

デジタルアーカイブのWeb公開に向けて

—デジタルアーカイブに関連する各種情報技術の概観—

専修大学 経営学部
植竹 朋文

1. 背景

近年、学術研究においては、各研究分野の高度化、国際化とともに、既存の分野や専門性にとらわれない学際化という新たな動向が顕著である。これに伴い学術研究を進める中で生み出される情報や、研究者が必要とする情報も飛躍的に増大している。一方、コンピュータや各種情報技術の急速な進展とインターネットの普及、及びブロードバンド化により、情報の処理や遠隔地に存在する情報へのアクセスが容易にできる環境が整ってきている。学術審議会学術情報資料分科会学術情報部会の「学術情報データベースの整備について¹⁾」で報告されているように、このような背景の中で行われる研究活動の基盤として、研究者が研究を進める上で必要とする情報を網羅的かつ効率的に入手できる環境を、いかに早く整えられるかが研究活動の発展にとって、大きな鍵となっていると言える。このため、必要となる情報を収集・蓄積し、整理・加工して提供する体制の整備充実が研究活動の発展にとって必須の条件となる。そして、さまざまな情報を電子化し、体系的に整理し、コンピュータ処理が可能な形にデータベース化して容易に利用できるようにすることが、益々必要となっている。

このような状況のもと、人文科学系の各分野で、各種学術・文化資料等に対して情報処理技術を用いたさまざまなアプローチが試みられ、提示され、議論され、対象となる資料をデジタル化した上で保存・流通・共有するための研究が進められてきている。そして、その一環として、デジタル化されたアーカイブあるいはデジタルデータのアーカイブ、すなわち「デジタルアーカイブ」が注目され、すでに多種多様なデジタルアーカイブが構築され、公開されている。

各種学術・文化資料などをデジタルアーカイブ化し、データベースとして提供することにより、以下のような効果が期待される。

1. 精緻・網羅的かつ迅速・効率的な情報の収集・提供が可能となる
2. 情報の検索・加工、整理・分析といった処理が容易となる
3. 情報を国内外へ発信することが容易となる

また、デジタルアーカイブの意義は、研究面に止まるものではなく、文化財ないし学術財として、研究活動の成果を電子的に記録し、後世への保存・伝承を可能とするとともに、教育面でも、デジタル化された情報は、その活用を一層容易にする。また情報の利活用の面ではデータを利用して各種の作品をつくり、研究、知育、教育、趣味、鑑賞など私たちの生活やビジネスでのさまざまな場面で広く活用できるというメリットもある。

デジタルアーカイブ白書2005によると、デジタルアーカイブを構築、公開している主な組織は図書館、博物館、美術館、各種研究・教育機関である。ここでは、その代表的な組織である図書館についてみると、JDAA²⁾の調査でも示されるように、公共図書館では、デジタルアーカイブの公開率が2004年の9.5%から2005年は42.6%と大きな飛躍を遂げている。このほか公文書館でも33.3%から53.1%と著しい増加が見られる。

¹⁾ http://www.mext.go.jp/b_menu/shingi/12/gakujutu/toushin/970601.htm

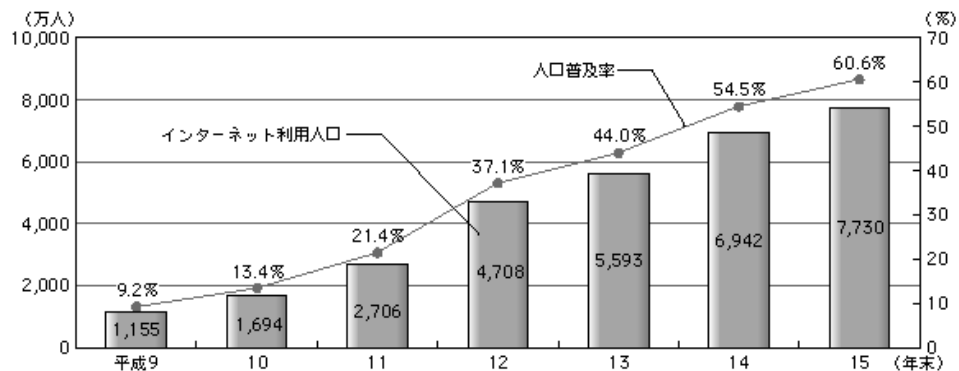
²⁾ Japan Digital Archives Association、デジタルアーカイブ推進協議会

また、ほとんどの組織がWebサイトを開設している。しかし、多くのデジタルアーカイブは、いまだシステム指向の傾向が強く、その目的や理念が必ずしも明確でないことが多い。また、利用者の立場から見ると、必要な情報が公開されていなかったり、インターフェースが貧弱であったりと、利用者を十分に意識したつくりになっていないものが多い。ただその一方で、一部では先進的事例ともいえるデジタルアーカイブの積極的な利活用を可能としているサイトが存在しているのも事実である。

そこで、本論文では、上述したような状況を踏まえ、平成17年度に選定されたオープン・リサーチ・センター整備事業である『Anglo-Saxon 語の継承と変容³』において、中世英文学における手稿写本 (Manuscript) の文字研究の成果をデジタルアーカイブとして公開していく上で必要となるインターネットにおける情報公開の特徴と、既存のデジタルアーカイブ及び関連する情報技術の調査・分析を行い、全体を概観する。

2. インターネットを用いた情報公開の特徴

ブロードバンドによるインターネット接続が急速に広まっている現在、日常生活における情報探索にインターネットが広く活用されている。平成16年版情報通信白書によれば、インターネット利用者は、平成15年末現在で7730万人に達し、人口普及率は60.6%と初めて6割を超えてきており、インターネットは着実に普及していると言える(図1参照)。また、多くのインターネット利用者は、多様な端末からインターネットを利用しており、利用場所も限定された場所から複数箇所での利用へと広がっているとともに、インターネットの平均利用時間も長時間化し、いつでも、どこでもインターネットを利用する傾向にある。



- ※1 上記のインターネット利用人口は、パソコン、携帯電話・PHS・携帯情報端末、ゲーム機・TV機器等のうち、1つ以上の機器から利用している6歳以上の者が対象
- ※2 平成15年末の我が国の人口普及率(60.6%)は、本調査で推計したインターネット利用人口7,730万人を、平成15年末の全人口推計値1億2,752万人(国立社会保障・人口問題研究所「我が国の将来人口推計(中位推計)」)で除したもの(全人口に対するインターネット利用人口の比率)
- ※3 平成9～12年末までの数値は「情報通信白書(平成12年までは通信白書)」より抜粋。平成13年末、14年末の数値は、通信利用動向調査の推計値
- ※4 推計においては、高齢者及び小中学生の利用増を踏まえ、対象年齢を年々拡げており、平成12年末以前の推計結果については厳密に比較できない(平成11年末までは15～69歳、平成12年末は15～79歳、平成13年末から6歳以上)

