

集合知相互運用のための 分散デジタルアーカイブプラットフォームの標準化提案 standardization proposal of distributed digital archive platform for interoperability collective knowledge

グローマン・ヒロスケ

Hirosuke GROHMANN

Telecommuting-Lab. / Portside-Station Ltd.

要旨

本研究は、横浜市の委託研究の一環として、産学官の協力により、市民団体や公的機関がそれぞれの保有するリソース（資料や記録）を適切な形態で備蓄し（アーカイブ化）、積極的に相互運用を行い、施策策定や市民活動のための集合知として活用するための、「デジタルアーカイブ（デジタル書庫）」の運用の方法論や相互運用のプロトコルなどの標準化について調査、検討および実証実験し、提言したものである。

横浜市や各自治体での取り組み例を紹介しながら、より実効性のある集合知プラットフォームへ発展させるための試みを、運用面および技術面から紹介する。

1. はじめに

「デジタルアーカイブ (Digital Archives)」は、1994年に、「世界の文化を未来に継承するデジタルアーカイブ国際会議」が石川県で開催され、それを契機に1996年にデジタルアーカイブ推進協議会 (JDAA) が文化庁、通商産業省 (当時)、自治省 (当時) の支援を得て設立され、2005年に解散されるまで、国家支援プロジェクトとして自治体などを中心に普及の計られた概念である。その概念や実施の記録は「デジタルアーカイブ白書」として纏められている。

ただ、知材 (知的財産) 一般のアーカイブという意味では、比較的文化的遺産的価値の高いリソースに限って、そのデジタル化やインターネット、DVDメディアなどによる公開が行われるに留まり、博物館や美術館など公的資料館などに納められたごく一部のリソースがデジタル化 (電子化) されたに過ぎず、国家的プロジェクトとしての啓蒙の成果は大きかったものの、予算を使い切ったあとの継続的な事業として積極的に活動のつづけられている例は少ない。また、予算処置上止むを得ないとはいえ、電子化の成果を誇示することに傾注するあまりに、インターネットや電子メディアとしてのビジュアルな成果物のみを力点をおいてしまい、収録リソースを継続的に追加し、一般活用するための運用面に関する研究や技術的な配慮については、残念ながら、特筆する程の成果があがったとは言いがたい。ただし、デジタルアーカイブという事業に一時的とはいえ、予算がついたことで、自治体毎のデジタルアーカイブ化への潜在需要が掘り起こされ、市民団体などが中心になりデジタルアーカイブに興味の注がれる契機となったことは評価できる。

我々が横浜市に提案したのは、そういった、デジタルアーカイブ熱が消えてしまわないうちに、持続可能な市民事業として、市民参加の自治が取りざたされる中、市民の参加による市民のための民学官のデジタルアーカイブを構築運用する「プラットフォーム (独占的なアプリケーションの作成という意味ではなく、運用に関するノウハウの確立と公開、および、多くのコミュニティで利用してもらうためのフォーマットやプロトコルの策定と公開を念頭に置いている)」として定着させることであった。そして、そのことを念頭に、2006年度は横浜市の人、モノの往来や景観を題材に、運用面や技術面での調査、検討、研究、フィールドワークを行った。そして2007年度には、その成果を踏まえて、プラットフォームの運用論の確立、フォーマットやプロトコルの策定し広く公開する準備を進めるとともに、例題として実装したものをフリーウェアとして配布するための作業に入っている。その中間結果として、要点を紹介、解説する。

2. 集合知活用のプラットフォームとしてのデジタルアーカイブ

「集合知」ということで、まず最初に思い付くのは **Wikipedia** であろう。数多くのキーワードに対しての解説がネチズンとしての衆人によって行われ、その内容の信憑性についての緒論はあるものの、膨大な収録キーワード数や多言語に跨がる網羅的な知識源としての評価は定着してきている。ただ、言語を中心とする知識ベースであること、中央集権的なアーキテクチャーであることなどによる利点があると共に限界もあり、多くのメディア形態を扱うことができたり、分散指向であったりといったオルタナティブな知識ベースのプラットフォームがもっと出現してくるべきだと思われる。

