

知のデジタルアーカイブへの期待 Toward Development of Digital Archives of Comprehensive Intellectual Resources

杉本重雄

筑波大学・図書館情報メディア系・知的コミュニティ基盤研究センター

Shigeo Sugimoto

Research Center for Knowledge Communities, Faculty of Library, Information and Media Science

University of Tsukuba

sugimoto@slis.tsukuba.ac.jp

概要

地域の記憶と記録を将来に残すこと、大規模災害から文化的資産を守ること、そして地域振興を進めることに関してデジタルアーカイブは大きな役割を持っている。図書館や博物館等におけるデジタルアーカイブの重要性は広く認められているが、我が国においては大規模館や一部の先進的な機関を除いてはあまり積極的に進められているとは言えない。総務省が平成 23 年 2 月から平成 24 年 3 月まで組織した「知のデジタルアーカイブに関する研究会（知デジ研）」の提言では、デジタルアーカイブを推進する上でのいくつかのキーポイントを示している。本稿では、デジタルアーカイブとは何かを議論した後、知デジ研の提言を簡単に紹介するとともに、筆者の視点からデジタルアーカイブの開発推進のための重要な話題として、デジタルアーカイブとクラウド（Cloud）環境、ネット上でのデジタルアーカイブのためのメタデータ情報基盤に関して述べる。

キーワード： 総務省・知のデジタルアーカイブに関する研究会（知デジ研）、メタデータ相互運用性、メタデータ情報基盤、デジタルアーカイブの長期利用、デジタルアーカイブネットワーク

abstract

It is widely recognized that digital archives have a large potential for a community to keep their memories and records for the future, to protect their valuable resources from natural and human originated disasters, and to promote the regional economy and culture. In Japan, most digital archives are developed only by major national memory institutions and some regional institutions. However, many medium or small regional institutions are reluctant to develop digital archives because of the lack of their human and financial resources. Ministry of Internal Affairs and Communication (MIC) organized a research group on digital archives (Chideji-ken) from February 2011 to March 2012. The group published a recommendation to promote digital archives in Japan where some key issues are discussed. This paper first overviews digital archives to help understand the fundamental features of

digital archives. It briefly shows the recommendation of the group. It discusses cloud computing environment for digital archives and metadata infrastructure for networked digital archives followed by concluding remarks.

Keywords: Research Group on Digital Archives at Ministry (Chideji-ken) hosted by Internal Affairs and Communication (MIC), Metadata Interoperability, Metadata Information Infrastructure, Long-term Use of Digital Archives, Digital Archive Network

1. はじめに

総務省における「知のデジタルアーカイブに関する研究会（知デジ研）」[1]の提言が出されてからちょうど1年がたつ。知デジ研では、地域の図書館、博物館、美術館、文書館など文化や社会の記憶と記録にかかる機関や組織におけるデジタルアーカイブを中心に議論した。この問題意識の背景には、我が国におけるデジタルアーカイブ構築が先進諸国に比べて決して進んでいるとは言えないこと、特に地域の MLA による構築が進まないことがある。加えて、数年前から始まった電子書籍の普及、オープンデータの動きなど、デジタルアーカイブに関係する様々な動きがあったこともある。そして、知デジ研での議論の開始直後に東日本大震災が起こり、震災の記録、災害への備えとしてのデジタルアーカイブの役割が議論された。

デジタルアーカイブは、MLA が持つ文化財をデジタル化、データベース化して提供するものととらえられる。こうした種類のものは 1990 年代に始まったデジタルライブラリ開発ブームの当初から中心的なテーマであり、決して最近始まったものではない[2]。たとえば、国立国会図書館の近代デジタルライブラリ、国立公文書館のデジタルアーカイブ、文化庁の文化遺産オンラインのほか、地域 MLA が進めるものはそうした種類のものである。これらは、比較的一様な標準化された技術を用い多数の文化財をデジタル化し、提供するものである。その一方、高度なセンシング技術、グラフィックス技術、画像処理技術などを用いて、高精細、高機能なデジタルコンテンツを提供するものもある。Web 上のコンテンツを収集、蓄積、保存する Web アーカイブもデジタルアーカイブの代表例である。電子書籍や電子ジャーナルのような電子出版物の場合も、長期の利用を考えるとデジタルアーカイブのすぐ隣に位置づけられる。

知デジ研の提言によると、デジタルアーカイブを、「何らかの方針に基づき、デジタルコンテンツを選択、収集、組織化、蓄積し、長期にわたって保存するとともに利用に供するシステム又はサービス」、そして知のデジタルアーカイブを、「人間・コミュニティの知的活動を支えるためのデジタルアーカイブであり、デジタル・ネットワーク社会の知識インフラ」と定義している。また、後述するように、提言では、デジタルアーカイブ化による知的資産の公開を進めること、デジタルアーカイブ推進のための人的基盤を整備す

