

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

2018年3月

積 佑典

## 目次

1. はじめに .....	1
2. デジタルアーカイブの利活用性向上.....	4
2.1. 東日本大震災デジタルアーカイブ .....	4
2.2. デジタルアーカイブにおけるリソース集約.....	7
2.3. 東日本大震災デジタルアーカイブにおける利活用性の課題.....	8
2.4. 東日本大震災デジタルアーカイブにおけるリソース集約.....	10
3. 関連研究 .....	11
4. 研究対象アーカイブのメタデータ分析.....	12
4.1. 研究対象アーカイブ .....	12
4.2. メタデータ分析 .....	14
4.2.1. タイトルの分析 .....	14
4.2.2. 主題情報の分析 .....	15
4.2.3. 時空間情報の分析 .....	17
5. リソース集約.....	18
5.1. 時空間情報と作成者情報を用いた集約.....	18
5.2. 主題情報を用いた集約 .....	23
5.3. オントロジーを用いた集約 .....	27
5.4. 複数手法の併用 .....	29
5.4.1. 時空間情報と主題情報の併用 .....	29
5.4.2. 主題情報とオントロジーの併用 .....	29
6. リソース集約システム .....	31
7. 考察 .....	32
8. おわりに .....	33
謝辞 .....	34
文献リスト .....	35
付録 .....	37

## 図表一覧

図 2-1 震災アーカイブのインタフェース例（青森震災アーカイブ <sup>[14]</sup> ）	4
図 2-2 Europeana における集約リソースの例（[20]より引用）	7
図 2-3 震災アーカイブのリソースとメタデータ例	9
図 2-4 リソース集約による地域の歴史の発見	9
図 2-5 複数手法を用いた階層的集約のイメージ	11
図 4-1 震災アーカイブごとの収集リソースの地理的範囲の違い	13
図 4-2 類似するタイトルの例	14
図 5-1 時空間情報と作成者情報を用いた集約の例 1（一部）	19
図 5-2 時空間情報を用いて集約されたリソース群の例 2（一部）	21
図 5-3 主題語クラスタの例	24
図 5-4 主題語を用いた集約の例	24
図 5-5 サイズが大きい集約の主題語クラスタの例	26
図 6-1 リソース集約システムを用いた集約の閲覧（オントロジーを用いた集約例）	31
表 2-1 「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（2016年10月版）」 <sup>[15]</sup>	6
表 4-1 研究対象アーカイブの運営団体とリソース数	12
表 4-2 アーカイブ別リソース種別	13
表 4-3 主題語の語数の比較	15
表 4-4 河北新報の頻出語（上位10）	16
表 4-5 各震災アーカイブの頻出語（上位10）	16
表 5-1 写真リソース数と地理情報の入力状況	18
表 5-2 時空間情報を用いた集約における集約対象リソース数と集約数	19
表 5-3 集約されたリソースのメタデータ例 1（一部）	20
表 5-4 時空間情報を用いて作成した集約数とサイズ別集約数	20
表 5-5 集約されたリソースのメタデータ例 2（一部）	22
表 5-6 主題情報を用いた集約における集約対象リソース数と集約数	23
表 5-7 主題情報を用いて作成した集約の数とサイズ1の集約数	24
表 5-8 主題情報を用いて作成した集約の数と孤立したリソース数	25
表 5-9 オントロジーを用いた集約における集約対象リソース数と集約数	28
表 5-10 オントロジーとマッチした固有名詞とリソース数（青森震災アーカイブを除く）	28
表 5-11 主題情報を用いた集約における集約対象リソース数と集約数	29
表 5-12 固有名詞とリソースとのリンキング	30

# 1. はじめに

後世に残し伝えるべき資料を長期に渡り保存・共有するため、資料をデジタル化し蓄積したデジタルアーカイブが多数構築・公開されている。1990年代初期より、米国議会図書館によるデジタル図書館プロジェクト American Memory<sup>[1]</sup>は、歴史的貴重書をデジタル化し提供している。日本においては、明治以降に刊行された図書・雑誌のデジタル化リソースを公開した国立国会図書館の近代デジタルライブラリー（後に国立国会図書館デジタルコレクション<sup>[2]</sup>に統合された）、国内の多数の博物館や美術館の文化遺産のデジタル化リソースを公開する文化庁の文化遺産オンライン<sup>[3]</sup>、日本やアジア近隣国に関する歴史資料を公開する国立公文書館のアジア歴史資料センター<sup>[4]</sup>が挙げられる。

近年、多数の機関・分野のデータセットを横断検索可能とするポータルサイトの構築が行われてきている。ヨーロッパ地域の文化財・歴史資料を検索できる Europeana<sup>[5]</sup>、アメリカの文化財・歴史資料を検索できる Digital Public Library of America<sup>[6]</sup>が挙げられる。日本においても国立国会図書館サーチ<sup>[7]</sup>が構築され、先述の国立国会図書館デジタルコレクション、文化遺産オンライン、アジア歴史資料データベースを含む約 100 のデータセットから資料を横断検索することができる。国立国会図書館は、さらに多くの分野・機関との連携を進め、2020年までに国の分野横断統合ポータル「ジャパンサーチ（仮称）」を構築する方針である<sup>[9]</sup>。

2011年の東日本大震災発生以降、東日本大震災に関するデジタルアーカイブ（以下、「震災アーカイブ」と呼ぶ）が、震災の記憶を後世に残し防災・減災、復旧・復興事業に役立てる目的で多数構築された。震災アーカイブに関しても、約 60 のデータセットから横断的にリソースを検索することができるポータルサイト「国立国会図書館東日本大震災アーカイブ ひなぎく」<sup>[10]</sup>が国立国会図書館によって公開されている。しかし、類似のリソースが検索結果として大量に並び一覧性が低いといった問題がある。また、大量に存在するリソースに対して個別にしかアクセスできず、一般利用者にとって東日本大震災の記憶と結びつきにくいものとなっている。

本研究室では、関連する複数のリソースの集合のメタデータを一つの集約リソースとして記述し、検索対象とすることで、リソースの一覧性やリソース間の関連の発見性を向上できると考え、研究を行ってきた。

リソース集約という手法は、デジタルアーカイブにおいて先行事例が存在する。Europeana では、特定の文化財・芸術品に関する複数の写真や文書といったリソースを集約リソースとして記述し、検索対象とできる Europeana Data Model(EDM)<sup>[11]</sup>を採用している。ただし、文化財系アーカイブにおけるリソース集約と震災アーカイブにおけるリソース集約とは、アーカイブの対象物の性質とメタデータ作成過程が異なる。

