



コミュニティアーカイブとしての東日本大震災アーカイブーオープンデータ連携による利用性の向上

Digital Archives of Great East Japan Earthquake as a Community Archive - Enhancing Usability through the Linking of Archives in Open Data Environments

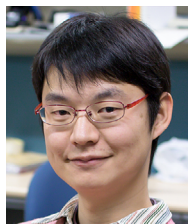
杉本 重雄
SUGIMOTO Shigeo

筑波大学
図書館情報メディア系



三原 鉄也
MIHARA Tetsuya

筑波大学
図書館情報メディア系



永森 光晴
NAGAMORI Mitsuharu

筑波大学
図書館情報メディア系



抄録：東日本大震災に関する記録を集めたデジタルアーカイブが多く作られ、国立国会図書館はそれらのポータルとして東日本大震災アーカイブひなぎくを提供している。現在、こうしたデジタルアーカイブの利活用性の向上が求められている。本稿では、筆者等がこれまでの研究から得た知見に基づきメタデータの視点から東日本大震災アーカイブにおける利活用性の課題に触れ、さらに、オープンデータ環境を利用したアーカイブ内ならびにアーカイブ間でのコンテンツ集約による利活用性の向上について述べる。最後に、アーカイブのコンテンツを結ぶことによる利活用性向上について考察する。

Abstract: There exist many digital archives for disasters caused by Great East Japan Earthquake, and Japan's National Diet Library (NDL) provides a central portal for these archives named Hinagiku. The aim of this paper is to discuss issues regarding the improved usability of the digital archives. This paper first discusses usability issues based on the authors' experiences in their research projects mainly from the viewpoint of metadata in Open Data Environments. Then, it presents a few experimental results from the aggregation of archived contents within an archive as well as across archives. Lastly, it discusses the importance of the linking and aggregation of archived contents in order to improve the usability of the digital archives of disaster resources.

キーワード：震災記録デジタルアーカイブ、Linked Open Data、アーカイブ間連携、メタデータ集約、地域指向アーカイブ、コミュニティの記憶

Keywords: Digital Archives of Disaster Records, Linked Open Data, Collaboration across Archives, Metadata Aggregation, Regional Archives, Community Memory

1. はじめに

東日本大震災の発災後、震災記録資料を集めたデジタルアーカイブ（以下、震災アーカイブと記す）が多数作られた。これらの提供組織は、地方自治体、大学、報道機関、情報産業等の企業、そして地域のNPO等さまざまである。また、震災アーカイブには、発災から復旧、復興過程の写真や文書等の原資料を収集提供するもの、それらの上に何らかのテーマでまとめたコンテンツを提供するもの、被災地の震災以前の写真等を収集提供するものなど、様々なものがある。国立国会図書館（NDL）による東日本大震災アーカイブポータル「ひなぎく」によりそれらの横断的な検索が可能になっている^[1]。ひなぎくからは、震災アーカイブ以外に、震災関連資料として震災に関連する記録や論文等を含むデータベース等も参照されている。震災につ

いて知るうえでこれらはいずれも重要な資料である。

個々の震災アーカイブには、記録資料の探しやすさと使いやすさといった課題、そしてアーカイブ間にまたがった利用による付加価値といった課題がある。前者の解決は、個別のアーカイブの機能の向上によるところが大きく、後者の解決には震災アーカイブ間の連携だけでなく、それらの枠組みを超えていろいろなアーカイブや他の資料をWeb上で結びつける機能が必要とされる。

筆者等は、震災アーカイブに関するこうした課題に関してメタデータの視点から研究活動を進めてきた。この研究過程において、青森震災アーカイブ（青森県八戸市、三沢市、おいらせ町、階上町）、久慈野田譜代震災アーカイブ（岩手県久慈市、野田村、普代村）、みちのく震録伝（東北大学災害科学国際研究所）、郡