ること、デジタルアーカイブ構築を支えるためのシステム基盤構築を進めること、コンテンツ流通のためのメタデータ基盤を整備することが提案されている。

デジタルアーカイブということばは広く用いられており、知識インフラとしてのデジタルアーカイブの重要性も広く認められている。一方、デジタルアーカイブには、非常に高度な情報技術を用いて作られる超高品質なものから標準的な品質のコンテンツを非常に大量に持つものまで、そしてデジタル化した歴史資料のアーカイブからボーンデジタル資料のアーカイブまでというように、いろいろな種類のアーカイブを表すことばでもある。本稿では、概念整理のために、はじめにデジタルアーカイブということばに関していくつかの視点から述べる。その後、知デジ研提言に関して述べ、クラウド環境とメタデータ基盤について議論する。

2. デジタルアーカイブについて

2.1 デジタルアーカイブということばを考える

頭文字をとって MLA と総称される図書館 (Library)、博物館や美術館 (Museum)、公文書館、文書館 (Archives) は、人間の知的な活動を支え、記録する機能を持つため知の記憶機関 (Memory Institution あるいは Organization) と呼ばれる。MLA はデジタルアーカイブを実現する主たるプレイヤーである。MLA 連携ということがよく話されるが、デジタルアーカイブは MLA 連携のキーとなるものである。その一方、M、L、A それぞれの性格や置かれている環境の違いによってデジタルアーカイブへの取り組みは同じではない。

本稿では、知デジ研に倣ってデジタルアーカイブを「何らかの方針に基づき、デジタルコンテンツを選択(Select)、収集(Collect)、蓄積・組織化し(Organize)、利用に供する(Access)とともに、長期にわたって保存する(Preserve)システム又はサービス」と定義する。頭文字で表す SCOAP は MLA に共通の基本的な機能であるということもできる。

電子図書館、デジタルライブラリ (Digital Library) や電子博物館(美術館)・デジタルミュージアム(Digital Museum)といったことばとデジタルアーカイブということばを区別することに大きな意味があるとは思えないが、敢えて違いを考えてみよう。デジタルライブラリやデジタルミュージアムといったことばが、ライブラリやミュージアムの機能全般の意味合いを持つのにに対してデジタルアーカイブということばの場合は、デジタルコンテンツに関する SCOAP に特化しているといえよう。公文書館に関しては、電子公文書館の意味でのデジタルアーカイブズもデジタルライブラリやデジタルミュージアムと同様に考えることはできる。しかし、電子的な記録管理を意味する Electronic Records Management や Electronic Records Archive という名前で呼ばれることが多いようである。いずれにしても電子公文書館としてデジタルアーカイブズと呼んでしまうと、より一般的な意味で広く使われているデジタルアーカイブということばと区別することが難しく文脈で判断するしかなくなるので。わかりにくくなる。あまり名前にこだわってもことば遊び

になってしまうので、ここではこれ以上議論しないことにする。いずれの場合にも、デジタルアーカイブはデジタルライブラリ等の中核的な機能である。

欧米では、デジタルコンテンツを長期にわたって使い続けられるようにする取り組みを表すデジタルキュレーション (Digital Curation) ということばが広く用いられている[3]。長期にわたって利用できるようにすることで付加価値を作り出すことが重要な目的であり、先のデジタルアーカイブの定義と大きな違いはない。デジタルキュレーションでは文化財のみならず、科学技術情報や政府情報なども重要な領域として視野に入っているのも、我が国でのデジタルアーカイブという言葉が持つ響きよりは広いかもしれない。いずれにしても、重要な資源を集めて、長期にわたって保存し、利用し続けられるようにすることが目的である[4]。

2.2 いろいろなデジタルアーカイブ

ここでは、デジタルアーカイブのパターンを大別してみたい。

- ・ 高度な計測技術と記録技術を利用して、重要な文化財の高精度なデジタル化を行い、高度なグラフィックス技術を利用して利用者に提供するもの。たとえば、アンコール遺跡や平等院のデジタルアーカイブ化といったものは典型である。
- ・ 書籍や文書、文化財を対象として要求に応じた精度のデジタル化を行い、一般的な環境で利用できるコンテンツとサービスするもの。一般的に大量のコンテンツを蓄積提供する。
 - イメージデータ(静止画、動画)としてデジタル化したもの
 - 電子テキストとしてデジタル化したもの
- ・ 電子的に作られ流通するボーンデジタル (Born Digital) リソースを収集、蓄積、保存、提供するもの
 - 典型的なものは電子文書のアーカイブ
 - 文書に限らず電子書籍、ゲームなどいろいろなコンテンツがある
- ・ インターネット上で提供されるリソースを自動収集することで形成されるもの
- ・ これらのほか、個人やグループが持つリソース

デジタルアーカイブに蓄積されるリソースは様々であり、文書や図画のような平面的なリソース、3次元物体、動画、ビデオ、音や踊りなどの無形物、そしてゲームやデータベースなどのソフトウェアもある。共通する視点は先に示した SCOAP である。特に保存(P)はデジタルアーカイブにとって重要である。大規模な資料のデジタル化の場合、落ち着いた標準的な技術が用いられることが多いので、保存の面からは問題が少ないといえる。一方、先端的な技術を用いるものの場合、できあがったコンテンツの寿命は用いられた技術の寿命に依存するので一般的には長くないと考えられる。そのため、コンテンツそのものの保存に加えて、コンテンツ作成にかかわる情報、そして素材データの保存が欠かせないと考えられる。ボーンデジタルなリソースやインターネットリソースのアーカイブの場合、収集し