図1 インターネット利用人口及び人口普及率の推移(平成16年版情報通信白書より抜粋)

インターネットでの情報探索は、従来のマスメディア接触型の情報探索と違い、以下に示す特徴を持っている(植竹,2004)。

³ 本研究プロジェクトは、Anglo-Saxon 語が通時的にどの要素が継承され、どの要素が変容したかを検討するために、情報科学の成果をイギリスの大学図書館、大英図書館等に収蔵されている多くの写本群に光を当て、従来、日本国内で軽視されてきた中世英文学の基礎資料を包括的に集積するとともに、21世紀の科学性にふさわしくデジタルベース化して研究態勢を整える予定である。また、デジタル写本を作成して、写本の語彙、文体、統語構造等を分析し、中世英文学の分野において、国際的な貢献を目指すものである。

- ① 時間と場所の制約を受けない
- ② テレビなどの受動的な情報ではなく、情報探索やコミュニティやメーリングリストの参加など能動的な行動である
- ③ 比較がしやすい
- ④ 双方向的
- ⑤ 今までは手にいれることが難しかった口コミ情報も容易に入手可能

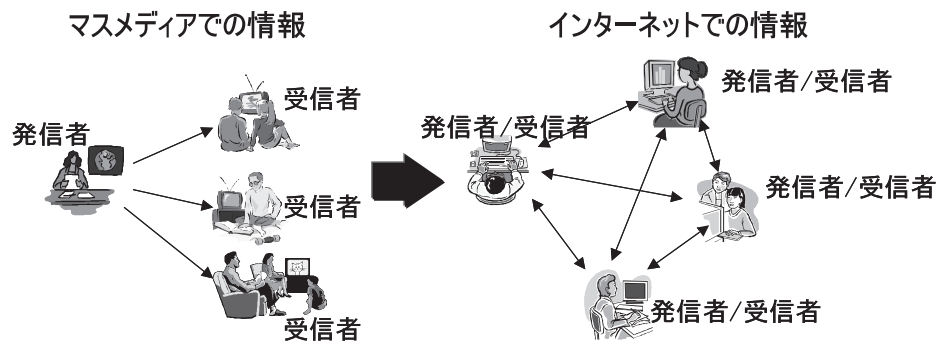


図2 情報の非対称性の解消 (植竹、2004 より引用)

つまり、図2に示されるように、従来型のマスメディアを利用した情報探索で存在した送り手から受け手へと情報が一方向に流れる「情報の非対称性」が、インターネットを利用することにより解消されたと言うことができる。

さらに、古川らも指摘しているように、インターネットを活用すれば、リアルの世界では見つけにくい同好の士を見つけることが容易である。インターネットは薄くて浅い関係から、狭くて濃い関係まで、自分たちにあった距離感の関係を構築する上で非常に便利なツールとなる (古川、2001)。

このような状況のもと、情報の利用者と提供者の関係性は大きく変化してきた。図3に示されるように、情報の提供者と利用者が持つ情報格差がほとんどなくなり、さらに、利用者が能動的に情報発信をしていく消費者生成メディア (CGM⁴) を作りだしてきた。そのCGMのひとつに、ネット上の口コミサイトがあるが、現実の場とどのような関連を持つのかまだ明らかにされていない。



図3 インターネットの情報と現実社会の関係

⁴ Consumer Generated Media

インターネットが普及し始めてほぼ10年がたち、インターネットはもはや新しいメディアではなく、私たちの生活に慣れ親しんだメディアとなってきた。グリー株式会社代表取締役社長の田中氏も述べている⁵ように、インターネット革命の序盤で言われていた「個人へのパワーシフト」は、ついにこれから、CGMの力を借りて、現実のものとなっていくと考えられる。

3. デジタルアーカイブについて

ここでは、デジタルアーカイブの定義、及び一般的な構築方法について述べる。

3.1. デジタルアーカイブとは

JDAAによるデジタルアーカイブの定義は以下のとおりである。

デジタルアーカイブとは散在している有形・無形の文化財、膨大な遺跡、自然環境などをデジタル映像や文書として記録・保管するシステムのこと。遺跡の劣化、書物など文化財の老朽が懸念される中で、これらの永久的保存に役立つ手段として注目されてきている。遺跡の状況や自然の風景を、くまなくデジタル映像として撮影・収録し、その映像をテーマや時系列、関連資料、研究成果とともに分類、データベース化し、ネットワークを通じて見るができるようにする。

3.2. デジタルアーカイブの構築方法

JDAAによると、デジタルアーカイブを構築するためには、以下に示す5つのステップを踏む。

- 計画層
- 記録層
- データベース層
- プレゼンテーション層
- 利用層

その詳細をデジタルアーカイブ白書2005より抜粋し、以下に各ステップごとに示す。

3.2.1. 計画層

計画層で考えるべきポイントは、どのような目的、使い方をするデジタルアーカイブをつくるか検討し、まとめることである。たとえば、単に所蔵品の記録・保存用としてつくる場合と、作品をつくりハイビジョンなどの高精細画面で来館者に鑑賞してもらう場合とでは記録層以後の作業の規模、必要なハードウェアのシステム、運営管理体制、予算などが大きく異なる。したがって、事前にどのような目的及び使い方をするのかをしっかりと検討しておかないと失敗する可能性が高い。また将来の収蔵作品量の拡大やハードウェア及びソフトウェアの技術的進歩などを考慮し設計することも重要である。



図4 計画層 (デジタルアーカイブ白書2005より引用)

⁵ <http://it.nikkei.co.jp/internet/column/literacy.aspx?i=20050714bi000bi&cp=2>

3.2.2. 記録層

記録層におけるポイントは、従来使用していた作品管理台帳の項目の再整理と新規項目の追加を十分に検討することである。また作品解説など作品情報としての文章については、データベースの作成を意識し、記述文字数として、短・中・長文の3段階くらいのレベルに整理しておくとならば後々便利である。

また、デジタルカメラやスキャニング等の技術的知識を持った人材の確保やスタジオ、利用目的に合わせた各種機材などの利用も必要となる。



図5 記録層 (デジタルアーカイブ白書2005より引用)