文化財系アーカイブでは、実世界で容易に識別可能な有形・無形の事物を対象としている。例えば、「日光東照宮」という有形物の写真や、「阿波踊り」という無形のイベントに関する

写真やチラシといった物である。一方、震災アーカイブでは、実世界上で識別が困難な事物を対象とした資料を多く扱っており、何らかの風景を撮影した写真リソースが数多く作成されている。例えば、宮城県石巻市門脇町4丁目9で2013年8月23日に撮影された、建物や道路が写っている写真が、何を記述したリソースなのかを定義することは難しい。そのため、作成した集約リソースに対して、リソースの記述対象を機械的に明確に定義することが困難となる。

Europeana では、予め集約モデルを定義し、それに合わせてトップダウンでメタデータを作成している。震災アーカイブでは、リソースの記述対象が曖昧なリソースに対して、各機関が異なるガイドラインによってメタデータ作成を行っており、機関によって記述内容の粒度や記述項目の用法が異なる。本研究は、このように既に作成されたアーカイブリソースに対して、アーカイブごとにボトムアップにリソース集約を行う点において、Europeana と異なる。

著者ら(2015年)<sup>[12]</sup>は、震災アーカイブのリソース作成傾向を把握するためのメタデータ分析と試験的集約実験を行った。同研究室の横山ら(2016年)<sup>[13]</sup>は、リソースの作成状況に関するメタデータである時空間情報と作成者情報を用いて集約を行なった。本研究室でのこれら二つの研究を通し、我々は特定の手法では集約することができないリソースが数多く発生してしまう問題を課題として残した。また中山(2016年)は、語の意味に基づくリソース集約手段として、ヒューリスティックによる震災アーカイブオントロジーの構築を目指した。

本論文では、引き続き震災アーカイブオントロジーの構築を行うとともに、時空間情報と作成者以外のメタデータによるリソース集約手法の適用実験を行う。また、単体手法では集約できないリソースを集約するため、複数手法を階層的に適用し、利活用性の高いデータセットの構築を行う。

2章では、デジタルアーカイブの利活用性向上のためのアプローチの一つであるリソース集約について述べる。そのために、震災アーカイブはどのようなものであるかを紹介し、リソースに付与されるメタデータ、利活用性の問題をあげる。そして、リソース集約の先行事例と本研究との相違、および本研究における研究手順を述べる。

3章では、デジタルアーカイブにおける利活用性向上のための研究や事例を紹介する。オントロジーを利用したデジタルアーカイブの利活用性向上のための研究、リソース集約以外の震災アーカイブの利活用に関する研究を挙げる。

リソース集約を行うにあたり、震災アーカイブのリソースにはどのようなメタデータが付与されているか、また、どのような手法で集約を行えるかを検討するためのメタデータ分析をする必要がある。4章では、研究対象のアーカイブそれぞれについて分析した内容を述べる。メタデータ分析に基づき5章では、複数のリソース集約手法でデータセットの構築を行う。

集約した結果としてどのようなリソースのまとまりができているのかを確認し、一般ユー

ザに対して、どれほど利活用性が向上するかを評価する必要がある。そのために 6 章では、5 章で構築したデータセットに対してメタデータを付与し、集約単位でのアクセスを可能とする。また、集約に実際にアクセスするためのインタフェースを構築する。

7 章では、構築したデータセット間の比較とデータセットの利活用像について考察する。

## 2. デジタルアーカイブの利活用性向上

### 2.1. 東日本大震災デジタルアーカイブ

2011年3月11日に発生した東日本大震災の記憶を記録として残し、将来起こり得る災害の防災・減災、復旧・復興に役立てることを目的として、多数の東日本大震災デジタルアーカイブ（以下「震災アーカイブ」と呼ぶ）が構築・公開された。震災アーカイブを運営するのは、行政機関、企業、大学など多種にわたる。

震災アーカイブの概要を、青森県八戸市、三沢町、おいらせ町、階上町が運営する「青森震災アーカイブ」<sup>[14]</sup>を例に紹介する。震災アーカイブごとに異なった特徴があるが、ここでは概ね共通する特徴をあげる。

多くの震災アーカイブでは、ホームページに検索ボックスが配置され（図 2-1）、検索テキストを用いた全文検索が用いられていることが多い。アーカイブによっては、アーカイブ構築者が設定した分類に基づいた絞り込みや、地図を用いた作成場所の絞り込み、タイムライン状のグラフを用いたリソース作成時間の絞り込みといった機能を備えたものもある。



図 2-1 震災アーカイブのインタフェース例（青森震災アーカイブ<sup>[14]</sup>）

検索を実行すると、検索結果が表示される。検索結果画面では、ヒットしたリソースの一覧が表示されるが、サムネイルを主体としたタイル状での表示や、サムネイルに加えていくつかのメタデータをリスト状にした表示方法がある。ヒットしたリソースの一つをクリックすると、そのリソースの詳細画面へ遷移し、リソースの内容や詳細なメタデータの閲覧ができる。多くの震災アーカイブでは、共通するメタデータ項目として、タイトル、作成者、作成時間・時間、主題語（あるいはキーワード）、権利情報が記述されている。

国立国会図書館は 2013 年、複数の震災アーカイブを横断的に検索できるポータルサイト「国立国会図書館東日本大震災アーカイブ ひなぎく」<sup>[10]</sup>（以下「ひなぎく」）を公開した。ひなぎくでは、約 60 のアーカイブやデータセットからメタデータを収集しており、約 360 万件のリソースを検索することができる。収集したメタデータを統一のメタデータ形式「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」<sup>[15]</sup>（以下「NDL 大震災メタデータスキーマ」）にマッピングすることによって、各機関の震災アーカイブのメタデータを統一的に検索可能にしている。

総務省は 2013 年、震災アーカイブ構築者に対して「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」<sup>[16]</sup>を示した。このガイドラインでは、アーカイブ間の連携を考慮し、既存のメタデータスキーマを使用することを推奨しており、ひなぎくが使用している NDL 大震災メタデータスキーマを参考にすることで、ひなぎくと連携した際の検索性が向上すると述べている。

NDL 大震災メタデータスキーマは、Web 上での高い共有性を実現する Linked Data 技術に準拠している。Linked Data 技術は Tim Berners-Lee(2006 年)<sup>[17]</sup>によって提案された、文書間のみならずデータ間の関係を辿れるようにする技術である。Linked Data の基本 4 原則を以下に示す。

1. ものごとを URI で名前付けする
2. これらの名前を調べて見つける（参照解決する）ことができるように、  
http:スキームの URI を使う
3. 名前付けした URI がたどられたら、有用な情報を返す
4. ほかの URI へのリンクを加えて、より多くのものごとを見出せるようにする