たリソースを正確に再現するための情報を獲得することが重要であるが、リソースそのものには必ずしもそうした情報はない。そのため、ファイル形式に関する情報を収集保存する仕組みが重要な役割を持つ。イギリス国立公文書館によるフォーマットレジストリ PRONOM はそうした情報を蓄えてくれている[5]。

MLA はそれぞれ異なる機能を持つため、リソースのデジタル化、デジタルアーカイブの組織化、特にメタデータの持ち方に関しては、それぞれの伝統をベースにした方法がある。このように、デジタルアーカイブと総称されるものの、その内容や構成方法は多様である。このことは、個別の MLA が実際にデジタルアーカイブを構築しようとする際に、どのような方法で、どのような技術を用いればよいのかが分からない、あるいはわかる人材がいらないといった問題につながることになる。

2.3 デジタルアーカイブを支える環境

何らかの条件が生じない限りコンテンツへのアクセスをせず、蓄積のみを行うダークアーカイブと呼ばれるものもあるが、一般にデジタルアーカイブはアクセスして利用すること、特にインターネット経由でのアクセスを前提としているものが多い。そのため、利用権限の認証やアクセシビリティなど種々の問題に対応しなければならない。また、現在のインターネット環境では、人間向けのユーザインタフェースのみならず、コンピュータ向けの Application Program Interface (API) を用意して第三者によるサービスを作りやすくすること、横断的な検索を行えるようにするといったことが進められている。

デジタルアーカイブはコンテンツの検索や管理のためのメタデータを持つ。ネットワーク上ではデジタルアーカイブ間の相互運用性を高めることが重要であるため、標準的なメタデータを用いることが望まれる。その一方、それぞれのデジタルアーカイブの特色を出すための独自要求を満たすことも重要であり、それにはデジタルアーカイブごとの要求を満たすメタデータが必須である。こうした相互に矛盾する要求を満たすことがデジタルアーカイブの発展には不可欠である。こうした問題の解決には WWW 上の共通でオープンな標準の上にメタデータを流通させ、意味的な相互運用性を向上することが重要である。また、そうしたメタデータ流通の環境を実現するには、メタデータの語彙やスキーマ等のメタデータ情報を流通させるための基盤の構築が求められる。

先にも述べたように、アーカイブの長期利用性とコンテンツの長期保存性がデジタルアーカイブにおける重要な課題である。デジタルコンテンツの長期保存、特にボーンデジタルリソースの長期保存は、バイナリデータの保存だけでは不十分であり、マイグレーションやエミュレーションを行う環境を必要とすることがよく知られている。デジタルコンテンツの保存（デジタル保存、Digital Preservation）は、デジタルアーカイブ、デジタルキュレーションにおけるキーである。デジタル保存において重要なことは、保存対象の情報のみならず保存の過程で作り出される情報を残していくこと、そしてそうした情報を保存することである。すなわち、保存対象とそれにかかわる情報を記述したメタデータを

長期にわたって生かし続けることが課題となる。

デジタル保存は100%解決することの難しい問題である。基本的には、物理的な物体の保存と同様に、保存のポリシーと維持管理タスクの明確化、そして保存に責任を持つ組織による持続的な作業によって解決していかざるを得ない。デジタル保存が未確立であるため、デジタルアーカイブによる保存に対する危惧をもたれることがある。有形の文化財をデジタル化して残すことは、ある意味で、デジタル形式のレプリカを作ることである。有形物に比べてデジタルデータはコピーが作りやすいので、多重のコピーを持つことによって、より安全文化財の保存につながるといえる。また、デジタル化によって、レプリカとはいえ貴重な文化財に対するアクセス性が高まることも重要なメリットである。

3. 「知のデジタルアーカイブに関する研究会（知デジ研）」より

3.1 「知のデジタルアーカイブに関する研究会」提言より

知デジ研は、2012年2月から全部で9回の会合を持ち、デジタルアーカイブに関していろいろな観点からの議論を行った。特に、中小規模のMLAによるデジタルアーカイブ開発の推進における課題、インターネットの時代におけるデジタルアーカイブを支えるメタデータ情報基盤の重要性、東日本大震災に学んだデジタルアーカイブの役割等について議論した。表1に、知デジ研提言概要を示す。これは提言より抜き出したデジタルアーカイブ推進にかかわるキーポイントである。

この概要が示すように、知デジ研は、MLA機関等が持つ知的資産公開の推進と、そのための人的基盤、システム基盤、メタデータ基盤の3つの基盤構築を提言している。知的資産公開の推進自体は当然のことと思えるが、小規模な博物館や資料館にとっては、所蔵資料のデータ管理を電子的に行うことが最初の障壁となり得る。そして、所蔵資料に関する情報をインターネット上に公開していくことでそれぞれの館の認知度は上がると考えられるが、そこまでのステップを独自に行うことが難しいため、そうした館をサポートすることが重要である。中小規模のMLAにとっては、情報技術のことを知る人材を常時確保することが難しく、また、必要な機材を長期にわたって安定的に確保し、維持することが難しいため、人的側面とシステムの側面からのサポートが必要である。この点からは、文化庁の文化遺産オンラインのサービスを利用することは、インターネット上に文化財の情報を提供していくことに関する敷居を低くしてくれることが期待できる。