近年、機材については、高精細なデジタル画像を取得する各種機器の開発が一段と進展している。一般向けの分野でも、2003年ごろから10万円台のデジタル一眼レフカメラ (たとえば、Nikon D70等) が登場し、市場が急速に拡大しつつある。高画質デジタルカメラは高額なこともあり、国内ではまだ導入例が少ないが、今後はデジタル一眼レフカメラ製品にも影響を及ぼし、デジタルアーカイブ構築の中心機器になることが予想される。

3.2.3. データベース層

データベース層においては、完成した美術館資産としてのデジタルアーカイブを運営・管理する体制と要員がポイントである。図6に示されるように企業や個人がデータを利用して作品や商品を開発することが可能となるため、データの貸出方法、使用目的や使用条件のチェック、使用料金設定、課金・入金方法、権利関係者への支払、クレジット表記、不正複製のチェック、契約締結、違反時のペナルティ、コンピュータなどのハードウェアのメンテナンス等々を管理・運営する体制が必要となる。



図6 データベース層 (デジタルアーカイブ白書2005より引用)

3.2.4. プレゼンテーション層

プレゼンテーション層のポイントは、一般利用者にわかりやすい内容と操作性の良い作品を制作することである。特に最近の映像文化で育った利用者に満足いく作品を提供するには、内容もさることながら音や動きを組み入れたゲーム性やエンターテインメント性を持たせる工夫が必要不可欠となる。

3.2.5. 利用層

利用層においては、先のデータベース層と同様の管理・運営体制が必要となる。

4. インターネットにおける情報共有手法の動向

次に、近年のインターネットにおける情報共有の技術動向と、情報の標準化の動向について概観する。

4.1. インターネットにおける情報共有の技術

ここでは、近年インターネットにおける情報共有技術として注目を集めている「コンテンツマネジメントシステム」、「ブログ」、「RSS」と、情報を検索する際の技術である「セマンティックウェブ」についてふれる。

4.1.1. コンテンツマネジメントシステム

コンテンツマネジメントシステム (CMS: Content Management System) は、データベースを利用してウェブサイトを構築、管理するシステムである。PHP⁶ や MySQL⁷ などを利用して構築されており、コンテンツ管理システムともいわれる。モジュールにより機能を追加したり、テンプレートにより全体のデザインを容易に変更することができる。Wiki⁸ や XOOPS⁹ などフリーソフトウェアとして提供されているものも多い。後述するブログを構築するソフトウェアも CMS の一種である (出典: フリー百科事典『ウィキペディア (Wikipedia)』)。

4.1.2. ブログ

ブログ (ウェブログ、Blog、Weblog) とは狭義には World Wide Web (Web) 上のウェブページの URL とともに覚え書きや論評などを加え記録 (Log) している Web サイトのことである。“Web を Log する” という意味で Weblog と名付けられた。現在、より頻繁に用いられている広義には作者の個人的な体験や日記、特定のトピックに関する必ずしも Web に限定されない話題などのような、時系列で比較的頻繁に記録される情報についての Web サイト全般を含んでいる。このような Web サイトの作成機能を提供するソフトウェアやサービスなどを指して呼ぶ場合もある。また、写真画像を主体として更新されるブログはフォトログ (photolog、fotolog)、ビデオコンテンツを配布するビデオブログはヴログ (vlog) と呼ばれ、通常、投稿にはテキストやイメージが添えられる (出典: フリー百科事典『ウィキペディア (Wikipedia)』)。

4.1.3. RSS

RSS は、ブログなど各種のウェブサイトの更新情報を簡単にまとめ、配信するための幾つかの文書フォーマットの総称である。

● RDF Site Summary

⁶ 動的に Web ページを生成する Web サーバの拡張機能の一つ。(http://e-words.jp/ から引用)

⁷ TCX DataKonsultAB 社などが開発している、オープンソースのリレーショナルデータベース管理システム (RDBMS)。(http://e-words.jp/ から引用)

⁸ Web ブラウザから簡単に Web ページの発行・編集などが行なえる、Web コンテンツ管理システム。Web サーバにインストールして Web ブラウザから利用する。電子掲示板 (BBS) に近いシステムだが、BBS が時系列に「発言」を積み重ねるコミュニケーションツールであるのに対し、Wiki は、内容の編集・削除が自由なこと、基本的に時系列の整理を行なわないことから、誰もが自由に「記事」を書き加えていくコラボレーションツール、もしくはグループウェアと言える。柔軟性が高く、手軽に始められて操作が簡単なことから、メモ帳代わりに使ったり、簡易なコンテンツ管理システムに利用したりする人も多い。(http://e-words.jp/ から引用)

⁹ XOOPS (ズープス、eXtensible Object Oriented Portal System) は、GPL に基づいて開発された CMS。(フリー百科事典『ウィキペディア (Wikipedia)』から引用)

- Rich Site Summary
- Really Simple Syndication

上記の様に“RSS”という語には、名称の異なる複数の規格が存在しており、それぞれ記述方法や用途が異なるために、多少の混乱が見られる。日本においては、普及している RSS のほとんどが RSS 1.0 を用いているため、一般には、RSS は RDF Site Summary として認識されている。RDF とは Resource Description Framework の略である。

ウェブログでの更新情報の配信として用いられている場合が大半を占めているが、新聞社などのニュース配信サイトでは最新ニュースを、テレビ局・ラジオ局では番組情報を、その他各種企業においてプレスリリースや新製品情報、サポート情報を、RSS を使ったヘッドライン情報として配信する事例も増えている（出典：フリー百科事典『ウィキペディア (Wikipedia)』）。

最近では、携帯電話でも RSS 情報が利用可能であり、今後はメールと並ぶ新しい情報配信ツールとして急速に利用が拡大していくと考えられる。

4.1.4. セマンティックウェブ

Web 上には現在数え切れないほど多くの文書やデータがあるが、これらの中から有益な情報を得ることや、必要なサービスを有効に利用するのは意外と難しい。通常必要な情報を検索する際には、その情報を端的に示すいくつかのキーワードを抽出し、その論理演算によって実行するが、必ずしも欲しい情報だけが手に入られるわけではない。Google (www.google.co.jp) や Yahoo! (www.yahoo.co.jp) などの検索エンジンなどにもいろいろな工夫がなされているが、検索結果の中には、不要な情報が多く含まれていることが多い。ガートナー社も、Web 上の情報量は 6 ヶ月で倍になっているので検索結果は悪化する一途であると発表している。また、キーワードの組み合わせだけでは検索することができない情報も多数存在しているのも事実である。コンピュータに関して十分知識のある人はインターネットの情報を有効に活用できるが、それ以外の人はインターネットの情報を十分に活用できなくなり、デジタルデバイド¹⁰が拡大するのではないかと懸念ももたれている（情報処理、2002）。

