

集合知を活用したデジタルアーカイブの構築とその課題

研谷紀夫

東京大学大学院情報学環

〒 東京都文京区本郷7丁目3番1号

Tel: 03-5841-3500 Fax: 03-5841-3500

E-Mail: togiya.norio@iii.u-tokyo.ac.jp

概要

文化資源統合アーカイブでは、コミュニティ機能を使用して、資料に関する議論を行い、メタデータの修正や加筆などを行える機能を実験的に実装した。本研究では、これらのコミュニティ機能を使用して集合知を集め、それらをデジタルアーカイブのコンテンツに反映させるデジタルアーカイブの運用フローとその課題について検討する。

キーワード

デジタルアーカイブ、メタデータ、オンラインコミュニティ

Architecture and subjects of Digital Archive Using Collective Intelligence

Norio Togiya

University of Tokyo

Graduate School of Interfaculty Initiative in Information Studies

7-3-1 Hongou, bunkyoku, Tokyo Japan

Phone: +81-03-5841-3500, Fax: +81-03-5841-3500,

E-Mail: togiya.norio@iii.u-tokyo.ac.jp

Abstract

The user community function in which user could discuss and correct metadata was implemented in the Integrated Digital Archive of Cultural Resources. In this study, management flow and subjects of digital archive using collective intelligence in the user community was researched.

Keywords

Digital Archibe, Metadata, Online-community

1. はじめに

歴史資料や美術・工芸品などの文化資源をデジタル化して公開するデジタルアーカイブの課題は、各資料内容を記述したメタデータの信頼性を保つことである。特に歴史学などの学術研究の使用に耐えうるデジタルアーカイブを構築する場合は、極めて信頼性の高いメタデータを保存・公開する必要がある。しかし、文字などの記述が少ない画像資料などにおいては、資料に関係する情報が不足しているために、年代や被写体の同定・検証が少数の研究者だけでは困難な場合がある。そのため、様々な人々がオンライン上でコミュニケーションをすることが可能なデジタルネットワークメディアの特性を活用し、より多くの集合知を集めることによって、資料に関する情報を集約し、メタデータを随時更新していくことが必要である。しかし、その一方でそれらの情報内容の信頼性を保つ審査システムも同時に必要となる。本研究では集合知を集めるコミュニティ機能を利用しつつ、そこにおいて発生した様々な意見をどのように審査してデジタルアーカイブに取り込んでいくべきか、運用フローについて検討する。

2. 集合知を用いるデジタルコンテンツの例

オンライン上で複数のユーザから提供される様々な集合知を集めるサイトは、オンライン百科事典として知られる Wikipedia[1] 以外にも多くの試みを見ることができる。しかし、一次資料を対象に扱い、そのメタデータの編集を多面的な観点から行うことや、それに対して様々な観点から意見や修正意見などを記述できるシステムを導入しているデジタルアーカイブなどの例は極めて少ない。資料に関するユーザの記述という意味では、イギリスの Maritime Museum [2] において登録ユーザが自分でデジタル化した資料を博物館のサイトにアップロードすることや、コメントを入力することが可能である。また、米国では、The Drexel Digital Museum Project [3] などにおいて、服飾などに関するデジタル化資料に対してコメントなどを入力する機能が付与されている (Martin[2004]) [4]。しかし、これらの事例以外で、一次資料のメタデータなどについて意見を出し合い、それらをオンライン上で複数の人間で編集するシステムを構築する例は、極めて少数である。そのため今後これらの特徴を活かしたメディアの構築が今後増加すると考えられる。

しかし、一方でこれらの集合知を用いた、デジタルアーカイブを構成していく上で最も重要なことは、各資料に関するメタデータの更新をどのようなプロセスで行っていくかという点である。Wikipedia などの百科事典の場合は、様々な事項に関して、一定程度の文章で叙述する形式を採用しているが、メタデータの場合は作品名のほか、作者、成立年代、資料の種別など予め決められた項目に関して、極めて簡潔で短い信頼性のある情報を記述する必要がある。そのため、集合知を集めた上で、それらの情報の中から必要なものを選択し、編集し、掲載する過程が必要である。これらに関しては、非公開の作業によって編集され、かつ公開後も利用者からの修正や付加情報を受け付けることを前提としていない刊本の目録や図録とは異なる情報の編集・更新フローが必要である。そのため本研究では、コミュニティを用いて集合知を集約した上で、それらの内容をメタデータに反映させる具体的な運用フローを検討し、その内容について実証実験を行う。