3.2 デジタルアーカイブの機能階層

個別の中小MLA機関や組織が独自の要求仕様に基づいてデジタルアーカイブの構築を進めるのは効率的ではない。従来、独自のサーバの上にシステムを構築してサービスを提

表 1 知デジ研提言概要

1. 大福帳からデジタルへ。知的資産の公開
 - ・ 中小の MLA 機関が所蔵する資料等のデジタル化、ネット上への公開の推進。
 - ・ 「デジタルアーカイブ推進構築のためのガイドライン」の周知・普及及び技術・利用環境の変化に即した見直し。
 - ・ デジタルアーカイブ構築・連携を推進するための財政措置の拡充。
2. 人的基盤の構築
 - ・ デジタルアーカイブの技術・知識・ノウハウ等を収集・蓄積・共有する「デジタルアーカイブ支援ネットワーク（DAN）」の設立。
 - ・ デジタルアーカイブの推進に理解のあるリーダーの獲得及び構築・運営していく専門的な人材の必要性への理解。
 - ・ デジタルアーカイブの有機的連携を推進する「デジタルアーカイブ・スペシャリスト」の育成。
3. システム基盤の構築
 - ・ データ形式の標準化・耐災害性の観点からデータ蓄積基盤として、デジタルアーカイブ・クラウドを推進する。
 - ・ 「文化遺産オンライン」の利用促進・機能拡充（多言語対応等）。
 - ・ 東日本大震災の記憶を伝承し、将来の災害対策等に活用するため「東日本大震災アーカイブ」の構築。
 - ・ デジタルコンテンツ長期保存技術の開発を進めるとともに、ノウハウの共有を推進する。
4. コンテンツ流通基盤の構築
 - ・ 組織・コンテンツを一意に識別するための識別子の普及を進めるため、知的資産 ID の導入。
 - ・ メタデータ情報基盤（MetaBridge）の利用普及・機能拡張を進め、メタデータの流通を促進する。

供することが行われてきたが、人的資源の問題、維持管理の問題からこの方法がどんな組織にも適合するとは言えない。また、個別のシステムを作り上げることは、インターネット上でのサービス間連携、情報資源(リソース)の相互運用を難しくすることもある。また、インターネット上では、できるだけリソースを相互に利用できるようにして、利用者の目にはシステム間の壁がないように見せることが重要であることを考慮しなければならない。

MLA のデジタルアーカイブのコンテンツは多様であり、その利用者サービスもコンテンツの種類に依存する。その一方、デジタルコンテンツはバイナリデータと見れば共通であるため、データを蓄積保管するために必要な機能は共通であることに容易に気付く。そして、コンテンツの検索や管理のために必要なメタデータは、コンテンツとサービスに依存す

る部分とコンテンツ流通のために相互に共通化が必要な部分があることも理解できる。すなわち、利用者に近いところほど個別化が必要であり、反対にデータ蓄積機能のように利用者から直接見える必要のない部分は共通化が行いやすいことが容易に理解できる。こうしたことを考慮して、ここではデジタルアーカイブの構成を、以下のように3階層でとらえる。

(A) デジタルアーカイブのコンテンツを利用者に提供するためのサービスを実現するための機能

(ア) 個々のデジタルアーカイブごとの利用者要求を満たすサービスの実現

(イ) 個々のデジタルアーカイブの要求に適したコンテンツの組織化と管理

(B) デジタルアーカイブ間連携のためにコンテンツにかかわる情報、すなわちメタデータの流通、蓄積のための機能

(ア) メタデータを流通させるための標準フォーマットに基づくメタデータ交換

(イ) メタデータスキーマとメタデータ語彙の定義と情報の提供

(C) デジタルコンテンツやメタデータの実体の蓄積、保存、提供のための機能

(ア) データの実体を安定的に蓄積し、提供する機能

(イ) 個々のデジタルアーカイブからの要求に応じて、データを長期にわたって保存する機能

階層的にとらえることで、個々のデジタルアーカイブごとの要求に応じて開発すべき部分と共通化された技術や情報基盤機能を用いて実現すべき部分が明確化できる。人的資源を多く持たない中小MLAにとっては、個別に要求されるところだけに注力できればありがたい。もちろん、共通部分を他の組織と共有することによってかえってコストが大きくなるようでは意味がないので、共通部分、すなわち下位の層のコストの低減は必須の条件である。