こうした懸念を克服する Web 技術としてセマンティックウェブが注目されている。セマンティックウェブとは、コンピュータに情報の意味（セマンティクス）を理解させることで、情報機器間のやりとりを人間を介さずにコンピュータ自身が解釈できるようにする技術である。ウェブページ上のコンテンツの内容に関する情報である「メタデータ」を持たせることで、コンピュータにその内容が理解できるようにし、情報を自動処理させようというもので、この技術により、検索結果の精度向上や、Web 上の情報の有効利用が実現できるとされている。

4.2. 情報の標準化動向

デジタルアーカイブの置ける標準化は、利用者がストレスなく自由にさまざまな情報を取り出せるようにするために、(1)ファイル形式の標準化と、(2)目録情報の記載方法の標準化、の2方向から進められている（デ

¹⁰ パソコンやインターネットなどの情報技術 (IT) を使いこなせる者と使いこなせない者の間に生じる、待遇や貧富、機会の格差。個人間の格差の他に、国家間、地域間の格差を指す場合もある。若者や高学歴者、高所得者などが情報技術を活用してますます高収入や雇用を手にする一方、コンピュータを使いこなせない高齢者や貧困のため情報機器を入手できない人々は、より一層困難な状況に追い込まれる。いわば、情報技術が社会的な格差を拡大、固定化する現象がデジタルデバイドである。また、先進工業国が情報技術によりますますの発展をとげる一方で、アフリカなどの途上国が資金難や人材不足、インフラの未整備などで情報技術を活用できず「置き去り」にされ、経済格差が拡大するのは、国家間、地域間のデジタルデバイドと言える。デジタルデバイドは、もともと貧富や機会の差が激しかったアメリカで問題となった現象で、2000年夏の沖縄サミットでは議題として取り上げられるなど、地球規模の新たな問題として注目されている。(http://e-words.jp/ から引用)

デジタルアーカイブ白書2005)。

アーカイブの世界の記述標準化には、大きく分けて2つのアプローチが存在する。1つは、北米を中心とした図書目録の記述ルールである MARC¹¹ をアーカイブの記述に適した形「MARC - ARC」に展開する方法である。この方法は、図書館の既存の仕組みを活用することで、十分とはいえないまでも、早くから情報の共有化を可能にしてきた。もう1つは、イギリスを中心とした資料群ごとに多様な階層構造を持つアーカイブの特性をよく反映した記述標準が整備されてきた。そうした中で、国際文書館評議会 (ICA¹²) は1980年代末に目録記述の国際標準を策定しようと動き出し、1993年には最初の標準である ISAD (G)¹³ が提示された。

一方で、アメリカでは SGML¹⁴ を使ったアーカイブ記述のためのタグ開発プロジェクトが1993年から始まり、EADが提案された。現在は2002年に公開された XML¹⁵ バージョンが普及している。EADでは、ISAD (G)、MARC、Dublin Core¹⁶ のそれぞれの記述項目に公式に対応させており、アーカイブ記述情報のオンライン共有が現実のものとなっている。

5. 著作権との関係

デジタルアーカイブ白書2005でも指摘されているように、デジタルアーカイブを進める場合、層が進むほど一般的に権利関係者が増えていくことが予想される。したがって、重要なことは3章で説明した各層で、関係者と明確な契約・覚え書きなどを結び、使用目的、どのような媒体に使用するか、いつまで使用するか、使用料金などを明確にしておくことである。

また、著作権との関係も考慮する必要がある (表1参照)。

表1 著作権との関係

複製権	著作権法第21条：「著作者は、その著作物を複製する権利を専有する」
公衆送信権	著作権法第23条：「著作者は、その著作物について、公衆送信（自動公衆送信の場合にあっては、送信可能化を含む）を行う権利を専有する」
複製権の制限規定	著作権法第30条：「私的使用のための複製」 著作権法第31条：「図書館等における複製」 著作権法第42条：「立法又は行政の目的のために内部資料として必要と認められる場合」

総務省デジタル資産活用戦略会議の資料¹⁷を基に表を作成

著作権と関係するところでは、電子データをダウンロードしたり、複製したり、蓄積する際の複製権があ

¹¹ アメリカ・アーキビスト協会が開発し、階層構造を持つ資料の記述などを実現した。

¹² International Council on Archives

¹³ General International Standard Archival Description

¹⁴ Standard Generalized Mark-up Language。文書の論理構造、意味構造を記述する言語。タイトル、引用部分、著者など文書の中で特別な意味をもつ部分にマーク付けをすることができる。(http://e-words.jp/ から引用)

¹⁵ eXtensible Markup Language。文書やデータの意味や構造を記述するためのマークアップ言語の一つ。マークアップ言語とは、「タグ」と呼ばれる特定の文字列で地の文に構造を埋め込んでいく言語のことで、XMLはユーザが独自のタグを指定できることから、マークアップ言語を作成するためのメタ言語とも言われる。もともと、同じく独自のタグを指定可能な「SGML」のサブセットとして考案され、任意のデータをHTMLと同様の感覚で送受信できることを目標に作成されたものである。XMLはその性質上、他のマークアップ言語の骨組みとして使用されることが多い。(http://e-words.jp/ から引用)

¹⁶ Dublin Core Metadata Element Set (Dublin Core) はインターネット上での情報資源の発見 (Resource Discovery) を目的として提案されたメタデータである。Dublin Coreは多様な分野の様々な情報資源に対応するため基本的な15要素を定め、現在ネットワーク上での情報資源を表すためのコアメタデータとして注目されている。

¹⁷ http://www.soumu.go.jp/joho_tsusin/policyreports/chousa/digital/

げられる。複製権はその目的に応じて規定が違う。また、掲載ホームページからの利用や提供の際に関係する公衆送信権などもある。さらに、リンクの張替えや、文字コードを変換する際に同一性保持権が関係する可能性もある。