山震災アーカイブ（福島県郡山市）を中心にメタデータの分析等を行った。本稿では、そうした活動から得た知見を基礎として、メタデータと Linked Open Data (LOD) の視点から、震災アーカイブの利用性を向上するための課題について述べる。

2. デジタルアーカイブとメタデータ

2.1 メタデータに関するいくつかの基本概念

ここではメタデータに関してあまりなじみがない読者のために、いくつかの基本概念についてごく簡単に説明する。

メタデータの基本的な定義は「データに関するデータ」であり、「何らかの対象に関する記述」である。震災アーカイブのメタデータと言った場合には、震災アーカイブに収集蓄積された記録資料に関する記述（たとえば、記録資料の目録データ）である。メタデータ記述の際に、主題や分類等を表すためにあらかじめ定められた語彙（統制語彙、Controlled Vocabulary）を用いることがある。これらも何らかの概念に関する記述であり、広い意味でのメタデータである。

メタデータは、ごく簡単には、記述対象の属性を表す語と属性値の対の集合として実現される。属性は Attribute、Property あるいは Element（記述項目）といった語で示される。Web 上でのメタデータの流通性・相互運用性のために目的に応じたメタデータ標準が作られている。そうした標準の中で、ネット上の情報資源（リソース）に関する基本的な属性集合を決める Dublin Core、人や組織等を表すための FOAF（Friend of a Friend）のように広い領域で用いられるものもある。NDL は Dublin Core を基礎として DC-NDL を定めている。ひなぎくでは、震災アーカイブ開発のために定められたガイドライン^[2]に基づき、DC-NDL を基礎として定めたメタデータ記述項目を利用している^[3]。一方、World Wide Web コンソーシアムは、Web 上でのメタデータ交換のために実体関連モデル（Entity-Relationship model）に基づくデータモデルとそのテキスト表現形式の標準 Resource Description Framework (RDF) を定め、そのテキスト表現のために XML 形式、JSON 他の形式を定めている。また、ひなぎくでは、OAI-PMH（Open Archives Initiative – Protocol for Metadata Harvesting）と呼ばれるメタデータ収集の標準を利用して、多数の震災アーカイブからメタデータを収集し、アーカイブ横断的な検索機能を実現している。

LOD は、Web 上で公開されたデータをリンクづけることで、データの価値を高める取り組みとして広く知られている。データ同士を結びつけるには、属性と

属性値の対応関係を陽に定められるメタデータが重要な役割を果たす。震災アーカイブの場合、ひなぎくのメタデータ記述項目の範囲でのメタデータの横断的な利用は行いやすい。一方、属性値がプレーンテキストで記述される場合が多く、属性値による結び付けにはそれなりの処理が求められる。また、アーカイブ資料と Wikipedia の記事や Web ページを結びつける場合には、異なる規則の下に作られたメタデータ間でのマッピングが必要になる。

2.2 東日本大震災アーカイブとそのメタデータ

個々の震災アーカイブには写真を中心とする多様な資料が蓄積されている。震災アーカイブ開発のガイドラインでは、単純な構造のメタデータによるアーカイブ構築を前提としている。震災アーカイブ構築においては、大量の資料に対し、厳しい時間的かつ予算的制約の中でメタデータを作成することが求められた。こうした環境下で作られたメタデータの記述品質は必ずしも高いとはいえない。たとえば、写真の内容を表すタイトルをつけることは第三者にとっては容易ではない。また、1 点の資料ごとにメタデータを付与することを原則とするので、1 枚の写真、1 件の文書毎にメタデータが付与される。写真の場合、1 か所で撮影された連続した写真であっても、それらの関連性は必ずしも陽に記述されていない。写真のメタデータが緯度経度、撮影時刻等の時空間データをそれなりの精度で持っている場合は写真間の関連性が推定可能である。しかしながら、そうした時空間データがメタデータに含まれているとは限らない。このように、いろいろな問題があることが容易に理解できる。