3. 文化資源統合アーカイブでの実証

3.1 文化資源統合アーカイブの概要

前述のようなオンラインコミュニティを用いてメタデータの更新を行う機能を、東京大学大学院情報学環の文化資源統合アーカイブに実装し、その実証実験を行った。文化資源統合アーカイブは、弥生式土器の発見者の一人であり、日本の人類学・考古学の礎を築いた坪井正五郎資料、日本の営業写真師の草分け

である上野彦馬撮影の歴史写真資料、江戸時代の地理学者長久保赤水製作の古地図である赤水図、戦前から戦後にかけて活躍した岩石学者坪井誠太郎の邸宅に関する建築資料の主に四資料を統合的に格納したデジタルアーカイブである [5]。

文化資源統合アーカイブ全体に関しては拙論において既にその概要が解説されている（研谷 他 [2006]） [6]。また、本アーカイブで活用されている、オントロジの構造とその機能に関する内容についても拙論（研谷 他 [2007]） [7] において概説され、またコミュニティのユーザ側から見た機能とユーザによる評価概要に関する筆者による研究発表においてその内容が解説されている（研谷 [2007]） [8]。本研究ではさらに、コミュニティ機能を用いて、資料に関する集合知を集め、実際のメタデータを変更する運用フローについてその概要と課題について考察する。

3.2 コミュニティを利用したメタデータの更新フロー

前述した文化資源統合アーカイブのメタデータに関しては、大学院の博士課程を中心としたリサーチアシスタントによって、資料整理時にその内容情報が記述された。しかし、必ずしもその資料内容を完全に把握できなかったため、一部の資料に関しては、その記述が不足している点や、資料の成立年代が不確定のものなどが存在した。そのため、資料に関するより詳しい情報追加や、より正確な情報を記述するため、オンラインコミュニティを使用して、幅の広い有識者の意見を収集反映させることとした。

メタデータを更新する内部の体制としては「段階 1」と「段階 2」までを想定し準備を行った。「段階 1」は、内部の単数の管理人がメタデータの修正を行う権限をもち、コミュニティにおける指摘に対応し、メタデータを修正するフローである。また、「段階 2」では、専門的な知識を必要とする内容が発生するため、各資料の内容に精通する内部の担当者を決め、管理者と担当者との協議においてそれらの情報内容の修正を行う運用フローである。これら、2つの運用フローを適宜利用しながら、コミュニティにおけるユーザからの意見に対応することとした。

3.3 実証実験の結果

前述の更新フローに基づいて文化資源統合アーカイブの実証実験を、2006年10月3日から20日までの17日間行った。実証実験は50名余りが参加し、資料の閲覧や、オンラインコミュニティにおいての実証実験が行われた。コミュニティの指摘によるメタデータの修正に関しては、実証実験開始後、「談話室」においては、実証期間中合計で353件、1日平均で20件程度のコメントがあり、運営者側が予想したより多くのコメントがあった。発言の内容としては、(1)メタデータや事物・概念の記述に関するコメント、(2)資料に関する批評、(3)アーカイブに対する意見・感想など主に3つの内容のコメントが寄せられた。特に(1)のメタデータや事物・概念に関しては表1にあるように、誤字などの単純な誤りに対する指摘と、年代や被写体に関する記述の誤りなど内容面の誤りを指摘するものの主に2種類にわけることができる。前者に関しては、誤字などの指摘があり、実際3件の誤字の修正と4件のメタデータと写真の不一致が指摘され、後日修正された。これらの修正に関しては「段階 1」によって対応することが可能であった。

一方後者に関しては主に歴史写真資料に関して、表1にあるような内容が指摘された。これらの内容は専門的な内容に関する指摘であったため、一部「段階 2」に移行し、内部の研究者の協力を得て確認をとった。その結果、「No.9」「No.10」に関しては、コミュニティによる指摘の信憑性が高いとされ、メタデータの変更、もしくは併記などが望ましいことが確認された。一方で、「No.11」から「No.15」に関しては、指摘が事実である可能性が高いものの、内部の担当者だけではその可否を判定できないと判断した。そのため、これらの項目に関しては、引き続き外部の専門家をあわせた検討が必要であり、「段階 2」に加えて外部の協力者を加えた検討調査を行う、新たな「段階 3」を策定することが必要となった。特に歴史写真