また、**Google** など高性能の検索エンジン環境が整いやGIS（地理情報システム）の活用が進んだことにより、インターネットそのものを、検索エンジンや地理情報をアクセスの窓口とした大規模な知識ベースと見なすこともできなくはない状況が出現しつつあることも事実だろう。ただ、検索エンジンに収録されない公開情報が未だ圧倒的に多く、収録されたとしてもその情報があまりにも多彩で必要な情報にたどり着くための方策がまだまだ未熟なこと、そして、商業的なフィルターがかかりやすいことなどから、検索エンジンさえ使えば、ポータルにさえアクセスすれば、どんな知識でも手に入れることができると思い込んでしまうと、落とし穴に落ちることになってしまう。本論の調査でも明らかになっているが、インターネットの宇宙がいかに爆発的な膨張を続けているにせよ、そこでネットワークされている情報は人類の持つ知識のほんの一部分に過ぎないということを強く認識する必要があるだろう。インターネットのネットワークが驚異的なスピードで成長を続けている原動力は、人類の個人が持っている潜在的な知識があるからこそであるということに立ち戻って考えなければならない。

我々のアプローチは、「狭義のデジタルアーカイブ」という概念を発展させ、積極的に知識ベースとして登録、運用し、広く活用するための、集合地インフラとして、「広義のデジタルアーカイブ」という概念を作り出して「プラットフォーム」としてその運用を定着させ、インターネットによるネットワークにも積極的に接続するための方法論を確立し、その担い手を、専門家民間を問わず分野を問わずに、広く育成することである。

現存のデジタルアーカイブは、「文化財や歴史的遺産など、先人が生み出した有形無形の創造物を記録し保存活用して、新たな文化や知識の生成に役立てる」というデジタルアーカイブの定義を博物館的、図書館的に捉えており、絶えず成長し、知識のフィードバックループを作るといった柔軟な運用が行われているほとんどなく、システムもそのようには作られていない。今ある文化遺産を兎に角デジタル化することで劣化を防ぎ、博物館的展示になぞらえてDVDなどの映像メディアやインターネットでの公開を行うというどちらかという静的な「記録 (record)」という意味合いが強い。また、その公開に関しても体系的に格好よく見せることに重点が置かれている。

これを「狭義のデジタルアーカイブ」と仮定するならば、我々の考える「広義のデジタルアーカイブ」とは、博物館的図書館的な評価の確定した対象リソースだけでなく、評価の決まっていない創造物、自然の営みなど地域の記憶、博物館的価値が低くても市民にとっての意味のあるリソースをも持続的に取り込みつつ、二次創造物、三次創造物の作成を促し、それらをも取り込みつつ成長することのできる「集合知プラットフォーム」として、前向きに活用することを前提としてダイナミックな「記憶 (memory)」と定義したい。また、権威者が、その評価を行うのではなく、多くのコミュニティにより分散運用されることを前提とし、活用や交流が進む中で、評価が確定していくようなコミュニティ同士のネットワークによって強化される評価のダイナミズムを作り出すことができると考えている。

3. 広義デジタルアーカイブの運用についての考察

技術的な考察にはいる前に、デジタルアーカイブの実際の運用を想定して、システム技術という枠から離れ、本体の要求仕様がどこにあるかを明らかにしたい。

さまざまなリアルコミュニティがデジタルアーカイブを収集公開することを想定した場合、最初に認識しなければならないことは、リソースの在り所、所有者、権利者についてである。博物館や自治体が鳴り物入りで進めるプロジェクトの場合は、こういった条件がクリアされている場合が圧倒的に多く、問