知デジ研提言の第3と第4の項目は、それぞれ階層(C)と(B)に対応する。こうした階層的なとらえ方は情報システムの構造を考える上でごく一般的になされる問題整理の方法である。最も下の階層(C)を、データを蓄積・保存・提供するクラウド(Cloud)型のサービスととらえることで、共有の考え方が明確になる。階層(B)は、メタデータの流通のための基盤機能であり、個別の用途に合わせて作られるメタデータスキーマや標準化されたメタデータ語彙(注:メタデータの記述項目集合や主題語彙等の統制語彙)の流通のために必須である。WWWにおけるLinked Open Dataの活動は、Resource Description Framework (RDF)を共通のメタデータ表現形式として利用し、いろいろなデータセットを結び付けようというものであり、デジタルアーカイブの相互連携にとっては重要な活動である[6][7]。また、異なるデジタルアーカイブ間でメタデータ語彙を共有することは、ある意味では、デジタルアーカイブ間でメタデータを表すためのことばを共通化する取り組みである。従来、デジタルアーカイブ間の相互運用のためにメタデータのマッピングを行うことが多かったが、もともと同一語彙でメタデータを書くようにすれば、マッピングのためのコストを低減する

ことができ、相互運用性が向上する。また、デジタルアーカイブを利用するサードパーティにとっても、異なるメタデータ間のマッピング作業を減らすことができる。

4. クラウド (Cloud) とデジタルアーカイブ

人的資源を多く持たない中小規模の MLA に対してデジタルアーカイブ構築のための敷居を低くするためには、誰もが簡単に使えるクラウド型のデジタルアーカイブ構築環境が求められる。また、東日本大震災では、津波によって多くの文化財や公文書のみならず、オフィスにおいてあったパソコンやサーバも被害を受けた。このことは、デジタル化だけでは意味がなく、より堅牢な環境にデジタルアーカイブを構築することが必要であることを意味する。堅牢な環境でのデジタルアーカイブの実現は、人的資源を持たない機関にとっては非常に困難な要求であるため、さらに敷居が高くなることになる。一方、前述の階層構造からも理解できるようにシステム基盤は共通的な要素からなる。そのため、デジタル保存や堅牢な環境作りを MLA 間の共通基盤としてクラウド(Cloud)化することが望まれる。

図1 は、こうした観点から構成したデジタルアーカイブ構築のためのクラウド環境の単純なモデルを示している。一番下の階層は長期間の保存の役割を持ち、中央の階層は収集・組織化とアクセスにかかわる機能をもつ。収集・組織化・アクセスの機能は個別のアーカイブのサービス内容に依存する。しかし、基本的な機能は共通するものが多いと考えられること、そしてアーカイブ間の連携を進めるためのメタデータの相互運用等のための共通基盤が必要であると考えられる。長期保存の階層は、バイナリデータの保存が基本機能であ

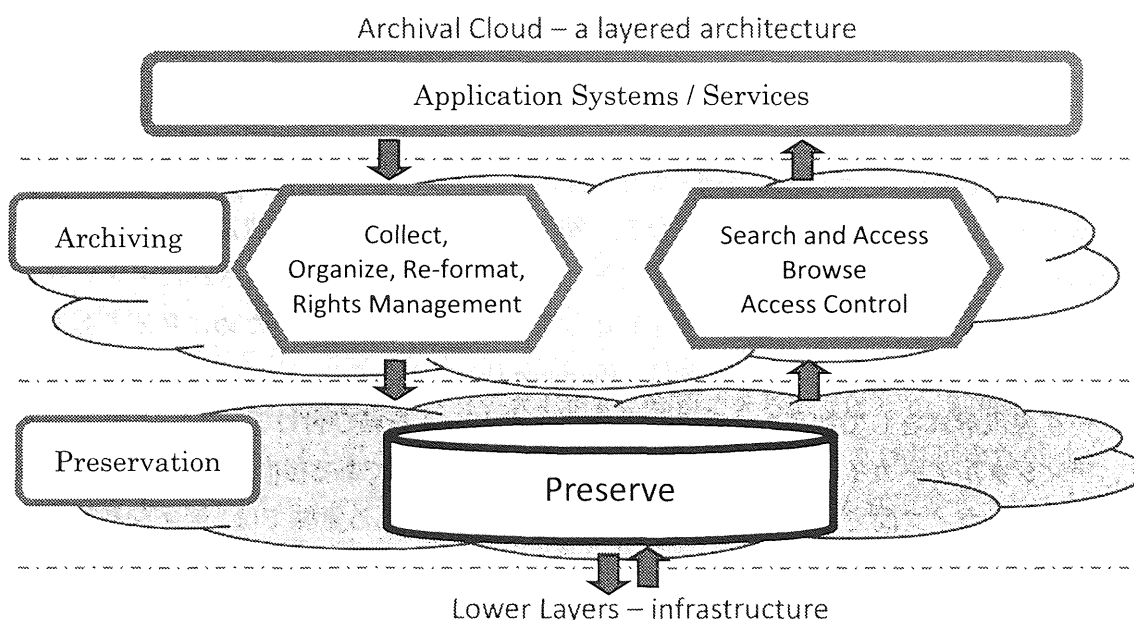


図1 デジタルアーカイブのための単純なクラウド環境のモデル

る。しかしながら、個々のデジタルアーカイブの要求に応じたデータの維持管理機能が求められる。

現在のクラウド環境の基本的な問題は、はベンダーごとの構成を持っていることと、公的な機関のリソースを蓄積、保存するに十分な安全性、信頼性を持つとまだ十分に認められていないことである。筆者等はデジタル保存のためのクラウド環境に関する分析をメタデータの視点から進め、必要とされる機能を含むモデルの提案を行った[8][9]。こうした研究はすぐに実際のシステムに結び付くというものとは言えないが、デジタルアーカイブをクラウド上に実現していくために必要なモデルに関する新しい視点を与えていると考えている。

