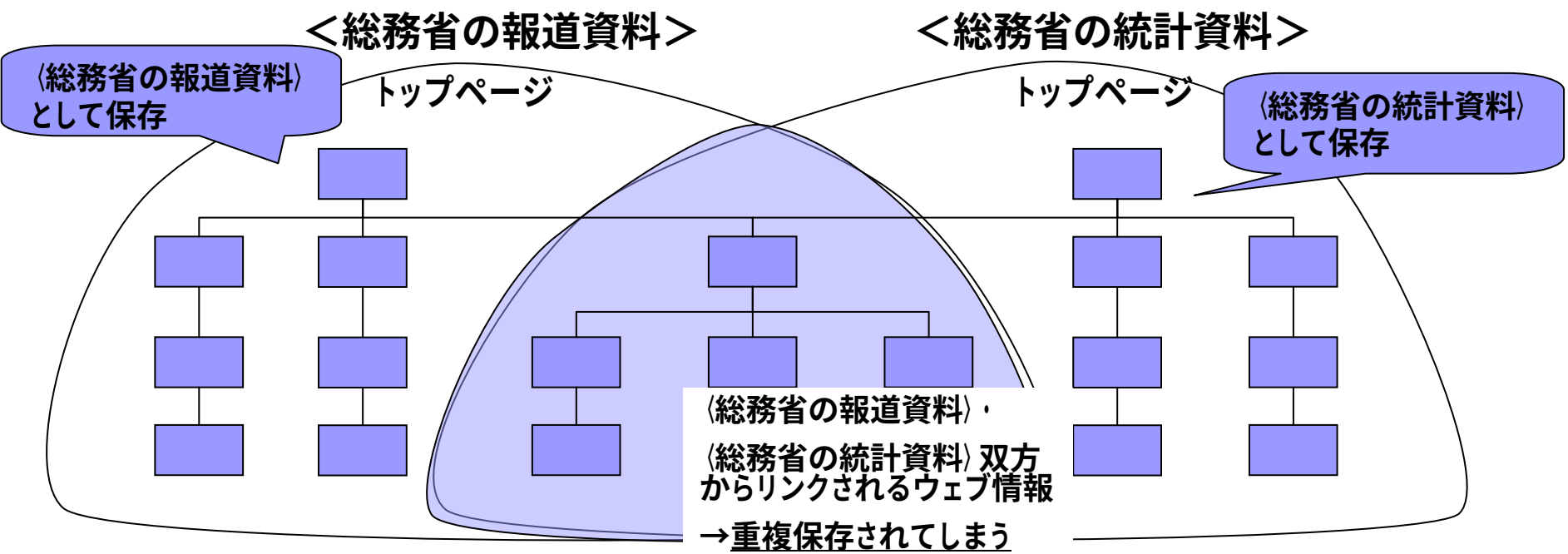


④同一のウェブ情報を重複して保存することを回避する技術に関する調査研究(1/2)

保存するコンテンツの量が膨大であるため、同一ウェブ情報の重複保存を回避することにより、容量を節約する必要性
→同一のウェブ情報を重複して保存することを回避する必要性がある。(1)、(2)-1、(2)-2)