題にならないことが多いのだが、小さなコミュニティが何の背景もなくリソースを収集、収録しようとした場合には、権利問題が最大の壁として立ちはだかることになる。それ故、フォーマットやプロトコルに最初から、権利の運用を円滑に進めるための仕組みを組み込んでおかなければならない。それは既存の販売のための Copyright 保護プログラムといった後ろ向きの規則ではなく、Copyleft といった新しい権利運用の潮流をも取り入れたものとする必要があるだろう。知材の既得権益を守りたい対象リソースもあれば、逆に積極的に活用流通してもらうことに意味を見いだしたい対象リソースも存在するに違いないからだ。対価が欲しいということではなく、中途半端に運用公開して欲しくないという心理のためにデジタルアーカイブへの収録を拒否する例の方が圧倒的に多いことが我々のフィールド調査でも感じられた。引用によるネットワークへの速やかな伝播、そして近年隆盛をみせている「マッシュアップ(複数の既存コンテンツを組み合わせ、新たなコンテンツを創造する手法)」を促進させるためにも、権利運用は非常に重要な意味を満っている。従って、クリエイティブコモンズⁱⁱのような既存の権利運用プログラムなどとの連携を計りながら、当初から権利記述と権利運用の組み込みを前提としてフォーマットおよびプロトコルを策定する必要がある。

次に考慮しなければならないのが、ヒューマンインターフェースである。デジタル化されたリソースを登録し公開するためには、当然、携わる人間とコンピュータ環境とのインターフェースが要求される。それが収集促進や活用促進の要ということになる。大抵のデジタルアーカイブプロジェクトでは、この部分に気のきいたビジュアルの要求を満たすことに重点を置いてしまい、特定のソフトウェア会社やデザイン会社に依頼をすることになる。そうすると、その時は満足する結果が出たように錯覚するが、実は、予算が切れたあと、流行遅れの成長することのない遺物が Web サイトや DVD 作品として残されるだけで、持続的な追加作業や活用には殆ど繋がらないことが明らかになるのが常だ。予算の豊富な公的機関のメモリアル事業として行うならそれでも良いかもしれないが、集合知として意味のある運用をするためには、ビジュアル面の整備は枝葉と考える方が正しいし、プラットフォームの提案側が行うことでもないだろう。より重点をおかなければならないのは、より多くのコミュニティや個人が参加することのできる、標準的なアーカイブの格納および交換フォーマットや運用のためのプロトコル作りのための環境づくりである。少なくとも、適切な入力、格納や検索、情報交換のためのフォーマットやプロトコルが策定されていれば、細部の実装やヒューマンインターフェースについては、各コミュニティ、運用者にとって一番親しみがあがり使い易いものを選択することが可能になる。その方が独占を防ぐことができ、予算もかからず、コミュニティの独自色を発揮するのに都合が良いことになる。市民中心の民学官ですすめるプロジェクトであれば、これは言わずもがなであろう。

運用面で考慮しなければならないもう1つの重要な点は、デジタルアーカイブ事業の担い手に対する教育と標準の適用である。図書館司書などの専門家が主導するプロジェクトであれば、分類やリソースハンドリングのリテラシーや共通認識が確立しており、その慣例や帳票類に従ったシステム設計やフォーマットを規定すれば、自ずと標準化が行われるだろう。それに対し、興味や視点そして活動の背景、メディアリテラシーの全く違う雑多な市民も参加する、分野横断的に運用される事業では、そういった定式は成り立たない。司書ならばそれほどバラツキなくできる分類作業は、さまざまな分野や仕事を背景とする人間が行った場合には絶望的なほどまちまちの分類基準が噴出することになる。だからといって、それが誤りと決めつけることはできないことに注意する必要がある。分野や目的が変われば分類基準も違い、それを共存させつつ集合知として寧ろ分野の違いをクロスさせることにこそ意義を見いだして行くというのが正しいアプローチだと言える。定型化されない知識の運用にこそ、雑多な要素を集合させるデジタルアーカイブの本領発揮の場面だと積極的に解釈することが必要になるだろう。