#### Linked Data の 4 原則（和訳：神崎）（<sup>[18]</sup>より引用）

NDL 大震災メタデータスキーマはこの 4 原則に準拠したスキーマとなっており、データ同士を繋いでいくことを想定した設計となっている。

主なメタデータ項目の定義の例を表 2-1 に示す。表中の「dcterms:subject」のような、「:」（コロン）で結合された文字列は URI の省略形であり、例えば「dcterms:subject」は「<http://purl.org/dc/terms/subject>」を意味する。記述例には、青森震災アーカイブのリソース<sup>[19]</sup>を用いた。

このスキーマで注意すべき点は、「作成者」と「寄与者」、「作成日時」と「リソースが対象とする時間」、「撮影場所」と「リソースが対象とする場所」の 3 組は、入力する値の型が同じであるということである。3 組のそれぞれは、異なる意味を表す項目である。例えば、「発災当時の体験談」という文書リソースがあり、その文書が作成されたのが 2013 年 2 月

28日であっても、リソースが対象とする時間は「発災時」となる。作成日時は、文字通りリソースが作成された日時を表すのに対し、リソースが対象とする時間は、必ずしもリソースが作成された日時とは限らない。

表 2-1 「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 (2016年10月版)」 [15]

内容	タグ名	記述例
タイトル	dcterms:title, dc:title	"H23年度 生活・活動復興支援助成金受付集計表"
シリーズタイトル	dcndl:seriesTitle	"平成23年度生活経済活動復興支援助成金①"
作成者	dc:creator, dcterms:creator	"おいらせ町"
寄与者	dcterms:contributor	"おいらせ町"
主題	dcterms:subject	"復興", "支援", "助成金", "集計表"
要約・抄録	dcterms:abstract	
リソースが対象とする時間	dcterms:temporal	"2011年7月～2012年3月"
作成日時	dcterms:created	(記述なし)
リソースが対象とする場所	dcterms:spatial	ラベル: "青森県上北郡おいらせ町中下田135-2" 緯度: 40.599087 経度: 141.397772
撮影場所	ma:locationLatitude, ma:locationLongitude	(記述なし)
資料種別	dcndl:materialType	"文書データ"

NDL 大震災メタデータスキーマによって記述された震災記録のリソースのメタデータ例を付録2に添付する。

## 2.2. デジタルアーカイブにおけるリソース集約

メタデータを多数の機関から収集し連携している Europeana では、リソース集約モデルである Europeana Data Model(EDM)<sup>[11]</sup>を採用している。デジタルアーカイブにおいて、ある事物に関する写真や文書といったデジタルリソースは複数作成される場合がある。EDM では、文化財や歴史資料といった事物ごとに、その事物を記述対象とするリソース群を一つの集約リソースとしてまとめて記述する。このモデルでメタデータを記述することで、ユーザは事物単位でリソースにアクセスすることが可能となる。これにより、全体としてのリソース数が減ることで一覧性が向上し、検索結果を俯瞰できるようになる。また、一つの事物が複数のリソースによって表現されるため、現実の文化財や歴史資料の理解も容易となる。

図 2-2 に Europeana の集約リソースの例を示す。これは、ある彫刻作品を表すリソースであり、9 枚の様々な方向から撮影された写真によってこの彫刻作品を説明している。仮にこれら 9 枚の写真が一つの集約リソースとしてまとめられていなければ、個別の写真単位でしかこの彫刻作品の情報を得ることができず、集約リソースと比較してその説明力は損なわれてしまう。

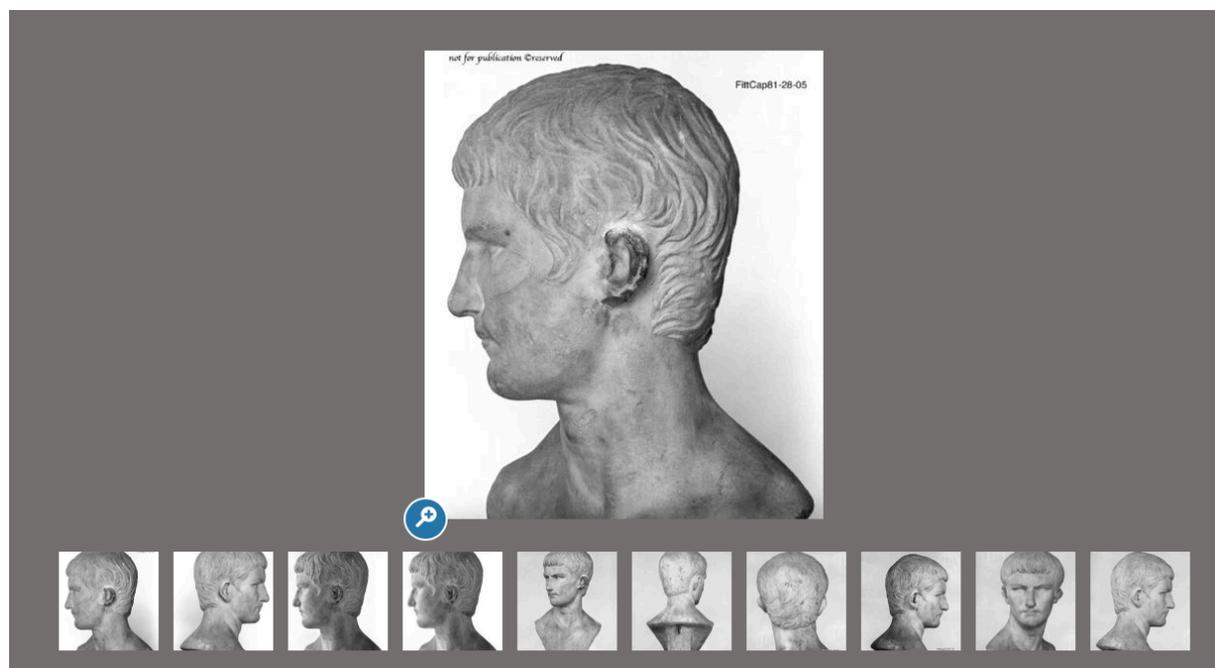


図 2-2 Europeana における集約リソースの例 ([20]より引用)

Europeana では、テキスト検索やタグによる絞り込みの他に、データリポジトリに対して直接クエリを送信しメタデータを取得することができる。2 件以上のリソースの集合から成る集約リソースを取得するクエリを付録 1 に添付する。