などの資料に関しては、製作者を特定することが難しく、写真だけではなく、様々な傍証資料を参照した上で、確認する必要がある。そのため、これらの資料に関してはすでに内部だけではなく、外部のより多くの関係者による審査や調査作業が必要であり、これら「段階3」を設けることとした。但し本実証実験の期間中でこれらのフローの構築を行うことは難しいため、これらの「段階3」を設けることは今後の課題となった。

4. まとめ

本実証実験で示したとおり、デジタルアーカイブのような様々な資料を集め、それらのメタデータに関する内容を、コミュニティなどを用いて更新する場合は、凡そその半分近くに関しては内部の担当者によって処理することが可能な指摘・コメントであった。しかし残りの半数近くの指摘・コメントに関しては、外部のユーザより想定範囲を超えた量と質のコメントが寄せられるため、メタデータに反映する判断を内部の人間だけでは行えない状況となった。そのため、オンラインコミュニティで様々なユーザの集合知を用いて情報の更新を行う場合は、内部の管理者や資料に関する知識を持った担当者だけではなく、より幅広い知識を持った外部の専門家などを含めた調査委員を当初より調査及び審査委員として配置する必要がある。これらの専門家を含んだ複数の担当者による審査が行われることによって、より信頼性の高いメタデータの記述が可能になると考えられる。

しかし、これらの人員を十分に配置することは人的リソースが限られているため容易ではない。そのためコミュニティの情報を資料のメタデータに反映させる段階に至るまでは、より多くの時間を要することとなる。そのため、コミュニティなどにおける様々なユーザ意見についてはそれらを、全面的に閲覧可能な情報とし、ユーザの責任においてそれらの情報の信憑性を判断し活用する一方で、より多くの時間と調査をかけた上で、メタデータ上には正式に反映させるという時間的、手続き的な段階を分けることが必要であると考えられる。

このような点からコミュニティにおいて集合知を集める機能を備えたデジタルアーカイブのメタデータ更新フローは、図1のような形態が現状では最も現実的な形式であると考えられる。図1で示したフローにおいては、コミュニティにおける情報はメタデータ内容とは一線を画し、ユーザの責任において、その内容を判断、利用する情報と位置づけられる。そしてそれらの情報の中でより正確性が高く、メタデータにも反映すべき情報があった場合は、内部の人間だけではなく外部の専門家を調査員もしくは審査員として、メタデータ更新の検討に対して協力を得ることが望ましい。そしてその後のフェーズにおいては、編集長にあたる内部の担当者によって責任を明確にした上で、追加・更新などが決定される必要がある。その場合も、誰のどのような意見によってどのような調査が行われ、どのような文献資料、および各種一次資料によって該当メタデータの更新が行われたかを示すことが必要である。この点はオンラインの百科事典のように匿名の複数のボランティアなどによる緩やかな審査よりも、より更新・編集に明確な根拠と責任を負わせたフローとなっている。このように編集責任を明確にし、誰がどのような理由によって、メタデータが変更・更新されたかを明確にすることによって、各メタデータが学术论文や学術文献に引用される信用を獲得する情報になると考えられる。

以上のようにコミュニティなどの活用をし、様々な知識情報を取得することと、学術研究基盤としての信頼性を保った情報を格納する点の二つを両立しようとする場合は、コミュニティにおいてはユーザによる自由な情報の書き込みを許容する一方で、それらの情報を資料のメタデータに反映する段階で、外部の専門家を含めた調査・審査委員の協力を仰ぎ、情報内容を調査し、編集責任と情報更新の理由を明確にした上で、メタデータを編集して行くフローを確立することが必要である。このことによってオンラインコミュニティを実装したデジタルアーカイブが学術研究に耐える研究基盤ツールとなることが可能となるが、このような運用フローを構築することは様々な人的リソースを必要とすることとなり、その構築には

一定程度の時間を要する。そのため、これらの運用フローが定着し、実際に間断なく運用されていくためには、集合知を用いながら、運営されるデジタルアーカイブの理念と内部運用フローをよりオープン化し、学術コミュニティや一般社会における理解を十分に深化させていくことが必要である。

