

平成22年6月3日現在

研究種目：基盤研究(B)  
 研究期間：2007～2009  
 課題番号：19300081  
 研究課題名(和文) デジタル資料の長期保存とデジタルアーカイブの長期利用性に関する総合的研究  
 研究課題名(英文) Comprehensive Study on Long-term Preservation of Digital Resources and Long-term Use of Digital Archives  
 研究代表者  
 杉本重雄(SUGIMOTO SHIGEO)  
 筑波大学・大学院図書館情報メディア研究科・教授  
 研究者番号：40154489

## 研究成果の概要(和文)：

デジタル情報資源の長期保存とデジタルアーカイブの長期利用は、デジタル化が進む図書館や公文書館における重要な課題として広く認められている。本研究では、文書資料の効率的な保存方法、権利管理や利用者の視点からの適切な資料保存等の問題を含め、デジタル情報資源の保存とデジタルアーカイブの長期利用のための要求要件を総合的にまとめることを目的とし、電子資料の保存技術、情報管理技術、社会制度と利用性向上の視点からの研究を進めた。

## 研究成果の概要(英文)：

The amount of digital resources collected at libraries and archives is rapidly increasing. Preservation of digital resources and longevity of digital archives are widely recognized as crucial issues for libraries and archives. This project aimed to clarify technological aspects and requirements for preservation of digital resources and longevity of digital archives, i.e. technologies to efficiently preserve digital documents, resource and rights management in digital archives, and usability of digital archives. This project has developed technologies for long-term use of digital archives in three aspects – preservation technology, information management technology and usability issues in social and individual contexts.

## 交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2007年度	5,200,000	1,560,000	6,760,000
2008年度	4,800,000	1,440,000	6,240,000
2009年度	4,700,000	1,410,000	6,110,000
年度			
年度			
総計	14,700,000	4,410,000	19,110,000

研究分野： デジタルライブラリ

科研費の分科・細目： 情報・図書館情報学・人文社会情報学

キーワード： デジタルコンテンツの長期保存, デジタルアーカイブの長期利用, デジタルアーカイブ, メタデータ, メタデータスキーマ, Web アーカイブ, デジタルライブラリ

## 1. 研究開始当初の背景

インターネットの急速な発展の中で、電子的な文書の発信・流通が飛躍的に進み、デジタルアーカイブの構築が進められた。電子的な文書資料の長期保存の困難さと重要性が認識され、各国の国立図書館や公文書館を中心に組み込まれてきている。しかしながら、これまでの成果は特定の領域を指向したものや文書そのものの保存にとどまっている。一方、デジタルアーカイブには、個別の資料に関わる問題とアーカイブ全体に関わる問題があり、電子文書技術や資料情報管理の問題、社会制度や利用者面の問題がある。こうした背景の下で、本研究では、デジタル情報資源の長期保存とデジタルアーカイブの長期利用に関する総合的な研究を開始した。

## 2. 研究の目的

本研究では、文書資料のライフサイクル全体を見た上での効率的な保存方法、権利管理や利用者の視点からの適切な資料保存といった問題を含め、デジタル情報資源の保存とデジタルアーカイブの長期利用のための要求要件を総合的に研究することを目的とした。本研究では以下の3つの視点を中心に研究を進めた。

- ・ 電子資料の保存技術の視点
- ・ デジタルアーカイブの長期利用のための情報管理技術の視点
- ・ デジタルアーカイブの長期利用に関わる社会制度と利用性向上の視点

また、いろいろな視点からの議論を行うために、国際シンポジウムを開催し、デジタル資料の長期保存とデジタルアーカイブの長期利用に関する意見交換の場を設けた。こうした研究活動を通じて、デジタルアーカイブの長期利用に関するガイドライン開発を進めることを目的とした研究を進めた。

## 3. 研究の方法

本研究では、上の3つの視点を基礎にして、それぞれの領域での研究を進めるとともに、国際シンポジウムを開催していろいろな視点からの意見交換を行った。以下は、上に述べた視点に基づく研究内容の概要である。

- (1) 電子資料の保存技術の視点： Web アーカイビング技術を利用して、組織内の Web 文書を保存するもので、本研究では文書へのアクセス権等も含めた組織内