## 2.3. 東日本大震災デジタルアーカイブにおける利活用性の

### 課題

震災アーカイブの利活用性の課題を2点挙げる。1点目は、蓄積されたリソースを一覧的に閲覧することが難しく、外観できないという問題。2点目は、個別のリソースでは利活用の価値の判断が難しいという問題である。これらの問題について、震災アーカイブの検索結果や、震災アーカイブ利活用のワークショップの事例（佐藤, 2017）<sup>[21]</sup>を用いて紹介する。

#### (1) リソース群の一覧性が低い

震災アーカイブ「たがじょう見聞憶」を運営する宮城県多賀城市の教育委員会は、同アーカイブに蓄積された写真・地図・映像といったリソースを利用して、多賀城市防災教育副読本資料集「命をまもり 未来をひらく」を作成した。本資料には、担当教員がアーカイブから検索して選定したリソースを使用する予定であったところ、目的のリソースの検索が困難であることから、「アーカイブを所管している担当課の職員に「こういう写真が欲しい」というリクエストを投げかけて、「おすすめ」の写真を提示してもらおうという流れで、写真の選定が進んでいった」（佐藤, p.693）という。震災アーカイブから要求に合致するリソースを検索するには、どのようなリソースが蓄積され、どのようなメタデータが与えられているかを熟知した人間でないと難しいというのが現象である。

実際に震災アーカイブでリソースを検索した際、震災に関する写真や文書といったデジタルリソースは、複数同時期にまとまって作成される。例えば、復旧工事の施工記録写真や、移動しながら被害状況を連続に撮影した写真、一連の復興事業の行政文書といったものが挙げられる。このようなリソース群は、作成時のまとまりのまま、提供されるべきである。しかし、震災アーカイブでは全てが同列に扱われ、列挙されてしまっている。

#### (2) 個別リソースでの利用価値の把握の困難さ

また、佐藤は、ワークショップに使用するリソースを選択する際、『「（アーカイブされている）写真を見るだけ」では、決して行うことはできず、個々の写真について東松島市図書館の担当者からのレクチャーを踏まえ』（佐藤, p.693）た上で行なったという。このことから、震災アーカイブの個々のリソースを閲覧するのみでは利用価値を把握することが困難であると考えられる。

この問題の原因の一つとして、震災アーカイブで収集されているリソースの傾向として、単一のリソースが表現できる情報が少ないことがあげられる。例えば図 2-3 に示すリソース<sup>[22]</sup>では、主題語「切り株」「八幡神社」と作成場所から、宮城県多賀城市宮内にある八幡神社にある切り株を撮影した写真であることがわかる。また「津波被害」という主題語があることから、津波による被害の様子を撮影した写真であることがわかる。つま

り、この場所が津波被害を受けたことしか読み取れないのである。

一方、図 2-3 のリソースに登場する八幡神社のリソースを 4 つ収集し、主題情報を抜粋したものを図 2-4 に示す。この 4 つのリソースを閲覧したところ、「八幡神社に存在した鎮守の森は津波に耐えたものの、伐採されて切り株となった。その後、ボランティアの手で植樹を行い、鎮守の森の再生を図った」という地域の被災と復興に関するストーリーを読み取ることができた。このストーリーを発見できたことによって、図 2-3 の写真は、「切り株を撮影した写真」から「津波に耐えた森を伐採した後の切り株」という特別な意味を持つようになる。

現状の震災アーカイブでは、検索にヒットしたリソースが単純に並ぶだけとなっており、関連するリソースでまとめられていない。そのため、リソースの周辺情報を読み取ることができず、利活用性が低くなっている。



タイトル：みちのく・いまをつたえ隊(07 利府・多賀城)  
 主題：八幡神社, 切り株,  
 多賀城市宮内地区, 津波被害  
 作成日時：2013/09/10 14:55:29  
 作成場所：宮城県多賀城市宮内 1 丁目 2

図 2-3 震災アーカイブのリソースとメタデータ例

 <p>2013/09/10 津波被害, 切り株</p>	 <p>2011/10/25 松林, 伐採</p>
 <p>2013/09/10 津波被害, 切り株</p>	<p>2015/06/23 鎮守の森, 植樹祭, ボランティア</p> <p>(文書のためサムネイルなし)</p>

図 2-4 リソース集約による地域の歴史の発見

## 2.4. 東日本大震災デジタルアーカイブにおけるリソース集約

文化財系アーカイブにおけるリソース集約と震災アーカイブにおけるリソース集約とは、アーカイブの対象物の性質とメタデータ作成過程が異なる。

文化財系アーカイブでは、実世界で容易に識別可能な有形・無形の事物を対象としている。例えば、「日光東照宮」という有形物の写真や、「阿波踊り」という無形のイベントに関する写真やチラシといった物である。一方、震災アーカイブでは、実世界上で識別が困難な事物を対象とした資料を多く扱っており、何らかの風景を撮影した写真リソースが数多く作成されている。例えば、宮城県石巻市門脇町4丁目9で2013年8月23日に撮影された、建物や道路が写っている写真が、何を記述したリソースなのかを定義することは難しい。そのため、作成した集約リソースに対して、リソースの記述対象を機械的に明確に定義することが困難となる。

Europeana では、予め集約モデルを定義し、それに合わせてトップダウンでメタデータを作成している。震災アーカイブでは、リソースの記述対象が曖昧なリソースに対して、各機関が異なるガイドラインによってメタデータ作成を行っており、機関によって記述内容の粒度や記述項目の用法が異なる。本研究は、このように既に作成されたアーカイブリソースに対して、アーカイブごとにボトムアップにリソース集約を行う点において、Europeana と異なる。

横山ら（2017年）<sup>[13]</sup>は、写真リソースを研究対象とし、リソースの時間と地理座標が共に近いものを集め、作者が一致するリソース群を集約としてデータセットを構築した。しかし、得られた集約のうち、集約できないリソースが多数発生した。横山らは、このようなリソースを他のリソースと関連づけることを課題として残した。

本論文では、横山らと異なる手法でリソース集約をし、データセットの構築することを目的とする。4章で、研究対象とする震災アーカイブのメタデータについて分析を行い、それを踏まえ、5章でリソース集約を行う。横山らが行なった時空間情報を用いたリソース集約手法に加え、本研究では主題情報を統計的にクラスタリングしてリソースを集約する手法、オントロジーを構築して意味に基づいてリソースを集約する手法を用いる。

また、横山らが指摘した、関連づけられず孤立したリソースを他のリソースと関連づけるため、時空間情報を用いた集約手法、主題情報を用いた集約手法、オントロジーを用いた集約手法を段階的に適用し、孤立するリソースを少なくする（図 2-5）。