5. デジタルアーカイブの基盤としてのメタデータ

知デジ研提言の中においても、メタデータ情報基盤の必要性をうたっている。メタデータの相互運用性の重要性は言うまでもないであろう。Dublin Core Metadata Initiative (DCMI)は、メタデータの相互運用性を高めるための基本モデルとして DCMI Application Profile (DCAP)に関する Singapore Framework を定義している[10]。Singapore Framework では個別のサービスごとの機能要求等も含まれているが、メタデータの相互運用に直接関係するメタデータスキーマの視点からは Description Set Profile (DSP)を決めている。DSP は、個別の応用ごとに作られるものであり、メタデータで用いられる記述項目等の語の集合（メタデータ語彙）と構造的制約を決めている。図 2 にメタデータスキーマ、DSP とメタデータ語彙の関係の概念図を示す。

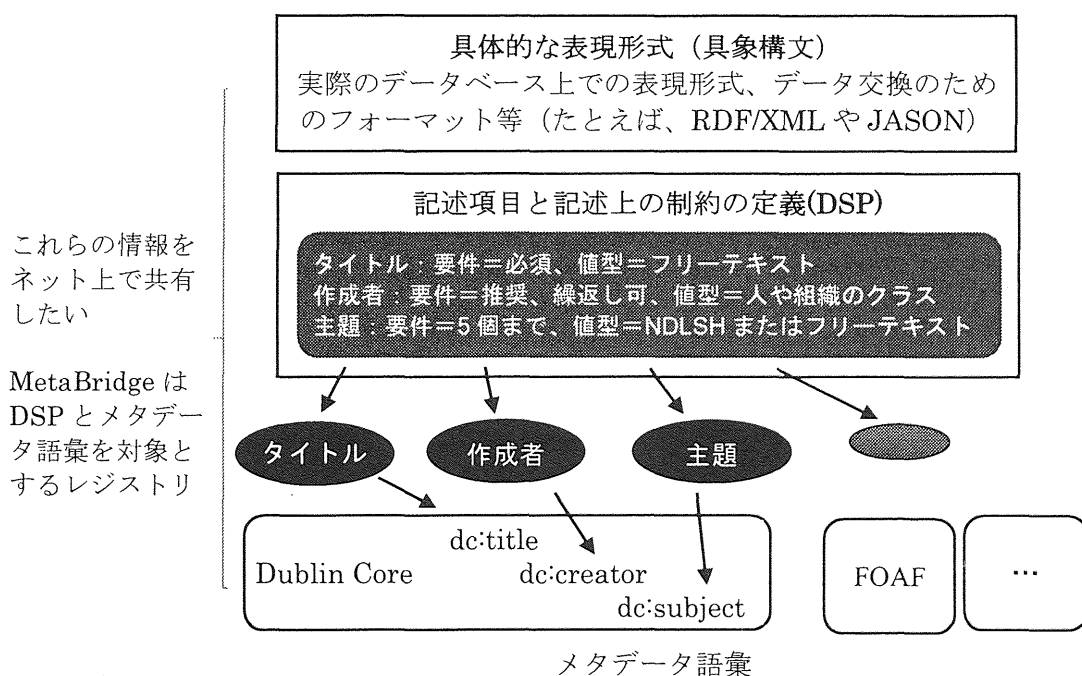


図 2 メタデータスキーマとその情報共有

DCAP にとってキーとなる点は記述項目のための語彙や記述内容として用いられる統制語彙を、応用ごとのメタデータスキーマとは切り離して定義している点である。記述項目や記述内容に用いられる語は、メタデータの意味内容に深く関連する。異なるメタデータスキーマを相互運用する際に作られる記述項目間のマッピングは、同じ意味を持つ記述項目を結び付けるものであり、いわば意味的關係づけの役割を持っている。意味的關係づけは一般に面倒である。そのため、異なるスキーマであっても、共通のメタデータ語彙を使って記述項目を決めてくれれば相互につながやすくなることは明らかである。言い換えると、DCAP とは独自の語彙はできるだけ作らず、使わず、できるだけアリモノを利用すれば、マッピングの手間を少なくし、相互運用性を高めることができるというものである。これは Resource Description Framework の考え方と合致するものでもある。したがって、メタデータに用いる記述項目や統制語彙についてはできるだけ広く使われる標準的なものから選ぶことが重要であり、たとえ独自に作ったとしても、それを他の応用でも利用できるようにネット上に公開しておくことが重要である。すなわち、メタデータを表すためのことばを共有することがメタデータの相互運用性を高め、デジタルアーカイブの相互運用性、連携性を高めるために役立つのである。