Web アーカイブ技術、Web 文書の一貫性管理の研究では、文書間の矛盾を効率的に発見、補修する Web 文書の一貫性管理技術に関する研究を進めた。

- (2) デジタルアーカイブの長期利用のための情報管理技術： メタデータスキーマの相互運用性管理技術、デジタルコンテンツの保存過程で利用する複数のメタデータスキーマの統合モデルに関する研究を進めた。これらはアーカイブの過程におけるメタデータおよびメタデータスキーマ管理のための情報管理技術要素として重要である。
- (3) 社会制度と利用性向上： 電子文書や電子書籍の効率的な管理と保存に関する研究を中心にして研究を進めた。ここでは、電子公文書のアーカイブに関する研究、デジタル出版物の流通、保存、再利用の視点から、デジタル形式のマンガのためのメタデータスキーマの基本モデルの研究を進めた。
- (4) 国際シンポジウムの開催： 以下の国際シンポジウムを行った。
  - ・ デジタルアーカイブの長期利用に関する国際シンポジウム（2008年3月、筑波大学）
  - ・ デジタル情報資源の長期保存とデジタルアーカイブの長期利用に関する国際シンポジウム（2010年2月、国立国会図書館）

## 4. 研究成果

本研究では、先に述べたように3つの視点の領域において研究を進めた。各領域において、システムやモデルの研究開発を進めた。また、シンポジウムを行い、関連する研究成果に関する情報共有と意見交換の機会を設けた。

以下の節では、本研究における主な研究成果について述べる

### 4.1 組織内 Web アーカイビング

現在、国立図書館を中心としてインターネット上の情報資源を収集、蓄積、保存する Web アーカイビングの開発が盛んに進められている。その一方、一般の Web アーカイビングの場合には、Deep Web あるいは Hidden Web と呼ばれる、情報資源収集ロボットには隠さ

れた資源に対してどのように対処すべきかといった問題があることも広く知られている。本研究では、こうした問題を、組織の中に開かれた Web のアーカイブの問題ととらえ、組織の中で、情報資源提供者とアーカイブ管理者が連携して収集、蓄積、保存、提供を行う Web アーカイビングシステムの研究を進めた。

本研究の Web アーカイビングシステムの特徴は、情報資源提供者が、資源の収集と保存そして提供に係る条件をメタデータとして記述し、そのメタデータとアーカイブ管理者が記述する保存のポリシーに関するメタデータを加えて、保存された情報資源の管理方法を制御する方法を提案した。

また、長期にわたるアーカイブの運営の場合、運営組織そのものが消滅したり、他組織と統合したりする可能性がある。そのため、本研究では、Web アーカイブの分割統合のための技術に関する研究を行った。分割統合時には、単に Web 情報資源のデータベースとして分割統合するのではなく、情報資源ごとの公開制限等のアクセス管理を継続して行わねばならない。そのため、本研究では、保存管理方針のメタデータを利用して、組織内アーカイブの分割統合管理を行う方法を提案した（参考：5の論文②）。

この研究を発展させ、公文書の公開時等に行われる文書の一部の墨塗りによるマスキングを管理するための方法の研究を進めた（参考：5の学会発表④）。図1に示すように、本システムでは、管理ポリシーと個別のリソースに与えられるメタデータを利用してマスキングに関する制御をおこなう。リソースの条件記述と利用時点での環境で決まる条件に基づいてマスキングされる部分が決定される。

#### 4.2 Web 文書の一貫性管理

HTML 文書はもちろんのこと Web 上で提供される多くの文書はハイパーテキスト構造を持っている。ハイパーテキスト構造は、文書間を柔軟につなぐ半面、リンク切れの問題を持っている。特に、アーカイブの観点からはリンク切れの問題は重要である。そのため、

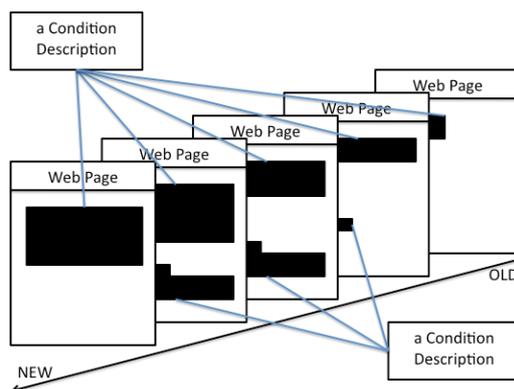


図1 ポリシー記述に基づくマスキング管理

本研究では、リンク切れを修復するための技術、関連を持つ文書間において内容に矛盾が生じないようにするための技術の研究を進めた。

コンテンツ間のリンクを自動的に修復するために、リンク切れを起こす原因に関する分析を基礎にして、リンク切れを起こした文書を見つけるための基本的な探索手続きとヒューリスティクスによる精度向上により、95%の信頼度で約70%以上のページの移動先を発見することができた（参考：5の学会発表③）。