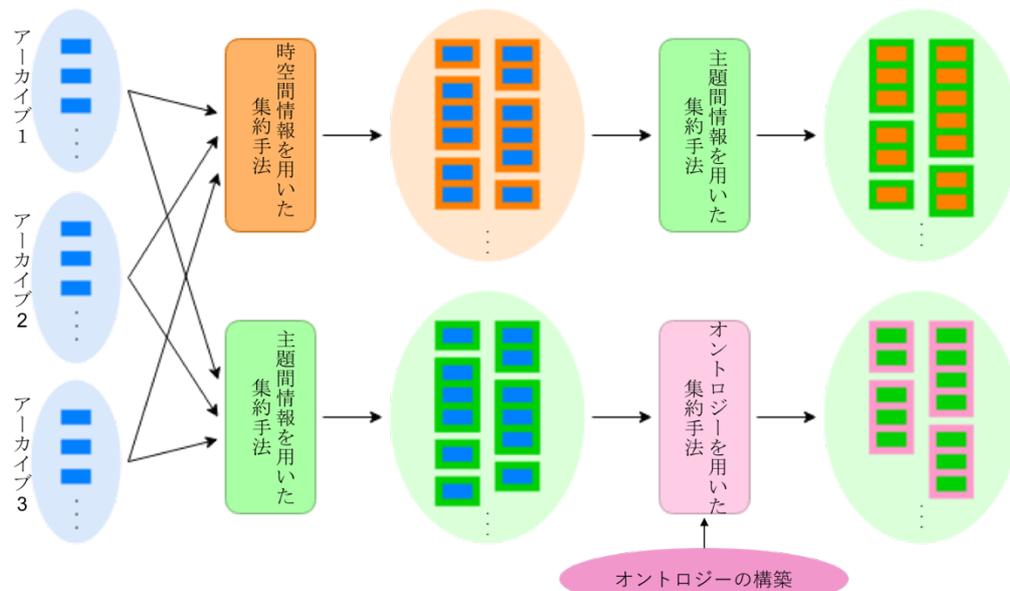


図 2-5 複数手法を用いた階層的集約のイメージ

### 3. 関連研究

関連するリソースをまとめる機能を有した震災アーカイブがある。ハーバード大学ライシヤワー日本研究所による「Japan Digital Archive」では、Web リソースをアーカイブしたものを検索できるだけでなく、ユーザは検索で得たリソースを選択し「コレクション」へ格納することができる。このコレクションは「コレクション検索」によって検索、閲覧することができる。Japan Digital Archive では、機械的なリソースの集約は行なっておらず、ユーザが個別に手作業で集約を行なっている。本研究では、ひなぎくと連携する5アーカイブを対象とし、関連するリソースの集約を機械的に行う。

研山ら(2006)は、歴史資料アーカイブの利活用性向上のため、「歴史資料に関連する、人・組織・時代・場所・物などの事物の関連性」をオントロジーとして構築した。事物の上位下位概念をツリー構造でたどりリソースを閲覧できるようにし、閲覧システムを用いてユーザ評価を行なった。特定の概念に結びつくリソースが大量に存在したことから、一覧性が低くなったとし、インターフェースの改善を対策としてあげている。本研究では、構築したオントロジーの概念に対して、個別のリソースだけでなく集約単位でのリンクングを行う。これにより、関連する大量のリソースをリンクングするとともに、インターフェースではなくメタデータの構造によって一覧性を高くする。

## 4. 研究対象アーカイブのメタデータ分析

### 4.1. 研究対象アーカイブ

本研究では、表 4-1 に示す 5 つの震災アーカイブを対象とする。

青森震災アーカイブは、青森県八戸市、三沢町、おいらせ町、階上町が運営する震災アーカイブである。震災発災後、最も早く構築されたアーカイブの一つであり、総務省は震災アーカイブの構築ガイドラインを示す上でモデルとして扱っている。自治体が収集するリソースは、復旧事業や復興イベントをくまなく記録した写真や震災関連の行政文書を多く含んでいる。

久慈・野田・普代震災アーカイブは、岩手県久慈市、野田村、普代村が運営する震災アーカイブである。青森震災アーカイブと久慈・野田・普代震災アーカイブは、システムを受託開発した事業者が共通しており、メタデータの記述方法が類似している。開発事業者の担当者は、青森震災アーカイブの経験が活かされ、メタデータ付与を改善したと語っている。

みちのく震録伝は、東北大学災害科学国際研究所が開発した震災アーカイブである。みちのく震録伝では、同研究所の研究者や同研究所の震災アーカイブプロジェクト「みちのく・いまを伝え隊」のスタッフが記録した写真が多くを占める。

河北新報震災アーカイブは、報道機関である河北新報社が運営する震災アーカイブである。システムの構築は東北大学災害科学国際研究所が行なっている。このシステムはみちのく震録伝の後に構築されており、東北大学災害科学国際研究所は、河北新報震災アーカイブはみちのく震録伝の経験を活かし、改良されたアーカイブになっているという。

郡山震災アーカイブは、福島県郡山市が運営するアーカイブである。青森震災アーカイブや久慈・野田・普代震災アーカイブの運営する自治体が海岸に面しているのに対し、本アーカイブは内陸に位置している。また福島県には、福島第一原子力発電所の事故に関するリソースが多く含まれる点でも青森震災アーカイブと久慈・野田・普代震災アーカイブと異なる。

表 4-1 研究対象アーカイブの運営団体とリソース数

アーカイブ名	運営団体	団体種別	資料数
青森震災アーカイブ	青森県八戸市, 三沢町, おいらせ町, 階上町	自治体	68,297
久慈・野田・普代震災アーカイブ	岩手県久慈市, 野田村, 普代村	自治体	127,500
みちのく震録伝	東北大学災害科学国際研究所	研究機関	125,414
河北新報震災アーカイブ	河北新報社	報道機関	115,867
郡山震災アーカイブ	福島県郡山市	自治体	49,605