デジタルアーカイブとしての利用性を高めるには、高品質なメタデータを用意することが重要であるが、コストを考えると現実には容易ではない。また、画像解析や認識技術、テキスト分析技術等の高度化によって原資料から機械的にメタデータを作り出すことが期待されるが、写真が撮られた場所や撮影目的といった資料のコンテキストに依存する情報や資料作成の意図といった意味的情報を扱うための技術の蓄積は十分とは言えず、そうした課題を解決する技術の研究開発が求められる。

2.3 コミュニティの記憶としてのアーカイブ

震災アーカイブでは、地理的、時間的、内容的に資料収集の範囲が自ずと定められる。その一方、地域コミュニティは震災以前からあり、震災後も将来に向かって存在し続ける。コミュニティのアーカイブという視

点から見た場合、震災アーカイブは、東日本大震災という歴史的な大災害によって定まる時間的、地理的、内容的な窓を介してコミュニティを記録するものであるといえる。「未来へのキオク」^[4]は、震災以前の地域の風景等を提供してくれている。津波にかかわる過去の報告書や学術的な記録を収集する「津波デジタルライブラリ」^[5]には、三陸地域での津波関係資料も含まれている。こうしたいろいろなアーカイブ資料や地域コミュニティにかかわる Web ページ等を、震災アーカイブと結び付けることができれば、時間的、地理的により広い範囲でコミュニティの記録を見ることができるようになり、コミュニティの記憶と呼ぶことのできる意味的な資料の結びつけを得ることができる。

2.4 コミュニティアーカイブの基盤としてのオープンデータ環境

どのようなアーカイブであれ、それ一つだけでコミュニティの全ての記録をカバーすることはできない。また、アーカイブ毎にコンテンツ収集や組織化の方針は異なり、そして利用者のニーズはさまざまである。そのため、アーカイブをオープンなものとし、複数のアーカイブを結んでニーズに応じたいろいろな機能を実現できるようにすることが求められる。

LOD 環境に適合する形式でアーカイブや Web ページ、並びにそれらのメタデータが提供されることが望まれる。ひなぎくに結ばれるアーカイブでは、メタデータが標準化され、LOD 環境に適合するように提供されているため、データの形式上の横断的利用は行いやすくなっている。しかしながら、記述内容に基づく意味的な結び付けにはそれなりの処理が必要であり、解決しなければならない課題が多くある。

3. 震災アーカイブの利用性向上を目指して

3.1 ひなぎくと震災アーカイブのメタデータに関して

本稿を書くにあたって、ひなぎくで「多賀城市 八幡神社 被害」をキーワードとして検索してみた。結果として、2011 年 3 月から 2015 年 6 月までの資料 87 件（内、写真 85 件、文書資料 2 件）を得た。その中には、多賀城市の広報公聴係による 2011 年 4 月 20 日作成の写真が 7 件含まれていた。この内の 1 件からのリンクをたどり、写真の提供元である「たがじょう見聞憶」にアクセスすると、これらは一連の写真として掲載されており、ひとまとまりのものであることがわかった。一方、この構造をひなぎくの検索インタフェースだけから知ることは難しく、撮影時刻と場所からこれらが一連のものであることを想像するしか

ない。また、検索結果には、複数のアーカイブに収集された 2011 年から 2015 年までに作られた資料が含まれていることから、多賀城市の八幡神社とその近辺の被災から復興に至る過程を見ることができるのではないかとことが想像できる。実際に内容をチェックすると、被災直後の写真、復旧作業にかかわる神社周辺の木の伐採作業とその後の神社の修理の写真、そして、鎮守の森復活への植樹に関する新聞記事までが検索結果に含まれており、時系列として並べれば八幡神社とその周りで起きたことを知ることができる。こうした点は、複数のアーカイブを横断的に利用できることの利点である。その一方、資料 1 点毎のメタデータの内容を理解した上で資料間をつないで見ることをしなければ、こうしたつながりを見つけることはむづかしい。