これらの結論に至る調査の経過やフィールドワークについては、横浜市と共同で「市民デジタルアーカイブを利用した市民参加による文化伝播と往来の研究」という報告書に纏められており無償配布を行っているので、参照されたいⁱⁱⁱ。

4. 集合知プラットフォームとしてのデジタルアーカイブ運用のための技術提案

前節の運用についての考察から、広く集合知プラットフォームとしてのデジタルアーカイブを普及させるために必要な技術要件について仕様策定を行い、広くさまざまなコミュニティで活用するための標準化の提案を行っており、その概要について下記に概説する。

詳しくは、本プロジェクトのホームページを参照するか、お問い合わせいただきたい^{iv}。

4.1 交換および格納フォーマットとインターフェースについて

あるアーカイブ対象となるリソースの所有権や運用権を持つコミュニティないし個人がデジタルアーカイブとして、インデックスならびにリソースのデジタル化されたインスタンスを保管、公開するためのフォーマットを情報の交換やロボットによる可読を前提としてXMLベースで策定した。RSSのようなシンプルなフォーマットを指向しながらも、前節の考察にあるように権利運用のための情報を積極的に取り込んである。また、ロケール（言語地域）ごとの情報を多重登録する仕組み、および、テキストデータだけではなく「非文字データ」を柔軟に登録するための仕組みを施してある。

また、実装を考慮し、フォーマットにデータを格納しないフォーマットからのデータの取り出し、フォーマットを交換するためのインターフェースを策定してある。

4.2 検索およびリスティングのためのAPIとインターフェースについて

アーカイブを集約したり検索するためのhttpベースのプロトコルを策定し、Ajaxをベースとした検索とリスティング、そして、アーカイブ提示、アーカイブ収集のためのAPIを策定してある。

また検索インデックスを運用するために、形態素抽出および完全転置インデックスを生成、連想キーワード登録を行うためのJavaインターフェースを実装を考慮して策定してある。

4.3 アーカイブフォーマット専用知識ベースエンジンについて

アーカイブデータをXMLデータのままハンドリングし、リソースとパッキングした上で、転置インデックスを使いながら、効率よく格納しとり出すことのできる専用の知識ベースエンジンの仕様を策定し、実装を考慮したJavaインターフェースを策定してある

4.4 分散環境を連携させるためのプロトコルについて

個々のコミュニティで構築するデジタルアーカイブを集合知として、このコミュニティの権利を尊重しながら外部に知識発信を行うための、インデックスやアーカイブデータの伝播プロトコルを策定し、実装のためのJavaインターフェースを策定してある。

4.5 Plug-in インターフェースによる、実装

上記に挙げた各Javaインターフェースについて、他のプラットフォームへの組み込みを行うことができるよう、Tomcatなどと連携するPlug-in モジュール仕様の策定を行っている。

5. まとめおよび今後の予定

デジタルアーカイブを戦略的に活用することにより、コミュニティおよび分野横断的な集合知プラットフォームを普及させるためのアイデアや方法論について解説させていただいた。

現在、実際にこのプラットフォームの普及を促進するために、前節で挙げたインターフェースおよびAPIの実装を行ってJavaおよびAjax周辺技術を利用して行っている。策定したフォーマットおよびプロトコル、完成したインターフェースはフリーウエアでの配布を計画しているので、ご期待いただきたい。

ⁱ 引用文献：デジタルアーカイブ白書 2001～2005（デジタルアーカイブ推進協議会発行）

ⁱⁱ 参考サイト：クリエイティブコモンズ (<http://creativecommons.org/>)

ⁱⁱⁱ 引用文献：市民デジタルアーカイブを利用した市民参加による文化伝播と往来の研究（横浜市＋市民メディア連絡会発行／横浜市中区日本大通3 4 番地中区民活動センター気付「横浜市民メディア連絡会（hara@pado.co.jp）」）

^{iv} 参考サイト：横浜デジタルアーカイブプロジェクト (<http://digital-archives.org/>, info@digital-archives.org)