コンテンツの一貫性を発見するためには、Web コンテンツから構造データを抽出する必要がある。本研究では、XML から DTD を自動抽出するために提案されている基準である MDL コストと、ラッピングの具体性を表す値とを組み合わせることにより、HTML から XML データを抽出するためのデータ変換ソフトウェアであるラッパが、自動生成可能であることを示した（参考：参考[1]）。また、分散管理された Web コンテンツの内容一貫性管理にも取り組み、データベース分野でこれまで扱われてきた包含従属性に着目し、既存 Web コンテンツに内在する包含従属性を効率よく発見するための支援についても研究を行ってきた。発見支援処理をビットシグネチャを用いて高速化する手法を提案し、これにより大幅な高速化が実現できる可能性を示した（参考：5の論文①）。

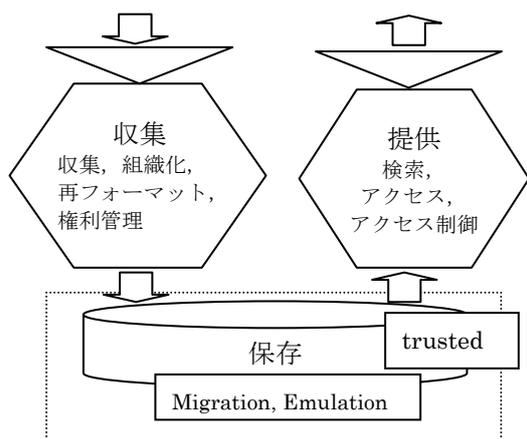


図2 デジタルアーカイブの概念構造

### 4.3 デジタルアーカイブのためのメタデータのモデル

デジタルアーカイブはいろいろな組織によって作られてきているため、その構成方法もいろいろである。本研究では、図2に示すような一般的なアーカイブのモデルを基礎として、電子公文書管理やデジタルライブラリにおけるアーカイブのいろいろな課題について検討してきた(参考: 5の図書①および参考[2])。このモデルでは、収集、保存、提供を明示的に分け、各フェーズでのタスクを明確化することで、そこで必要とされる機能、メタデータ記述内容を明確化できる。

公文書館、図書館は、それぞれのミッションの違いから資料の整理の仕方、目録の作り方が異なっている。また、図2に示すように、アーカイブのフェーズによって行われるタスクが異なるため、そこで必要とされるメタデータが異なる。そこで、本研究では、アーカイブされるリソースのライフサイクルとタイプに応じたアーカイブのためのガイドラインに関する研究を進めた(参考: 5の学会発表②)。こうした研究から、アーカイビングのタスクフローの中で用いるメタデータスキーマに関する記述と、メタデータスキーマ間のシステムティックなマッピングスキーマの必要性を強く感じた。そのため、タスクに応じたメタデータスキーマ記述のためのジェネリックなメタデータ記述方式に関する研究を開始し、現時点で研究を継続している。

公的機関や企業での電子文書のアーカイ

ブをできるだけライトウェイトな環境で実現する必要がある。こうした観点から、本研究では、既存の Contents Management System (CMS) と Record Management System (RMS) を橋渡しする軽量のシステムに関する研究を行った。ここでは、CMS と RMS のネットワーク経由の簡便な接続手法を、階層モデルとしてモデル化し、RSS を利用して実験システムを実現した(参考: 5の学会発表①)。

### 4.4 その他の成果

研究代表者等は、以前より Dublin Core Metadata Schema Registry の運営を行ってきた(参考: 参考[3])。デジタルアーカイブの長期利用にとって、メタデータの長期利用性は重要な要件であり、メタデータの長期利用性を高めるためにレジストリは重要な役割を持っている。レジストリは、メタデータの記述要素の定義だけではなく、統制語彙とその要素の語、リソースのクラスや値の型を表す語の定義も格納する。レジストリに登録された語のメンテナンスを効率的に行うための技術は重要な課題である。