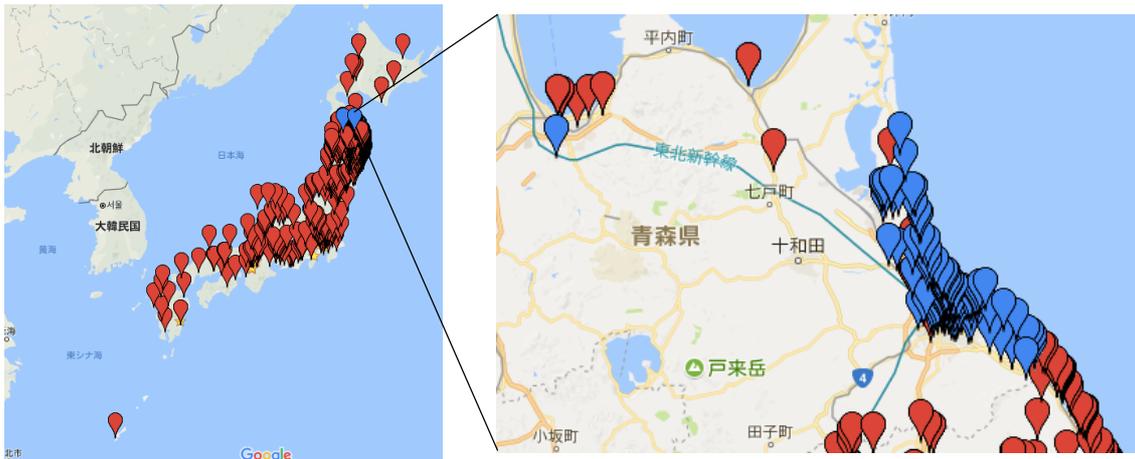


図 4-1 震災アーカイブごとの収集リソースの地理的範囲の違い

※赤：久慈・野田・普代震災アーカイブのリソース，青：青森震災アーカイブのリソース

表 4-2 アーカイブ別リソース種別

リソース種別	青森	久慈	震録伝	河北	郡山
写真	55,027	96,902	124,435	32,630	37,144
文書データ	11,728	27,710			9,609
映像資料	142	168		57	5
録音資料	142	2,020			
図書	399				
動画				57	
MovingImage				57	
新聞				83,175	
静止画資料				2	
電子資料			929		2,847
参考情報		609			
会議録		52			
政府刊行物		5			
和古書			17		
記事・論文			33		
合計	67,438	127,466	125,414	115,978	49,605

メタデータのタグ名「dcndl:materialType」に記述されている情報を基に、リソースの種別を集計したものを表 4-2 に示す。青森震災アーカイブ、久慈・野田・普代震災アーカイブ、

みちのく震録伝、郡山震災アーカイブでは、写真リソースが多くを占めており、報道機関である河北新報社が運営する河北新報震災アーカイブでは新聞リソースが多くを占めている。また、自治体が運営する青森震災アーカイブ、久慈・野田・普代震災アーカイブ、郡山震災アーカイブでは、写真について文書データが多くを占めた。この文書データは、復旧・復興事業等の行政文書や、被災状況に関する体験談が多く含まれた。

## 4.2. メタデータ分析

### 4.2.1. タイトルの分析

類似する文字列のタイトルは、一連に作成されたリソースか関連するリソースであると考えられる。特に、青森震災アーカイブ、久慈・野田・普代震災アーカイブ、郡山震災アーカイブでは、図 4-2 に示すタイトルのように、文字列の大部分が一致するものが多く存在する。図 4-2 の上 3 つのタイトルは「震災復興祈念おいらせほっき貝祭り会場」部分が一致しており、関連する内容のリソースであると推測できる。一方、図 4-2 の下 3 つのタイトルは、タイトルからリソースの内容まで推測することは難しいが、作成場所が近いリソースが集まっていると推測できる。

- |   |  |
|---|--|
| 1 | 震災復興祈念おいらせほっき貝祭り会場レイアウト図 (その 1)                  |
| 2 | 震災復興祈念おいらせほっき貝祭り会場設営準備作業スタッフ帳                    |
| 3 | 震災復興祈念おいらせほっき貝祭り会場設営等の作業協力依頼について                 |
| 4 | 20110602_階上町_大蛇漁港付近_屋外_青森県三戸郡階上町_復旧・復興_支援・ボランティア |
| 5 | 20110602_階上町_大蛇漁港付近_屋外_青森県三戸郡階上町_復旧・復興_被害        |
| 6 | 20110603_階上町_大蛇漁港付近_屋外_青森県三戸郡階上町_復旧・復興_被害        |

図 4-2 類似するタイトルの例

このような、タイトルが類似するリソースを集約するため、ここではタイトルのクラスタリングを行う。クラスタリング方法には、文字列の編集距離を用いる。編集距離とは、2つの文字列のうち片方の文字列をもう片方の文字列になるように編集した際の、編集に要した文字数を指す。ここでは、Levenshtein 距離を用いた。

ただし、短い2つの文字列の編集距離は、必然的に短くなる。そのため、類似するタイトルをまとめる基準として Levenshtein 距離の他に Levenshtein 距離を文字列長で徐算した値（正規化 Levenshtein 距離と呼ぶ）を使用する。Levenshtein 距離、及び正規化 Levenshtein 距離が閾値未満のタイトルを一つにまとめる。Levenshtein 距離と正規化 Levenshtein 距離の閾値はヒューリスティックに調整する。

青森震災アーカイブのタイトル 11,638 件に対して、タイトル間の Levenshtein 距離と正規化 Levenshtein 距離を算出し、Levenshtein 距離 5 未満、かつ正規化 Levenshtein 距離 0.2 未満のものをまとめた結果、6,880 件のグループにまとめられた。

図 4-2 に示した 5 番、6 番のタイトルは、他の 19 件のタイトルとまとめられた。一方、1~4 番のタイトルは他とはまとめられなかった。他にまとめられたタイトル群を確認してみると、体験談のテキストリソースは「階上町在住 20 代男性 震災体験談(震災後)」のように体験者の在住地、年齢、性別、「震災体験談」、体験時期を表す文字列で構成されており、300 件以上のタイトルがまとめられた。

図 4-2 の 1~3 のような、タイトルから関連は見られるが、編集距離が大きくなってしまいうタイトルをまとめるには、閾値を大きくするか他の条件を導入する必要がある。

他に考えられる手法としては、部分文字列ごとの出現頻度を算出し、出現頻度が高く、かつ長い部分文字列を用いてタイトルをまとめる手法が考えられる。

#### 4.2.2. 主題情報の分析

アーカイブによって、サブジェクトの出現頻度の傾向が分かれる。

(ユニーク数/資料数)を見ると、久慈、震録伝は豊富な主題語が付与されている一方、河北新報は少ない。主題語が少ないにもかかわらず、延べ数が多いことから、河北新報では同じ主題語を多くのリソースに付与していることがわかる。

表 4-3 主題語の語数の比較

主題語	ユニーク数	延べ数	資料数
青 森	14,491	334,047	68,297
久 慈	<b>32,911</b>	747,775	127,500
震録伝	<b>32,413</b>	837,226	125,414
河 北	<b>16,095</b>	<b>761,588</b>	115,867
郡 山	13,275	179,014	49,605