筆者等はメタデータ情報基盤として MetaBridge と呼ぶシステムを開発し、その運用を進めている[11][12]。MetaBridge は、メタデータ語彙の定義、DSP の定義を蓄積提供する機能を持っている。こうした機能を持つシステムはメタデータスキーマレジストリ (Metadata Schema Registry) と呼ばれる。(メタデータレジストリと呼ぶこともある。) MetaBridge 以外にもメタデータスキーマレジストリはあり、アメリカで運営される Open Metadata Registry[13]、筆者等が携わってきた DCMI Registry[14]は RDF をベースにしたものである。また、国立図書館等が中心に進めている件名や名前の典拠を RDF 化して提供するサービスもインターネット上でのメタデータの流通性を高めるための重要なサービスである[15][16]。このほか、先にも示したイギリス国立公文書館のファイルフォーマットに関するレジストリ[5]はデジタル保存のためのメタデータにとっては重要な情報を提供してくれる。

6. おわりにー将来に向けて

貴重図書や文化財の電子化に始まるデジタルアーカイブの開発は、国立図書館などの大組織を中心に進んできた。その一方、デジタルアーカイブの重要性が広く認められてはいるものの、実際にはなかなか手につかない MLA 機関も多くある。現在、Web 上にはさまざまな有用なコンテンツがあり、デジカメやデジタルビデオカメラなどの機材が身の回りにあふれている。そして、デジタル放送が定着し、電子書籍も広がりつつある。こうした我々の環境を考えると、物理的な資料をデジタル化して作り上げるデジタルアーカイブ (伝統的なデジタルアーカイブ) だけではなく、最初からデジタルコンテンツとして作られ、そのまま流通するコンテンツを含め、将来の世代に向けて収集、蓄積、保存するも

のも含めデジタルアーカイブの役割はますます大きくなっていくことは疑えない。

東日本大震災では多くの文化財や地域の記録文書等が被害を受けた。有形物、無形物にかかわらずデジタルコピーを作りアーカイブ化することが、貴重な資産を残していく上で有効な方法であることは疑えない。デジタルコンテンツの長期保存に関する信頼性がまだ十分に高くはないという批判もあるが、きちんと維持管理する環境を開発していくことでコンテンツの長期利用は可能になる。1000年後の世界を想像してコンテンツを保存することは難しいが、次世代にきちんと手渡せるように保存していくことは可能である。デジタルコンテンツであれ、物理的な実体であれ、きちんとメンテナンスされないものは残っていく保証はない。

国立国会図書館がホストしたデジタル情報資源ラウンドテーブルの議論からも、デジタルアーカイブは MLA 連携にとって極めて重要な要件であることは認められている [17][18]。一方、デジタルアーカイブの構築には機関間連携が重要である。特に独自に十分な人的資源を持ってない中小機関にとっては、長期間のデジタルアーカイブ運営の視点を持った機関間連携が必要である。デジタルアーカイブネットワークを実現して、人的サポートや技術情報の提供を行うことのできるコミュニティを作り上げることは重要である。知デジ研の議論においても、文化遺産オンライン等の研修機会を利用すること、総務省の地域情報化アドバイザーの制度を利用することなどが議論された。その一方、こうしたコミュニティを実際に作り上げるために発信力のあるハブを作り上げていくことが必要であろう。

知デジ研には、デジタルアーカイブに直接関連する MLA 関係者に加えて、電子出版関係や IT 関係からの委員の参加をいただいた。これは、電子書籍、電子出版物、電子公文書等が、すぐにアーカイブでデジタル保存されるべきコンテンツになること、出版者がいつまでもコンテンツを保存できる環境の実現には図書館等の公共機関、保存のための情報技術を持つコミュニティとの連携が必要であると考えたためである。絶版になった本は図書館等におかれた現物から復元できる可能性はあるが、電子書籍のファイルがなくなってしまうと復元は非常に難しくなる。そうした視点からも保存、アーカイブは重要であると思う。

デジタルコンテンツ化とは、入れ物と内容の分離である。たとえば、印刷された冊子体の小説は、小説そのものと冊子が一体化しているが、ネット配信される電子書籍の小説にとって入れ物は意味を持たず、データを再生する環境が入れ物の代わりに必要とされる。加えて、デジタルコンテンツでは 1 次コンテンツ（たとえば、小説そのもの）とメタデータと一緒に配信できることになる。同時に配信しなくてもネット上でメタデータあるいはメタメタデータにリンクすることもできる。こうした技術はすでに実現されているが、有効に利用されているかどうかはよくわからない。デジタルアーカイブはこのような柔軟性を持つデジタルコンテンツを有効に、かつ長期にわたって利用し続けるための基盤である。したがって、そこでは 1 次コンテンツのみならず、メタデータ、メタメタデータ、メタメタメタ…が有機的につながった環境にならねばならない。こうした視点からも Linked Open Data (LOD) の取り組みは、デジタルアーカイブの発展にとって重要な役割を持つものである。