3.2 アーカイブの利活用性向上を目指したコンテンツ集約

3.2.1 単一アーカイブの中でのコンテンツ集約

デジタルカメラがどこにでもあるという時代になり、非常に多数の写真が記録資料として得られるようになった。このことはアーカイブする記録資料が豊富になるという良い面を持つ一方、多数の写真をどう組織化するかといった問題の原因となる。先に述べたように、限られた人的、予算的資源の中でのメタデータ作成においては、関連する写真同士の関連性や、写真の撮影理由や目的を反映するようなタイトルをつけることは現実性に乏しい。その結果、検索結果表示が単なる写真の羅列となってしまう、使いやすさの低下の原因になると考えられる。

この課題に対し、メタデータを利用して関連する資料をひとまとめにすること、すなわち複数の資料をひとまとまりの資料として集約することが一つの解決方法として考えられる。

3.2.2 アーカイブ間にまたがるコンテンツ集約

Europeana では、多数の参加館から収集したイメージデータ等を集約するためのデータモデルを定めている^[6]。ジャパンサーチ構想でもメタデータの集約が重要なトピックになっている^[7]。ひなぎくは多くの震災アーカイブからメタデータを収集し、横断的利用を可能にしている。こうした複数のアーカイブからメタデータを集めるための基盤の上で、さらに内容に応じた集約が求められる。

コミュニティの記録は複数のアーカイブや Web ページに残されていると考えるのが自然であり、それらをつなぐことでコミュニティの記録を有機的につな

げ、コミュニティの記憶を残すことにつながる。一方、アーカイブ間にまたがるコンテンツ集約には、越えねばならない壁がいくつもある。たとえば、タイトルやキーワードの作り方はアーカイブ毎にポリシーが異なると考える必要がある。また、地名のようにアーカイブによる差が少ないように思えるものであっても、記述の粒度の違いや地域独特の呼び方といった問題がある。異種アーカイブとの連携には、解決すべき問題がさらに多くある。たとえば、明治期や昭和初期の三陸大津波の記録と震災アーカイブを結ぶには、市町村合併等による地名の変更、時代や地域による表現の違いなどいろいろな課題があり、そうした課題を解決するための地名辞書等のデータ資源を別途必要とする。

3.3 時空間情報、主題情報による集約

筆者等は、青森震災アーカイブ、久慈野田譜代震災アーカイブを中心にメタデータに書かれたキーワードの分析を行った。その結果から、主題語の出現頻度、写真が占める割合の高さとメタデータ記述品質等に関する基礎的知見からメタデータによるコンテンツ集約の必要性を認識し、下記の視点からの研究を進めた。

- (1) キーワードの共起関係を利用した資料の集約作成^[8]
- (2) 時間情報、地理情報（緯度経度）を利用した写真を中心とする資料の集約作成^[9]

(1) に関しては、出現頻度がメタデータから取り出したキーワード（主題語）をクラスタリングし、キーワードクラスタに含まれる語を持つメタデータの集合を、各クラスタに対応するメタデータ集約とした。(2) に関しては、緯度経度と作成日時、および作成者情報を利用し、同一作成者が連続して撮影したと判定できる写真の集合をひとつの集約とした。被災地調査やお祭りなどのイベントで連続して撮影された写真をひとまとめにするのには効果的な方法であるといえる。こうした手法によって作成した集約を1件のコンテンツとするために集約に対して適切なメタデータを機械的に生成すること、集約として適切なサイズ（集約中に含まれる個別資料の数）をどのようにとらえるべきかといった課題が残されている。

3.4 LOD環境でのアーカイブ間連携のための基盤

アーカイブ内もしくはアーカイブ間で資料を意味的関係の下に結びつけることを目的として、地域指向のオントロジーと市町村等の地名の変遷を表す辞書の開発を進めた。

前者では、記録資料のタイトルや内容記述から固有名詞を集め、それを施設や組織の分類と組み合わせて

オントロジーを試作した^[10]。図1に一例を示す。東日本大震災アーカイブには津波被害関連資料が多く蓄積されていることから、港湾や漁業、水産関連の施設や組織の名前がメタデータに現れることが多い。一方、その施設や組織がどういう役割のものであるかは多くの場合メタデータだけではわからない。このオントロジーに含まれる組織や施設の種類を表す用語は地域に限定されるものではなく、アーカイブをまたぐ意味的なメタデータ集約には有用であると考えている。