6. デジタルアーカイブを公開しているWebサイトの調査

図書館や各種研究・教育機関に対して JDAA が行った調査によると、Web サイトを開設する図書館のうち、デジタルアーカイブを公開しているところは大学図書館で13.9%、公共図書館で42.6%、公文書館で53.1%であった。Web サイト上にデジタルアーカイブのある図書館の運営母体の内訳は、大学図書館では国立大学が59.8%、公立大学が4.9%、私立大学では35.4%であり、公共図書館では、80.8%が都道府県立で、残りの19.2%が市立ある。公文書館でも82.4%が都道府県立であり、国立は11.8%であった。この結果から、デジタルアーカイブがまだ十分に Web 上に公開されていないという現状が明らかになった。

次に、デジタルアーカイブの内容だが、「書籍」をデジタル化し、公開しているところが多い。デジタルアーカイブを公開している図書館のうち、大学図書館と公共図書館では「手書き以外の書籍資料」のデジタルアーカイブが最も多い。また、表示方法だが、動画資料のデジタルアーカイブを公開している図書館は少なく、ほとんどが静止画である。デジタルアーカイブを公開している図書館のうち、動画資料があるのは、大学図書館で2.4%、公共図書館で15.4%、公文書館ではまったく存在しない。ただ今後は、コンピュータ技術（ハードウェア及びソフトウェア）の進歩により、誰でもより簡単に動画情報を作成できるようになることが予想されるとともに、ネットワークのブロードバンド化もより進むと考えられるので、動画資料の公開は進んでいくものと考えられる。

通常、デジタルアーカイブには公開している各種学術・文化資料に関する説明文が付け加えられていることが多い。デジタルアーカイブを公開している図書館のうち、デジタルアーカイブに関する何らかの説明文があるところは、大学図書館で73.2%、公共図書館で96.2%、公文書館で70.6%である。

また、デジタルアーカイブの資料表示における付加機能の導入率では、拡大表示が高い。ZOOMA¹⁸等により、高精彩を維持しながら、古文書や古地図などを自由に拡大・縮小して閲覧できる。資料の拡大表示機能は、大学図書館の48.8%、公共図書館の15.4%、公文書館の17.6%で提供されている。ただ、そのほかの付加機能（回転・多角表示機能、解説・補助画像、仮想空間、音声等）の導入率は低いことも明らかになった。

6.1. 国内の事例

次に、写本などの各種学術・文化資料をデジタルアーカイブとして公開している国内の主な Web サイトである国文学研究資料館 (<http://www.nijl.ac.jp/>)、HUMI プロジェクト（慶應義塾大学、<http://www.humi.keio.ac.jp/>)、KACHINA（可知納）プロジェクト（立命館大学、<http://www.arc.ritsumei.ac.jp/kachina/>) を概観する。

● 国文学研究資料館 (<http://www.nijl.ac.jp/>)

国文学研究資料館は、諸学会や研究者コミュニティのもとに、昭和47年5月に大学共同利用機関として設立され、(1)文学資源研究系、(2)文学形成研究系、(3)複合領域研究系、(4)アーカイブ研究系の4つの研究系とし、日本文学資料・歴史記録資料の総合的な研究に取り組み、他の機関とともに人間文化研究に対応する態勢をとっている。国文学研究資料館のサイトは、日本文学に関する研究情報と、テキスト及び画像による資料の提供を行っており、数量が豊富でよく整理されたデータベースが構築されている。また、更新もよくされており信頼度の高さがうかがえる。

¹⁸ ZOOMA は、高解像度画像データをスピーディに表示する画期的なビューワ。データを細かいメッシュ構造に分割して、必要なパーツを表示するため高精細な画像のままネットワークに負担をかけることなく高速表示が可能。

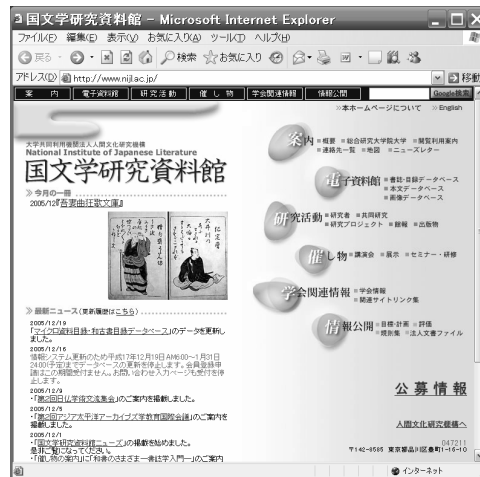


図7 国文学研究資料館のサイト

● HUMI プロジェクト (慶應義塾大学、<http://www.humi.keio.ac.jp/>)

慶應義塾大学で行われている HUMI プロジェクトの「HUMI (HUmanities Media Interface) ¹⁹⁾」は「人文科学研究における新しい情報伝達手段の導入」を意味している。このプロジェクトは、デジタル技術を積極的に導入することによって、新しい研究方法や研究分野を開拓することを目標として1996年の春に発足し、文学部のスタッフが中心となって稀覯書(歴史的価値のある古い書物)のデジタル化と研究を行うほか、理工学部のスタッフによる画像処理技術の研究や、非破壊検査技術を用いた古い書物の分析、あるいは、環境情報学部や法学部のスタッフによる情報マネジメント手法の研究、デジタルデータの知的財産権問題の研究など、多角的な研究が行われている。

HUMI プロジェクトが掲げる主なテーマは以下のとおりである。

- ・ 稀少本のデジタル化
- ・ デジタル化手法の研究
- ・ 海外図書館との協力関係
- ・ 画像データベースの構築
- ・ デジタル・アーカイヴィングと研究環境
- ・ 国際的なネットワーク

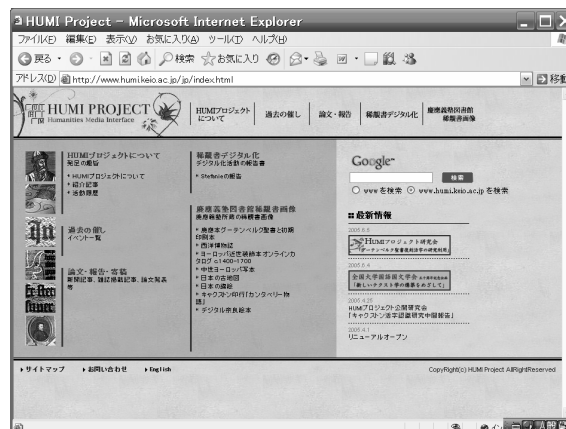


図8 HUMI プロジェクト

¹⁹⁾ 「HUMI (ふみ)」という言葉には日本語の「文、史、典」などの連想が働いているとのこと。

● KACHINA (可知納) プロジェクト (立命館大学、<http://www.arc.ritsumeit.ac.jp/kachina/>)