(1)複数のウェブ情報からリンクされるウェブ情報において発生する、同一情報の重複保存を回避する技術の調査研究

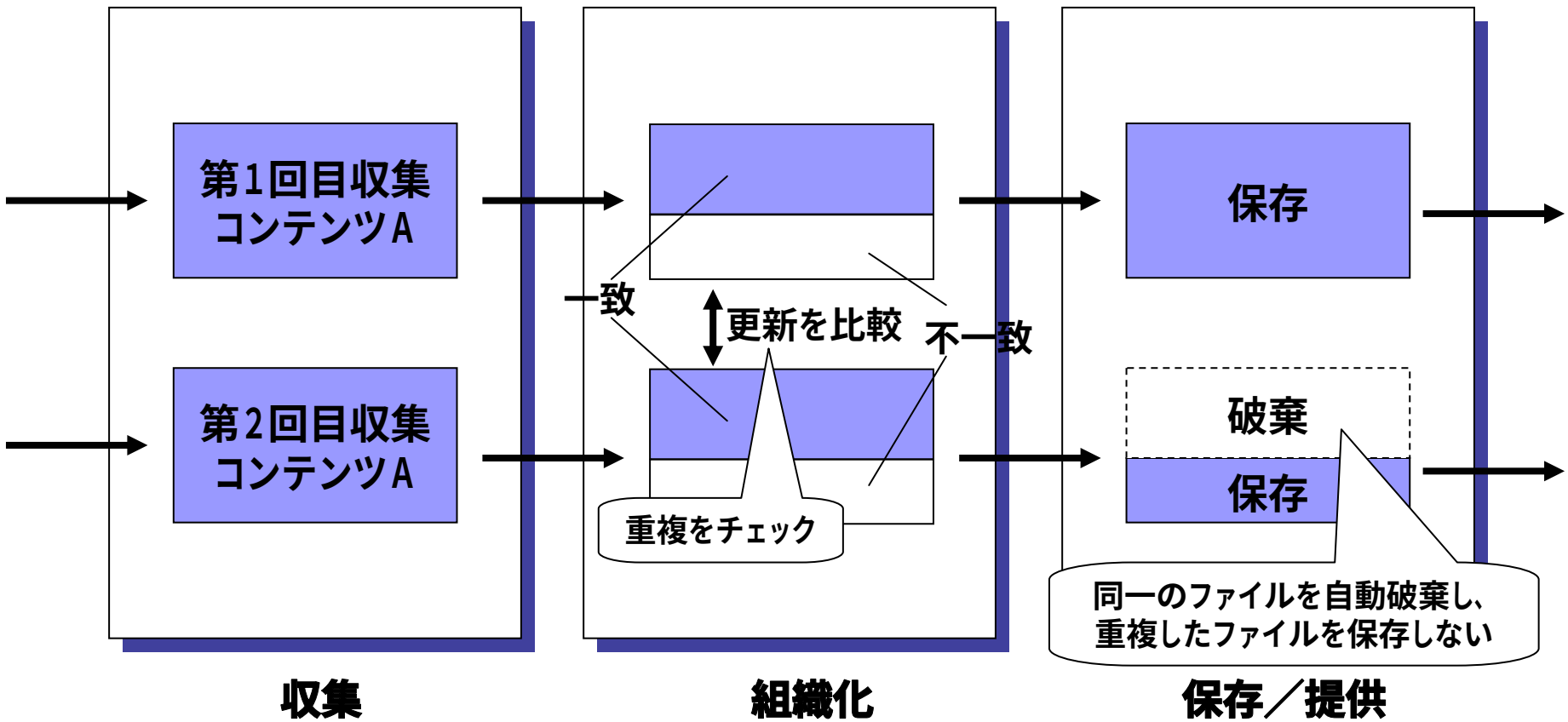
(例)



④同一のウェブ情報を重複して保存することを回避する技術に関する調査研究(2/2)