はじめにも述べたイギリスの Digital Curation Centre は、Digital Curation はデジタルリソースを長期にわたって利用することにより、リソースの価値をより高めることであると述べている。アーカイブされるリソースの種類も量も多く、デジタル保存にはいろいろな難しい問題もある。たとえば、大きなサーバを動かすためのエネルギー消費が問題にされることもあるが、物体の保存の場合であっても適切に空調された建物が必要であり、環境負荷はゼロではない。メタデータが解釈不能になるとデジタルコンテンツの再生ができなくなるが、メタデータとそのメタデータスキーマの長期保存のための基盤は十分には確立されていない。このようにいろいろな問題はあるが、現代の知的資源を将来に残していく上でデジタルアーカイブは不可欠であり、デジタルアーカイブを長期にわたって利用し続けることに関する取り組みを MLA に限定せず、より広いコミュニティが連携して進めていかねばならないと思う。

謝辞

1 年余りにわたるデジタルアーカイブに関する議論を、多数の有識者と一緒に続けられたことは筆者にとってこの上ない機会であった。しかも、東日本大震災という未曾有の災害の中でコミュニティの記憶と記録を長きにわたって保存し、使っていく基盤としてのデジタルアーカイブに関する理解を深めることができた。知デジ研のすべてのメンバーと知デジ研をホストしてくださった総務省情報流通振興課の関係者の皆様にお礼を申し上げたい。また、クラウドの重要性に関して気づかせてくれた大学院生 Jan Askhoj 氏、永森光晴先生ほかの研究室メンバーに感謝したい。なお、本論文で示した研究は一部科学研究費補助金（基盤研究(A) 22240023）による

参考文献

- [1] 総務省, 知のデジタルアーカイブ～社会の知識インフラの拡充に向けて～―提言及びガイドラインの公表―, 2012.3 (アクセス: 2013.2.24)
http://www.soumu.go.jp/menu_news/s-news/01ryutsu02_02000041.html
提言 http://www.soumu.go.jp/main_content/000156248.pdf
提言概要 http://www.soumu.go.jp/main_content/000153598.pdf
- [2] 笠羽晴夫. デジタルアーカイブ, 基点・手法・課題, 水曜社, 204p, 2010
- [3] Digital Curation Centre, <http://www.dcc.ac.uk/> (アクセス: 2013.2.24)
- [4] 杉本重雄. デジタル世界における図書館とアーカイブズ, 別冊「環」15, 藤原書店, pp.104-111, 2008
- [5] The National Archives, The Technical Registry PRONOM,
<http://www.nationalarchives.gov.uk/PRONOM/> (アクセス: 2013.2.24)
- [6] Linked Data - Connect Distributed Data across the Web, <http://linkeddata.org/> (アクセス: 2013.2.24)

- [7] LODAC Linked Open Data for Academia, <http://lod.ac/about> (アクセス : 2013.2.24)
- [8] Jan Askhoj, Shigeo Sugimoto, Mitsuharu Nagamori: A Metadata Framework for Cloud-Based Digital Archives Using METS with PREMIS. Proceedings of International Conference on Asia-Pacific Digital Libraries 2011、Springer LNCS 7008, pp.118-127
- [9] Jan Askhoj, Shigeo Sugimoto, Mitsuharu Nagamori. Preserving records in the cloud, Records Management Journal, Vol. 21 Iss: 3, pp.175 – 187, 2011
- [10] Mikael Nilsson, Thomas Baker, Pete Johnston. The Singapore Framework for Dublin Core Application Profiles, 2008, <http://dublincore.org/documents/singapore-framework/> (アクセス : 2013.2.24)
- [11] メタデータ情報基盤協議会, <http://www.mi3.or.jp/> (アクセス : 2013.2.24)
(MetaBridge <http://www.metabridge.jp/infolib/metabridge/menu/>)
- [12] Mitsuharu Nagamori, Masahide Kanzaki, Naohisa Torigoshi, Shigeo Sugimoto. Meta-Bridge: A Development of Metadata Information Infrastructure in Japan, Proceedings of International Conference on Dublin Core and Metadata Applications 2011 (DC-2011), Den Haag, Netherlands, pp.63-68, 2011, <http://dcpapers.dublincore.org/pubs/article/view/3632> (アクセス : 2013.2.24)
- [13] Open Metadata Registry, Supporting Metadata Interoperability,
<http://metadataregistry.org/> (アクセス : 2013.2.24)
- [14] The Dublin Core Metadata Registry, <http://dcmi.kc.tsukuba.ac.jp/dcregistry/>
(アクセス : 2013.2.24)
- [15] Web NDL Authorities 国立国会図書館典拠データ検索・提供サービス
<http://id.ndl.go.jp/auth/ndla> (アクセス : 2013.2.24)
- [16] Library of Congress Authorities, <http://authorities.loc.gov/> (アクセス : 2013.2.24)
- [17] 国立国会図書館デジタル情報資源ラウンドテーブル
<http://www.ndl.go.jp/jp/aboutus/roundtable.html> (アクセス : 2013.2.24)
- [18] 水谷長志編著. MLA 連携の原稿・課題。将来, 勉誠出版, 296p, 2010