立命館大学で行われている KACHINA (可知納) プロジェクトでは、インターネットを使ったデジタルアーカイブの知的発信技術に関する研究開発を行っている。

KACHINA (可知納) プロジェクトでは、知的エージェント技術と従来のデータベース技術やデジタルアーカイブ技術を融合させ、下記の機能を備えた、自己増殖型知的デジタルアーカイブの実現を目指している。

- ・ インターネット経由でデジタルアーカイブにアクセスする多様なユーザに対して、効果的な知識継承と共有を実現するためには、アーカイブを参照するユーザが持つ知識、趣向、ゴールなど深いレベルで理解し、それに適応した情報探索や加工技術
- ・ アーカイブに蓄積された知的資産をベースに新たな知識創造を行っていくためには、インターネットコミュニティに点在する多様な知識を獲得し、それを元にアーカイブを成長させる仕組み

具体的には、以下の基盤記述に関する基礎研究と実験を行っている。

- ・ ユーザの形式的／暗黙的知識を獲得・成長させるイシューベース技術
- ・ ユーザモデルに基づくユーザ適応型知的エージェント技術
- ・ 情報検索のためのマルチモーダルユーザインタフェース技術

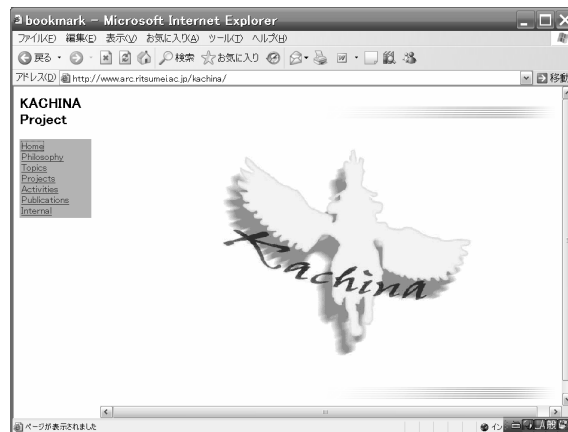


図9 KACHINA (可知納) プロジェクト

6.2. 国外の事例

次に、海外のデジタルアーカイブを公開している主な Web サイトである Oxford Text Archive を概観する。

● Oxford Text Archive (<http://ota.ahds.ac.uk/>)

オクスフォード・テキスト・アーカイブは1976年に Lou Burnard によって設立された、人文学研究のための多言語テキストアーカイブである。寄託を受けた種々の言語のテキストを集積して、磁気テープやフロッピーディスクの形で希望者に提供するサービスを行っているほか、現在はウェブサイトでテキストのカタログが閲覧できる。量は二千五百点強だが、東洋語をふくめ数十の言語が含まれている。しかし、寄託の条件などの理由で、利用には比較的制約が多く、またウェブサイトから直接テキストをダウンロードできるとは限らない。

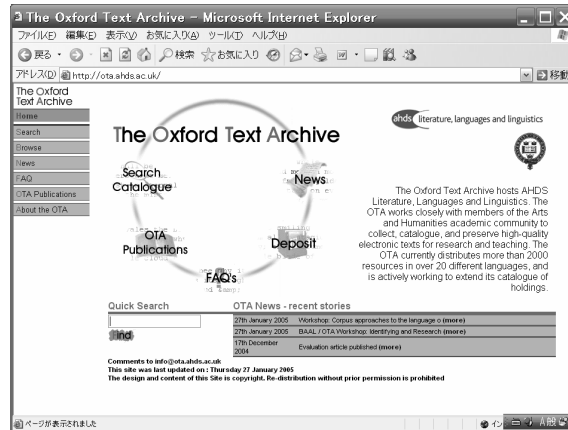


図10 Oxford Text Archive

この他に、現存する古英語の文献をすべて電子化したトロント大学の古英語コーパスである The Dictionary of Old English (<http://www.doe.utoronto.ca/>、図11参照) や、メルボルン大学の中世のテキスト及ぶ原稿に関するデジタルファクシミリと電子版のプロジェクトである Medieval Multimedia (<http://www.medieval.unimelb.edu.au/>、図12参照) などが存在する。

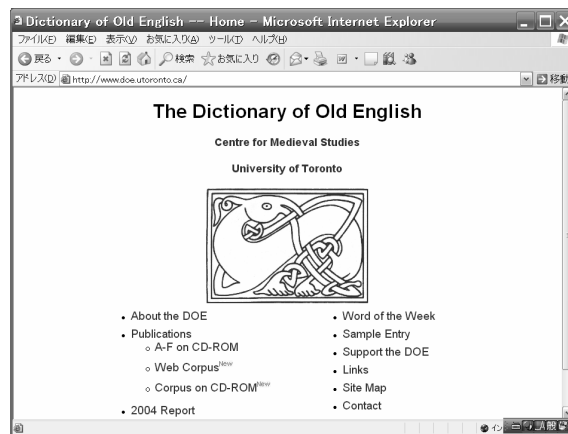


図11 The Dictionary of Old English

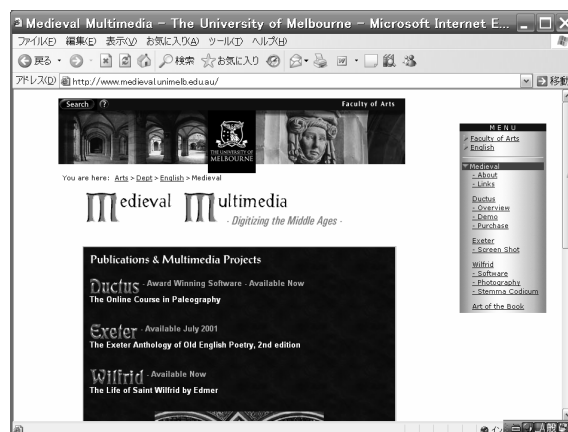


図12 Medieval Multimedia

6.3. 本学における事例

最後に本学におけるデジタルアーカイブ公開の状況を概観する。

本学においても、生田および神田の図書館において、貴重な書籍・資料、各種学術・文化資料等（表2に本学が保有している貴重な書籍・資料、各種学術・文化資料等を示す）をデジタルアーカイブとして公開するという試みが行われている。