(2)-1 再収集したコンテンツについて、コンテンツの非更新部分（重複部分）の保存を回避する技術の調査研究

(2)-2 閲覧･収集の際に、更新部分と非更新部分を組み合わせ、コンテンツを再構成する技術の調査研究

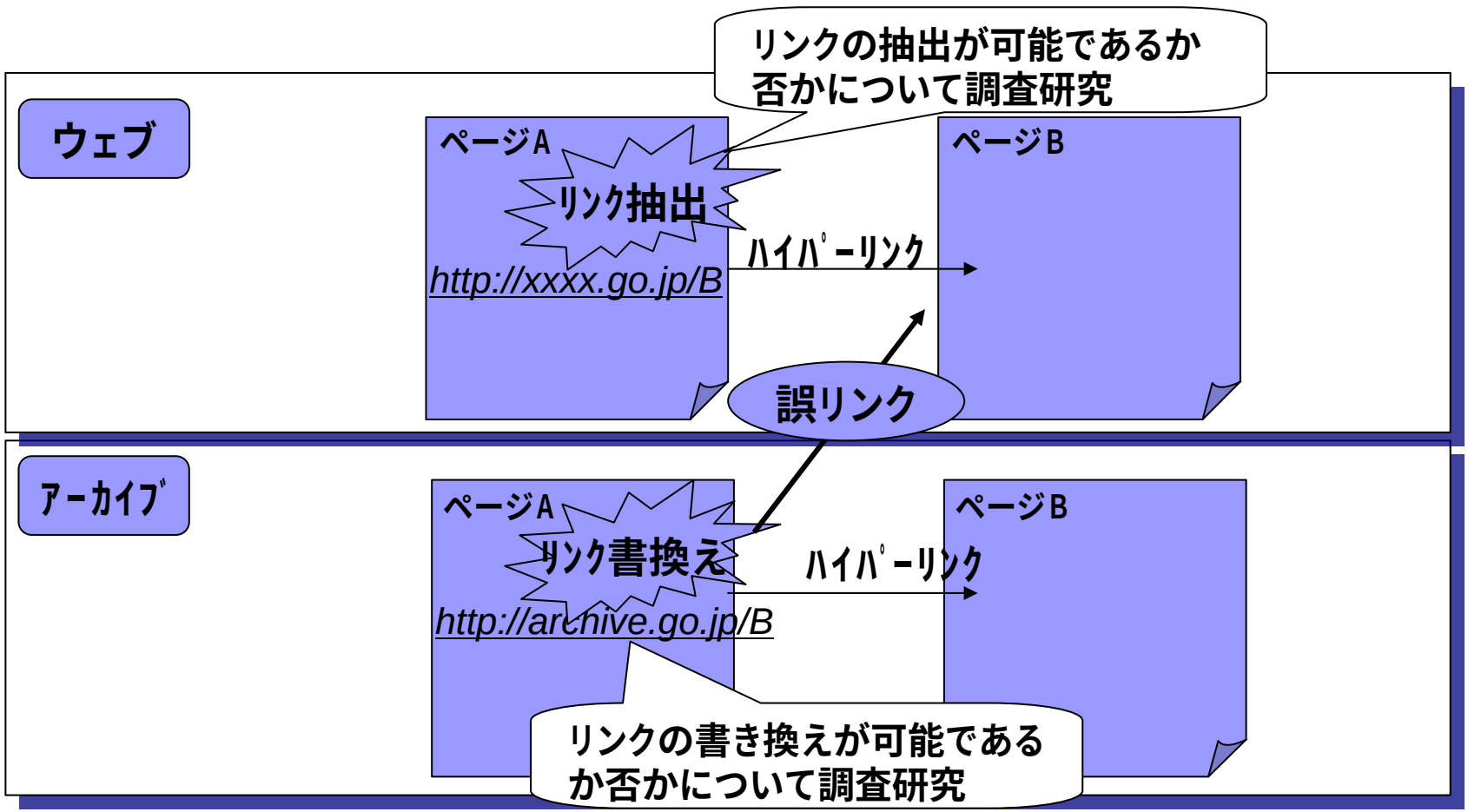


⑤収集範囲拡大に関する調査研究

・リンクの抽出
・リンクの書き換え

この2点が可能であるか否かについて調査研究
（特にHTMLでないバイナリ型オブジェクト（PDF、WORD、Flash等））

・ストリーミングデータ、Flash等のアーカイブ化技術について調査研究



資料６－２

ウェブ情報アーカイブの利活用方法に関する調査研究

- **目的**
  - ウェブ情報アーカイブの利活用モデルについての調査研究を実施する。

- **内容**
  - (1) 海外におけるウェブ情報アーカイブの利活用についての実態調査
  - (2) ウェブ情報アーカイブの利活用に関する情報収集
  - (3) ウェブ情報利活用モデルの確立に向けた今後の課題