本研究では、デジタルコンテンツの流通、共有、再利用の視点からマンガのメタデータに関する研究を進めた(参考: 5の学会発表⑤)。現在、マンガの制作環境のデジタル化が進み、マンガそのもののデジタル化も進んでいる。マンガは、イメージ、テキスト、シンボルなど、いろいろな要素の組み合わせとして実現される。そして、デジタル形式のマンガの場合は、動きや音を持つ表現もある。また、一つの作品を多種の閲覧環境で適合させることもある。こうした多様性を持つマンガのための制作作業、流通の効率化、保存されているデジタルマンガの検索性の向上は重要な課題である。

本研究では、こうした性質を持つデジタルマンガにおいてメタデータが持つ役割は大きいと考え、マンガのためのメタデータのフレームワークの研究を進めた。本研究では、IFLA の Functional Requirements for Bibliographic Records (FRBR), TV Anytime を基礎としてメタデータのモデルを設計し、さらにメタデータを支えるオントロジーに関する検討を行った。

#### 4.5 国際シンポジウム

本研究では、デジタルアーカイブに関する議論を進めるため、国際シンポジウムを2008年3月と2010年2月に開催した。

- (1) デジタルアーカイブの長期利用に関する国際シンポジウム(2008年3月、筑波大学)

ドイツにおけるデジタルアーカイブ開発とコンテンツ保存の中心の一つであるゲッチンゲン大学図書館よりデジタルアーカイブに関する専門家である Heike Neuroth 氏, Stephen Strathmann 氏を招き、同館におけるデジタルアーカイブプロジェクトを基礎とした組織的課題と技術的課題に関する講演を行った。

[http://www.kc.tsukuba.ac.jp/index\\_das08.html](http://www.kc.tsukuba.ac.jp/index_das08.html)

- (2) デジタル情報資源の長期保存とデジタルアーカイブの長期利用に関する国際シンポジウム(2010年2月、国立国会図書館)

本シンポジウムは国立国会図書館との共同で開催した。オーストリア・ウィーン工科大学 Andreas Rauber 准教授, シンガポール・ナンヤン工科大学 Paul Wu 博士, アメリカ・ミシガン大学 Margaret Hedstrom 准教授による招待講演を行った。Rauber 氏はデジタル情報資源保存のためのプランニングに関して, Wu 氏はアノテーションを利用した Web アーカイブの利用支援に関して, そして Hedstrom 氏はデジタル保存の現状と必要性に関して述べた。また, 国立国会図書館からも, 同館におけるデジタルアーカイブの取り組みについての説明があった。

<http://www.kc.tsukuba.ac.jp/symposium/symp20100219.html>

#### 4.6 まとめ - 将来への課題

デジタルアーカイブの長期利用性を高めることは、非常に重要な課題である。その一方、完璧なデジタルコンテンツの長期保存は非常に困難な課題である。デジタルライブラリの研究開発が活発化してすぐに、この問題への取り組みが進められてきた。たとえば、OAIS のような長期保存システムのための国際標準、PREMIS に代表されるデジタル情報資源に特化した保存のためのメタデータ規格も作り上げられてきた。また、デジタルライブラリの国際会議でもデジタル情報資源の保存は重要な技術トピックと

して認められており、保存やアーカイビングに特化した国際会議も継続的に開催されてきている。

本研究では、上に述べたようにデジタル情報資源の長期保存とデジタルアーカイブの長期利用のための技術面を中心として成果をあげることができたと考えている。また、アーカイブのためのメタデータスキーマを横断的に表現するための方法といった、本研究を進めることから新たに課題として現れ、現在も継続中のものもある。こうした新たな課題は、本研究において重要な目標とした総合的なガイドラインの開発にとってコアなものと考えている。

本研究の遂行期間の間に、公文書管理法が制定され、紙の文書、電子文書の区別なく適切に管理し、保存することが定められた。本研究の成果はこうした領域にも役立つと考えている。また、実際にアーカイブを運営していくには、それを支えるための人材が必要であり、そうした人材育成・カリキュラム開発といったことも、将来への課題としてとらえている。

参考 (5 に示すもの以外の発表論文その他)

- [1] 澤菜津美, 森嶋厚行, 飯田敏成, 杉本重雄, 北川博之. Web ページ移動先発見のための効率的なクローリング手法, 情報処理学会論文誌: データベース, Vol.48, pp.149-161, 2007
- [2] 杉本重雄. 電子記録の円滑な保存・利用に向けて, 国際公文書館会議東アジア地域支部 (EASTICA) 2007 年総会・セミナー招待講演, 2007
- [3] The Dublin Core Metadata Registry, <http://dcmi.kc.tsukuba.ac.jp/dcregistry/>