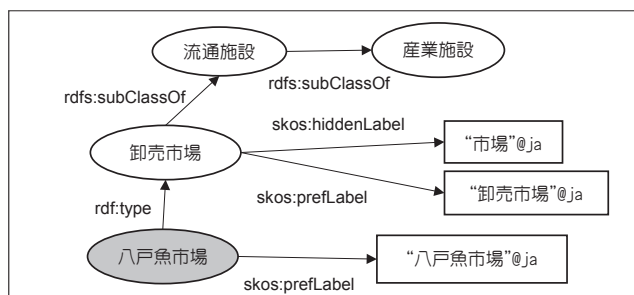


図1 地域オントロジー記述の一例

「どのような災害が同じ場所で過去にあったのか」、「災害の前はどのような場所であったのか」といったことを知りたいといった利用ニーズを満たすには、種々のアーカイブをまたいで資料を探す必要がある。先述のように、そこでは地名が重要な役割を果たす。しかしながら、長い時間がたつ間に地名は変化する。近代以降に限定しても、市町村名が自治体の合併等によって変化することはまれではない。そこで、筆者等は、LOD環境で利用可能な地名変遷辞書の開発を進めた^[11]。表1に地名変遷の種類の一覧を示す。

表1 地名の変遷種別

種別	概要
新設合併	複数の地名が統合され、新たな地名がそのいずれも引き継がない場合
編入合併	複数の地名が統合され、新たな地名がそのいずれかを引き継ぐ場合
分割	一つの地名を廃止し、複数の新たな地名のいずれもその名前を引き継がない場合
分立	一つの地名を廃止し、複数の新たな地名のいずれかがその名前を引き継ぐ場合
改称	一つの地名を廃止し、一つの異なる新たな地名をつける場合
町制施行	①村が町に移行する場合
市制施行	②町が市に移行する場合
区の設置	政令指定都市における区の新規設置
郡の設置	都道府県における郡の新規設置
郡の所属変更	ある町村の属する郡の変更

4. 「つなぐ」ことの大切さ

どのようなアーカイブであれ、地域コミュニティにおきた災害の全てを網羅的に収集することは不可能である。加えて、メタデータの記述は与えられた人的・予算的資源の範囲内で、ガイドラインに従って進めることになるので、自ずと記述内容には制約が生じる。ひ

なぎくが持つ網羅的な横断検索はカバーする範囲を広げてくれる半面、より深い内容や特定のトピックでの探索には別の視点を必要とする。そうした別の視点を与えてくれるものに、Wikipedia やボランティアによる地域情報のサイト、地域の自治体サイトとそのアーカイブ、さらには津波デジタルライブラリのような学術的サイトがある。こうしたサイトと震災アーカイブをつないで利用できるようにすることが重要である。また、先に示した多賀城市の八幡神社の例のように、複数の震災アーカイブを横断的に利用することで単独のアーカイブからは得られない情報を得ることができる。こうした機能を積極的にサポートすることも震災アーカイブをより深く利用するための重要な視点であろう。

個別のアーカイブは個別の目的のもとに作られる。他方、利用者は多様な目的でアーカイブを利用するので、個別のアーカイブで利用者の情報ニーズを十分に満たせないことが生じるのはごく自然なことである。利用者ニーズを満たすには複数のアーカイブや種々の情報資源を結ぶ付加価値サービスの実現が重要である。LOD 技術は、そうした付加価値サービスの実現のための有効な手段であると思われる。ジャパンサーチをはじめとして LOD 技術を基盤とするプロジェクトが進められているのは、そうした期待の表れであろう。一方、LOD 技術を利用して、有機的にアーカイブを結ぶには、アーカイブのメタデータに関する情報の共有はもとより、先に示した地名の変遷を表す辞書のようにアーカイブに特化した用語の意味やそれらのつながりを表すデータ資源が必要とされる。こうした資源は意味的な記述が求められるので、その作成には現時点ではどうしても人的努力が必要となるが、将来的にはそうしたデータ資源作成の自動化が進んでいくのであろう。