表 4-4 河北新報の頻出語（上位 10）

主題語	出現頻度	出現率 (出現頻度 / リソース数)
東日本大震災	57,847	49.93%
震災	32,859	28.36%
復興	24,956	21.54%
津波	24,080	20.78%
被災地	19,242	16.61%
被害	18,384	15.87%
被災者	15,620	13.48%
被災する	15,233	13.15%
津波被害	14,888	12.85%
がれき	14,808	12.78%

表 4-5 各震災アーカイブの頻出語（上位 10）

アーカイブ	頻出語
青 森	建物, 被災跡, 人, 重機, 瓦礫, 車, 破損, 漂着物, トラック, 海
久 慈	人, 建物, 瓦礫, 木, 海, 車, 山, 子供, 被災跡, 道路
震録伝	建物, 道路, 人, 瓦礫, 津波被害, 車, 樹木, 看板, 地面, 浸水域
河 北	東日本大震災, 震災, 復興, 津波, 被災地, 被害, 被災者, 被災する, 津波被害, がれき
郡 山	地震, イベント, 亀裂, 学校だより, 建物, 被害調査, 体験談, 放射線量, 除染, 工事

そもそも震災アーカイブでは震災に関わるリソースが収集されているという前提があれば、「東日本大震災」や「震災」という主題は冗長と思われる。しかし河北新報社の収集するリソースは、震災以外のものも多分に含まれるため、付与されていると推測できる。

青森、久慈、震録伝は、被写体や被写体の様子を表す語が多い。郡山は文書データ及び電子資料を多く含むアーカイブ、教育機関や行政機関が発行する文書に関連する語が多く含まれた。

### 4.2.3. 時空間情報の分析

NDL 大震災メタデータスキーマには、時空間情報に関するメタデータ項目として表 2-1 に示す 4 項目がある。このうち、空間情報には「資料の対象とする場所」と「撮影場所」がある。青森震災アーカイブと久慈・野田・普代震災アーカイブでは、これに加えてリソース作成者の所在地も記述している。青森震災アーカイブ及び久慈・野田・普代震災アーカイブのインタフェースでは「資料の対象とする場所」と「作成者の所在地」を分けて表示していることから、これらを分けてメタデータを記述していると考えられる。しかし NDL 大震災メタデータスキーマにマッピングされたデータでは、「資料の対象とする場所」と「作成者の所在地」が同一の項目にまとめて記述されてしまっている。

青森震災アーカイブと久慈・野田・普代震災アーカイブでは、写真リソースが多く収集されている（表 4-2）。これらの写真には、場所情報として、施設の対象とする場所と作成者の所在地の地理座標が記述されている。青森震災アーカイブには、地理座標に加えて住所も記述されている。作成日時としては日付と時刻が記述されている。一方、写真リソースに次いで数の多い文書リソースには、文書内に登場する場所と作成者の所在地の地理座標と住所が記述されている。作成日時としては、日付に加えて、時刻として「00:00:00」が記述されている。

みちのく震録伝は、収集しているリソースのほとんどが写真であり、場所情報として施設の対象とする場所と撮影場所の地理座標と、資料の対象とする場所の住所が記述されている。作成日時としては、日付と時刻が入力されている。

河北新報震災アーカイブのリソースには、地理座標は記述されておらず、住所のみの記述であった。作成日時として、写真リソースには日付と時刻が、新聞リソースには日付が記述されている。

郡山震災アーカイブのリソースには、地理情報として地理座標と住所が記述されている。作成日時としては、写真・文書いずれも日付のみの記述であった。

自治体が運営するアーカイブである青森震災アーカイブ、久慈・野田・普代震災アーカイブ、郡山震災アーカイブには、復旧工事等の委託事業者や市町村役所職員が、工事やイベント・行事等の様子を記録する目的で連続的に撮影した写真が多い。一方、みちのく震録伝や河北新報震災アーカイブは、研究者や記者が撮影したものが多く、撮影したすべての写真を公開するのではなく、特徴的な写真を選定して公開している。また、自治体が運営するアーカイブでは、特定の工事や開催された行事をくまなく撮影した写真が多いのに対し、みちのく震録伝では特定のイベントではなく撮影者の独自の観点で観測したものが多く含まれる。例えば、道路沿いの被害の様子を撮影したものや、特定の場所で時間をおいて定点観測的に撮影した写真が見られる。河北新報震災アーカイブでは、記事に載せる写真を想定しているためか、同一の被写体を撮影した写真が大量には発生していない。

## 5. リソース集約

### 5.1. 時空間情報と作成者情報を用いた集約

横山らは、写真リソースのみを対象として青森震災アーカイブ、久慈・野田・普代震災アーカイブ、みちのく震録伝の3アーカイブについて時空間情報を用いたリソース集約を試みた。この3アーカイブを対象とした理由は、写真リソースを多く含むアーカイブであるからである。郡山震災アーカイブは、時間情報の粒度が荒く、時刻のデータが入力されておらず日付のみであったため、対象外とした。

青森震災アーカイブ、久慈・野田・普代震災アーカイブは、写真リソースがリソース全体の70%以上を占める。次いで多いのが文書リソースであるが、この文書リソースに付与されている地理情報のほとんどは自治体の役場を指しており、また、時間情報として時刻の入力がなく日付のみ把握できるものが多く占めていた。このため、膨大な数の文書リソースが集約と判別されてしまったため、当アーカイブの文書リソースは本手法による集約には適さないと判断した。以上の理由から、集約対象を、リソースの大部分を占める写真リソースに限定することとした。

表 5-1 写真リソース数と地理情報の入力状況

アーカイブ	写真数	地理情報を持つリソース数	地理情報を二つ以上もつリソース数
青森	55,027	55,027	29,599
久慈	96,902	95,781	63,355
震録伝	124,435	96,441	0
河北	32,630	82	0
郡山	37,144	27,850	0

青森震災アーカイブと久慈・野田・普代震災アーカイブでは、一つのリソースあたり2つ以上の地理情報が付与されているものが多数存在した(表 5-1)。NDL 大震災メタデータスキーマによると、当該メタデータ項目は、リソースの主題とする場所に関する情報を記述すると規定している。しかしこの2つのアーカイブのリソースを確認した結果、リソースが対象とする場所、リソースの作成場所、リソースの作成者の所在地の3つが混合していることがわかった。混合の原因として、アーカイブ構築者がメタデータ作成の際に混同して入力した、あるいはNDL 大震災メタデータスキーマへのマッピングの際、項目を誤って統合してしまったという理由が考えられる。アーカイブのメタデータ連携の際にこのようなメタデータの意味の混合が起こるのは、検索性、利活用性の低下につながる。しかし、場所に関する3

つの情報のうち、どの値がリソースの対象とする場所であるかを機械的に特定することは困難である為、本手法においては全て同様に扱うこととした。また、時間情報が入力されていないものや、日付に加えて時刻まで入力されていないものは除外した。