謝辞

本研究は、21世紀COE次世代ユビキタス情報社会基盤プログラムのプロジェクト（リーダー：坂村健教授）として行われ、馬場章教授をはじめ、プロジェクトメンバーである倉持基、大島十二愛、山下大輔、松田好史、山根信二、添野勉の各氏の協力を頂きました。また資料の提供には小沢健志、上野一郎、関川敦之、坪井正道、坪井直道、長久保甫、三吉治敬、山本泰則、横山功、各氏と、学校法人

産業能率大学、国立民族学博物館、高萩市歴史民俗資料館、長久保赤水顕彰会、港区立港郷土資料館、(有)渡辺出版、ニューカラー印刷(株)のご協力を頂きました。さらにデジタルアーカイブの構築に際しては津田光弘氏、(株)堀内カラーの協力を頂きました。関係の皆様には厚く御礼を申し上げます。

[1]Wikipedia:<http://ja.wikipedia.org/wiki/> (2007.10)

[2]National Maritime Museum :<http://www.nmm.ac.uk/> (2007.10)

[3]The Drexel Digital Museum Project <http://digimuse.cis.drexel.edu/> (2007.10)

[4] Kathi Martin, The Role of Standards in Creating Community, Proceedings of the 13th international World Wide Web conference, pp35-41,2004

[5] 文化資源統合アーカイブ：<http://cr-arch.chi.iii.u-tokyo.ac.jp/> (2007.10)

[6] 研谷紀夫, 馬場章 他 ”オントロジとコミュニティを用いた統合型デジタルアーカイブの構築の場合” 情報処理学会シンポジウム じんもんこん 2007 論文集, 情報処理学会, pp57-62、2006

[7] 研谷紀夫, 馬場章 ”文化資源オントロジの構築とその活用” 情報知識学会誌, 17 巻 2 号, pp129-134,2007

[8] 研谷紀夫 ”デジタルアーカイブにおけるオンラインコミュニティの活用とその課題” 情報文化学会講演予稿集 15, pp115-118,2007

表1：コミュニティの指摘によって行われたメタデータおよびその他の機能の修正・更新事項

NO	修正の修理	内容	対応段階
NO1	誤記訂正	「事物・概念」の上野陽一郎の誤記を上野陽一に修正	段階1
NO2	誤記訂正	「メタデータ」の「坪井博士追悼記念1」の画像とメタデータが間違えていたため修正	段階1
NO3	誤記訂正	「メタデータ」の「坪井博士追悼記念2」の画像とメタデータが間違えていたため修正	段階1
NO4	誤記訂正	「メタデータ」の「坪井博士追悼記念3」の画像とメタデータが間違えていたため修正	段階1
NO5	誤記訂正	「事物・概念」の「二条基弘」の記述が途中で切れていたのを修正	段階1
NO6	誤記訂正	資料「兎と亀」の画像の形式が違うため表示できなかったのを修正	段階1
NO7	加筆	「事物・概念ツリー」の柴田常恵の経歴に加筆	段階1
NO8	機能改善	談話室にリンク機能を付与	段階1
NO9	内部で検討 情報併記	資料「奥州鎮撫振遠隊」の記述「上野撮影局前で撮ったと思われる」は正確には「上野邸の向かいの高木邸の前」	段階2
NO10	内部で検討 情報併記	資料「和装の男女」年代が明治ではなく「慶応年間」	段階2
NO11	外部の研究者を含めた調査が必要	資料「大浦天主堂」上野彦馬撮影ではない可能性がある	段階3
NO12	外部の研究者を含めた調査が必要	資料「諏訪神社の大鳥居」上野彦馬撮影ではない可能性がある	段階3
NO13	外部の研究者を含めた調査が必要	資料「雲仙小地獄」上野彦馬撮影ではない可能性がある	段階3
NO14	外部の研究者を含めた調査が必要	資料「高鈴島」上野彦馬撮影ではない可能性がある	段階3
NO15	外部の研究者を含めた調査が必要	資料「中島川と編笠橋」上野彦馬撮影ではない可能性がある。	段階3

図1：コミュニティを用いたデジタルアーカイブのメタデータ更新フロー