5. おわりに

本稿では、震災アーカイブの利活用性を高める観点から、アーカイブ内、アーカイブ間でのメタデータの集約と、それに基づくアーカイブ連携について触れた。メタデータ集約によるアーカイブ連携は、災害アーカイブだけではなく、文化的資源のアーカイブでも必要である。それらに共通することは、何らかのつながりを持つコンテンツを見出し、それらを結びつけて利用者に提供することであり、個別のコンテンツだけでは見逃しがちな内容を分かりやすく提供すること、そして、アーカイブを利用して新たに創られるコンテンツも含め利用者により豊富な内容を提供することである。

と考えている。

東日本大震災アーカイブには大量のデジタル写真が蓄積されており、これは阪神淡路大震災等の以前の災害アーカイブとの大きな違いであると感じる。また、ツイートや道路情報等、インターネットから得られた情報が災害の状況を知り復旧に役立てられたことも記憶には新しく、そうしたものも含めてデジタル情報をいかにアーカイブし、将来に渡る利用を可能にするかといったことの検討を進めなければならないと感じる。なお、本稿では、コンテンツに関する権利管理やデジタルアーカイブそのものの長期利用性に関しては述べていない。デジタルアーカイブの発展には、こうした観点も含めた研究開発を続けることが必要である。

謝辞

震災アーカイブに関する研究活動に貢献してくれた本研究室の学生諸君に感謝の意を表します。なお、ここで述べた研究は一部科研費基盤研究 A (16H01754) による。

(参考文献)

(web参照日は全て2018年6月18日です)

- [1] 国立国会図書館. 国立国会図書館東日本大震災アーカイブ ひなぎく. <https://kn.ndl.go.jp/>
- [2] 総務省. “震災関連デジタルアーカイブ構築・運用のためのガイドライン (2013年3月)”. http://www.soumu.go.jp/menu_seisaku/ictseisaku/ictriyou/02ryutsu02_03000114.html
- [3] 国立国会図書館. 国立国会図書館東日本大震災アーカイブ メタデータスキーマ. <https://kn.ndl.go.jp/static/metadata>
- [4] Google. 未来へのキオク. <https://www.miraikioku.com/>
- [5] 津波デジタルライブラリ作成委員会. 津波デジタルライブラリ. <http://tsunami-dl.jp/>
- [6] Isaac, Antione. “Europeana Data Model Primer”. 2013. https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf
- [7] 実務者検討委員会 (内閣府知的財産戦略推進事務局). “第一次中間取りまとめ”. https://www.kantei.go.jp/jp/singi/titeki2/digitalarchive_suisiniinkai/jitumusya/2017/torimatome.pdf
- [8] 積 佑典. “リソース集約を用いた東日本大震災デジタルアーカイブの利活用性向上に関する研究”. 筑波大学図書館情報メディア研究科修士論文, 2018, 46p, 修士論文.
- [9] 横山雄哉ほか. “シンプルなメタデータが付与された東日本大震災アーカイブの写真資料のための時空間情報を利用したコンテンツ集約手法”. 情報処理学会第79回全国大会, 1ZF-02, 2017.
- [10] 武田侑季ほか. “東日本大震災アーカイブのメタデータ集約を指向したオントロジーの開発”. 研究報告人文科学とコンピュータ (CH), 2018, vol. 2018-CH-116, no. 9, pp.1-7.
- [11] 松井慧ほか. “震災関連資料のリンクを目的とした地理的名称トレースのためのLODデータセットの開発”. 研究報告人文科学とコンピュータ (CH), 2018, 2018-CH-116, 11, pp.1-6.