表2 専修大学が保有している貴重な書籍・資料、各種学術・文化資料等

保管場所	資料
専修大学図書館（本館）	<ul style="list-style-type: none"> ●菊亭文庫 京都の今出川家に伝わる資料群。和歌に関する写本、詠草、雅楽に関する資料、宮廷の儀式関係、日記類を中心とする諸記録など3,448点。 ●蜂須賀家旧蔵本 鎌倉・室町・江戸期に書写された和歌・物語などの古鈔本で、16世紀から幕末にかけて阿波一国を領有した蜂須賀家旧蔵本の一部です。「長秋詠藻」、「三十六人和歌」など21点、129冊。「専修大学図書館蔵古典籍影印叢刊」として半数を復刻、出版。 ●ミシェル・ベルンシュタイン文庫 「人権宣言」をはじめ、「ルイ16世の遺言書」や当時の風刺画やポスターなどフランス革命期にフランス国内、周辺諸国で刊行、記録された原資料4万数千点からなる。「専修大学図書館蔵古典籍影印叢刊」として文庫資料の一部を復刻、出版。 ●『算術、幾何、比及び比例総覧』（略称『スママ』）ルカ・パチョーリ著 初版1494年 Su [m] ma de arithmetica geometria proportioni [et] proportionalita.
専修大学図書館神田分館	<ul style="list-style-type: none"> ●今村力三郎文庫 明治・大正・昭和にわたり法曹界で活躍した今村力三郎先生（元・専修大学総長）旧蔵の訴訟記録417件（大逆事件、金剛事件、虎ノ門事件、五・一五事件、神兵隊事件、帝人事件等）及びこれらの事件の弁護資料。 ●中川善之助文庫 家族法の権威として知られた中川善之助先生（東北大学名誉教授、元金沢大学学長）の旧蔵書で図書5,758冊、資料1,900点、雑誌65誌。J.Bentham 著 Traits de l'igislation civile et pnale, ... 3V. (1802)、わが国の民法典の元となった「民法草案人事編理由書」などの稀覯本も含まれる。 ●石井良助文庫 元本学教授で図書館長も勤められたわが国法制史学の泰斗、石井良助先生の旧蔵書及び文書・資料類を集めたもの。中心は約5,000点をこえる文書（近世の村方文書）類であり、わが国有数の法制史料集のひとつである。 ●ザイドル文庫 エルヴィン・ザイドルが長年かけて収集した古代法史・法パピルス学関係、ローマ法・法制史関係コレクション。抜き刷りなども丹念に収集されている。約5,000冊。

<http://www.acc.senshu-u.ac.jp/library/collection/ikuta.html> を基に表を作成

このうち、『算術、幾何、比及び比例総覧』（略称『スママ』、監修：商学部教授 佐々木重人）については、その概要や時代背景、人物紹介や参考文献など、書籍そのものの情報だけではなく関係する情報についての記述があることから、単なる広報手段ではなく、所蔵品を鑑賞したり、研究に利用することが可能になっている（図13参照）。ただ、そのすべてが公開されているわけではなく、検索機能などが提供されているわけではないので、誰もが簡単に利用できるという性格のものではない。

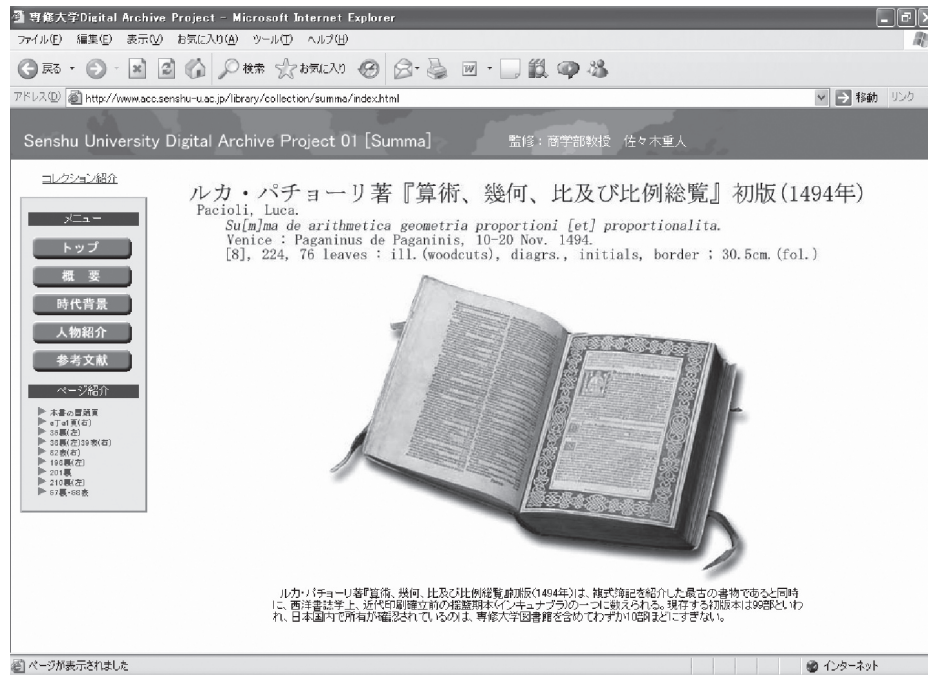


図13 『算術、幾何、比及び比例総覧』（略称『スママ』、監修：商学部教授 佐々木重人）

7. まとめ

本論文は、平成17年度に選定されたオープン・リサーチ・センター整備事業である『Anglo-Saxon 語の継承と変容』において、中世英文学における手稿写本 (Manuscript) の文字研究の成果をデジタルアーカイブとして公開していく上で必要となるインターネットにおける情報公開の特徴と、既存のデジタルアーカイブ、及び関連する情報技術を概観したものである。

まず、インターネットを使った情報公開の特徴についてまとめると、以下のように集約されると考えられる。

- インターネットは、単なる知識提供の手段ではなく、既存の資産をベースにした新たな知の創造の場になってきている
- インターネットを効果的に利用することで、個人の持つ知識や情報をナレッジとして関係者全体で共有し、有効に活用できる
 - この場合の知識・情報とは単なるデータである「形式知」だけではなく、経験則や仕事のノウハウといった、普段はあまり言語化されない「暗黙知」までを含む