#### 5. 主な発表論文等

(研究代表者, 研究分担者及び連携研究者には下線)

[雑誌論文] (計 8 件)

- ① 高橋公海, 森嶋厚行, 松本亜季子, 杉本重雄, 北川博之. Web コンテンツ管理のための一貫性制約発見支援, 日本データベース学会 Letters, Vol. 7, No. 3, pp. 25-30, 2008, 査読有
- ② 柘和佑, 阪口哲男, 杉本重雄. 分割・統合可能な組織内 Web アーカイブシステムの構成方法. 情報知識学会, Vol.18, pp.47-57, 2008, 査読有

[学会発表] (計 27 件)

- ① A. Morozumi, S. Nomura, M. Nagamori, S.

Sugimoto. Metadata Framework for Manga: A Multi-paradigm Metadata Description Framework for Digital Comics, Proceedings of International Conference on Dublin Core and Metadata Applications 2009, pp.61-70, 2009.10.15, Seoul, Korea, 査読有

- ② Wasuke Hiiragi, Tetsuo Sakaguchi, Shigeo Sugimoto. A Policy-based Institutional Web Archiving System with Adjustable Exposure of Archived Resources, Proceedings of the 9th International Web Archiving Workshop, 2009.10.1, 7p., Corfu, Greece, 査読有
- ③ A. Morishima, A. Nakamizo, T. Iida, S. Sugimoto, H. Kitagawa. Bringing Your Dead Links Back to Life: A Comprehensive Approach and Lessons Learned. The 20th ACM Conference on Hypertext and Multimedia (ACM Hypertext 2009), pp. 15-24, 2009.6.30, Torino, Italy, 査読有
- ④ 白才恩, 杉本重雄. デジタルリソースの保存方式選択のためのガイドライン, 情報処理学会情報学基礎研究会, 研究報告, 8p., 2007.11.8, つくば市, 査読無
- ⑤ Jan Askhoej, Mitsuharu Nagamori, Shigeo Sugimoto. Constructing a records archiving system using off-the-shelf tools - A light weight approach, Proceedings of the 7<sup>th</sup> International Web Archiving Workshop, 2007.6.23, 8p., Vancouver, British Columbia, Canada, 査読有

[図書] (計 1 件)

- ① 杉本重雄. デジタル世界における図書館とアーカイブ, 別冊「環」15 図書館とアーカイブズとは何か, 藤原書店, 8p, 2008

[その他]

ホームページ:

<http://www.slis.tsukuba.ac.jp/digitalarchive/>

## 6. 研究組織

### (1)研究代表者

杉本重雄 (SUGIMOTO SHIGEO)

筑波大学・大学院図書館情報メディア研究科・教授

研究者番号: 40154489

### (2)研究分担者

阪口哲男 (SAKAGUCHI TETSUO)

筑波大学・大学院図書館情報メディア研究科・准教授

研究者番号: 10225790

森嶋厚行 (MORISHIMA ATSUYUKI)

筑波大学・大学院図書館情報メディア研究科・准教授

研究者番号: 70338309

永森光晴 (NAGAMORI MITSU HARU)

筑波大学・大学院図書館情報メディア研究科・講師

研究者番号: 60272209

新保史生 (SHINPO FUMIO)

筑波大学・大学院図書館情報メディア研究科・准教授

研究者番号: 20361355

(2009年度: 連携研究者, 2009年4月より慶應義塾大学・総合政策学部・准教授)

永田治樹 (NAGATA HARUKI)

筑波大学・大学院図書館情報メディア研究科・教授

研究者番号: 40124200

(2008-2009年度: 連携研究者, 2009年度 筑波大学名誉教授)

川原崎雅敏 (KAWARASAKI MASATOSHI)

筑波大学・大学院図書館情報メディア研究科・教授

研究者番号: 70375517

(2008-2009年度: 連携研究者)

西岡貞一 (NISHIOKA TEIICHI)

筑波大学・大学院図書館情報メディア研究科・教授

研究者番号: 60436285

(2008-2009年度: 連携研究者)

### (3)連携研究者

鈴木誠一郎 (SUZUKI SEIICHIRO)

筑波大学・大学院図書館情報メディア研究科・教授

研究者番号: 70455939

(2009年度 連携研究者)

柊和佑 (HIIRAGI WASUKE)

筑波大学・大学院図書館情報メディア研究科・準研究員 (現在, 稚内北星学園大学・情報メディア学部・講師)

研究者番号: 80530659

(2009年度 連携研究者)