したがって、有形・無形の文化資産をデジタルアーカイブとして Web 上で公開することで、誰もが自由に閲覧・鑑賞できるだけではなく、人文科学や情報科学といった垣根を越えた学際的なコラボレーションも容易に行えるようにする研究基盤を構築することができる。また、このような研究基盤を構築することは、その国に存在する基礎資料や研究成果が活用可能な状態でどの程度整備されているかを示し、その国の学術研究レベルを測る指標ともなるものであり、非常に重要な意味を持つ。さらに、アナログ情報である有形・無形の文化資産を、記録精度が高く、再現性に優れたデジタル情報化し、世界規模の情報通信ネットワークを利用して情報の受発信を行うことにより、次の世代に正しく継承することも可能となる。

次にデジタルアーカイブの利用方法であるが、蓄えられたデジタルアーカイブの情報を探し出すには、名称やフレーズなどに関する言葉をキーワードとして探すことになる。したがって、必要とする情報を効率よく取得するためには、利用者が、ある程度専門的な知識を持つ必要がある。しかし、利用者すべてがそのよ

うな知識を持っているわけではない上に、類義語や同義語なども多数あり、またキーワードだけでは特定しづらい情報も多数存在しているため、利用者が必ずしも必要とする情報を容易に取得できるわけではない。したがって、誰もが利用可能な Web 上でデジタルアーカイブを公開する際は、専門知識がない子供などでも容易に使えるように、画像などを閲覧しながら似たものや関連するものを探す工夫や、類似するキーワードでも検索できるように、シソーラス²⁰ やセマンティックウェブ、連想検索等の技術を用いて実現する必要がある。さらに、作成しても利用されていない、もしくは利用しづらいデジタルアーカイブは、その存在価値がほとんどないと考えられる。利用しやすいデジタルアーカイブを作るためには、デジタルアーカイブ白書 2005でも指摘されているように、①誰に利用して欲しいのかという点と、②どのように利用して欲しいのかという点を意識し、以下に示す5つのポイントを抑える必要がある。

- 説明のしかた
- 目次や見出しの立て方
- 時事性や話題性の盛り込み方
- データ間のリンクの張り方
- 情報の知らせ方

また、総務省のデジタル資産活用戦略会議でも指摘されているように、各種アナログデータをデジタル化し、デジタルアーカイブとして Web 上で公開するためには、以下に示すような課題を考慮し、検討していく必要がある。

- 技術的課題
 - 媒体寿命：
記録される磁気ディスク、磁気テープ、CD-ROM、DVD 等の磁氣的／光学的記録媒体の寿命は紙・マイクロより短い
 - 技術の旧式化：
情報内容は、再生ソフトや機器に依存する。それらの寿命は媒体よりはるかに短く、数年で世代交代する（再生技術がないために利用できない場合が多発する）
- 経済的課題
エミュレーション、マイグレーション等の技術的コスト、権利処理コスト、システム開発コスト等、膨大なコストを要する
- 制度的課題
文化財の保護と知的財産保護の問題が拮抗する。保存のための改変を著作権上どのように取扱うかなどを検討する必要がある
- 組織的課題
膨大な電子情報の保存は、特定機関だけでできるものでなく、他機関（制作者、配信者、利用者、図書館等）との協力が必要である

今後は、本論文で概観したインターネットによる情報公開の特徴を活かしつつ、利用者を意識し、上述した課題などを十分に考慮した上で、中世英文学における手稿写本 (Manuscript) の文字研究の成果をどのよ

²⁰ シソーラスとは、単語の上位／下位関係、部分／全体関係、同義関係、類義関係などによって単語を分類し、体系づけた辞書。書籍の形態のシソーラスと、電子化されデータベース化されたシソーラスがある。電子化されたシソーラスは、自然言語処理などで使われる。日本語の電子化されたシソーラスとしては、分類語彙表、EDR 概念体系辞書などがある。英語では、WordNet が有名。(フリー百科事典『ウィキペディア (Wikipedia)』から引用)

うにデジタルアーカイブとして公開していくか方針を検討し、その方針に沿って公開を進めていく予定である。

参考文献

- 新井範子、植竹朋文、「インターネット上の情報の所在と広がりについての分析」、『第26回消費者行動研究コンファレンス予稿集』、日本消費者行動研究会、2003
- 植竹朋文、「学部ホームページの利用状況の分析及びその考察」、『専修経営学論集』、第79号、2004、pp. 41-67
- 植竹朋文、新井範子、「インターネット上の情報の所在と広がりについての分析—映画におけるリアルな情報とバーチャルな情報の関連性—」、『専修経営学論集』、第77号、2003、pp. 19-50
- 植竹朋文、新井範子、「電子掲示板上のユーザ評価の推移」、『経営情報学会2004年秋季全国研究発表大会予稿集』、経営情報学会、2004、pp. 502-505
- 笠羽晴夫、『デジタルアーカイブの構築と運用』、水曜社、2004
- 情報処理学会、『情報処理』、情報処理学会、Vol.43、No. 7、2002
- 総務省、『情報通信白書（平成16年版）』、総務省、2004
- デジタルアーカイブ推進協議会（JDAA）、『デジタルアーカイブ白書2005』、トランスアート、2005
- 古川一郎（編）、『デジタルライフ革命』、東洋経済新報社、2001

参考URL

- <http://anglo.fm.senshu-u.ac.jp/>（専修大学大学院社会知性開発研究センター 言語・文化研究センター）
- <http://e-words.jp/>（IT用語辞典 e-Words）
- <http://it.nikkei.co.jp/internet/column/>（NIKKEI NET IT プラス）
- <http://ja.wikipedia.org/>（フリー百科事典『ウィキペディア（Wikipedia）』）
- <http://ota.ahds.ac.uk/>（The Oxford Text Archive Website）
- <http://www.acc.senshu-u.ac.jp/library/collection/ikuta.html>（専修大学 図書館 コレクション紹介）
- <http://www.arc.ritsumei.ac.jp/kachina/>（立命館大学 KACHINA（可知納）プロジェクト）
- <http://www.dcaj.org/jdaa/>（デジタルアーカイブ推進協議会（JDAA）関連情報）
- <http://www.doe.utoronto.ca/>（The Dictionary of Old English, University of Toronto）
- <http://www.humi.keio.ac.jp/>（慶應義塾大学 HUMI プロジェクト）
- <http://www.medieval.unimelb.edu.au/>（Medieval Multimedia, The University of Melbourne）
- <http://www.mext.go.jp/>（文部科学省）
- <http://www.nijl.ac.jp/>（国文学研究資料館）
- http://www.soumu.go.jp/joho_tsusin/policyreports/chousa/digital/（総務省の情報通信政策に関するポータルサイト 調査研究会デジタル資産活用戦略会議）