本手法によって作成された集約数を、表 5-2 に示す。検索対象資料数の減少率は、8~14%程度となった。元のリソースと比較して集約数が減少することで、検索対象のリソース件数を減らすことができ、検索結果の一覧性は向上する。

表 5-2 時空間情報を用いた集約における集約対象リソース数と集約数

アーカイブ	集約対象 リソース数	集約数	リソースの 減少率
青 森	48,338	6,571	13.59%
久 慈	72,894	8,531	11.70%
震録伝	96,441	8,188	8.49%

本手法によって作成された集約の例を図 5-1、表 5-3 に示す。この例では、東北大学の研究者が、2011 年 3 月 29 日に宮城県石巻市の道路に沿って撮影した 60 枚の写真が集約されている。集約されたリソースには全て"東北大学研究者による震災写真"がタイトルとして与えられていた。主題としては、写真に写っているオブジェクトや被害の種類が付与されている。また、建物の名前といった固有名詞も記述されていた。



図 5-1 時空間情報と作成者情報を用いた集約の例 1 (一部)

(左：リソースのサムネイルとタイトル (一部)，右：地図へマッピングしたもの)

表 5-3 集約されたリソースのメタデータ例 1 (一部)

メタデータ項目	値	同じ値が付与されたリソース数
タイトル	"東北大学研究者による震災写真"	60
時間情報	"2011-03-29T13:21:29+09:00"	1
	"2011-03-29T13:19:57+09:00"	1
	"2011-03-29T13:20:03+09:00"	1
	(以下 57 件省略)	
地理情報 (緯度, 経度)	"38.419385", "141.30528166666667"	5
	"38.41518", "141.31188166666666"	2
	"38.41519", "141.31182"	2
	"38.41518666666666", "141.31182333333334"	2
	"38.415301666666664", "141.32010666666667"	1
	(以下 48 件省略)	
主題	"津波被害"	60
	"損壊"	55
	"建物"	54
	"堆積物"	52
	"(株)木の屋石巻水産"	7
	(以下 62 件省略)	

表 5-4 時空間情報を用いて作成した集約数とサイズ別集約数

アーカイブ	集約対象 リソース数	集約数	サイズ 1 の 集約数	サイズ 100 以上の 集約数
青森	48,338	6,551	1,599	102
久慈	72,894	8,514	2,023	189
震録伝	96,441	8,188	962	59

他のどのリソースとも時空間的に距離が離れおり、孤立していると判断されたリソースが数多く発生した(表 5-4, サイズ 1 の集約)。より利活用性を向上させるためには、これら孤立リソースを他のリソースと結びつける必要がある。

しかし個々の集約を閲覧する際、大量のリソースが一つに集約されている場合、集約内のリソースの一覧性が悪くなるため、さらなる細分化が必要である。

本手法で作成された、サイズが大きい集約の例を図 5-2 に示す。この集約のリソースには、

全て 2 点の地理座標{緯度: 40.678882, 経度:141.435056}, {緯度: 40.683078, 経度: 141.369032}が付与されていた。この 2 点の位置情報は、リソース作成者である三沢市役所の庁舎と、資料の作成場所であると推測される港を示していた。一方、時間情報には 872 種類の値が与えられており、2012 年 3 月 28 日 15 時頃から 20 時頃の範囲となっており、一連に撮影された写真であると推測できる。これらのメタデータや画像の類似性から、一連に作成されたものであると判断可能である。

集約されたリソースのタイトルを確認すると(表 5-5)、「20120328\_三沢市\_三沢漁港\_屋内\_青森県三沢市港町\_イベント\_産業」、「20120328\_三沢市\_三沢漁港\_屋外\_青森県三沢市港町\_イベント\_産業」、「20120328\_三沢市\_三沢漁港\_屋外\_青森県三沢市港町\_産業」の三種類の値しかないことがわかった。前者 2 つのタイトルは、「屋内」と「屋外」の文字列が異なっており、写真の内容に基づいて集約を分割することができる。

また、シリーズタイトルや写っているオブジェクトによる分割も可能と考えられる。

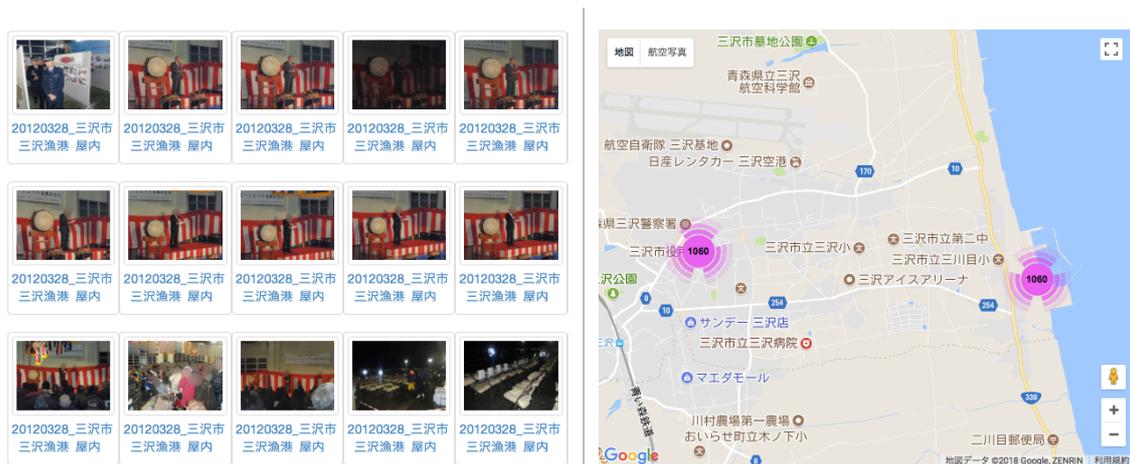


図 5-2 時空間情報を用いて集約されたリソース群の例 2 (一部)  
(左: リソースのサムネイルとタイトル (一部), 右: 地図へマッピングしたもの)

表 5-5 集約されたリソースのメタデータ例 2 (一部)

メタデータ項目	値	同じ値が付与された リソース数
タイトル	"20120328_三沢市_三沢漁港_屋内_ 青森県三沢市港町_イベント_産業"	910
	"20120328_三沢市_三沢漁港_屋外_ 青森県三沢市港町_イベント_産業"	146
	"20120328_三沢市_三沢漁港_屋外_ 青森県三沢市港町_産業"	4
シリーズタイトル	"12 03 28 ほっき祭り"	869
	"H24.3.28 復興祈念祭写真"	191
時間情報	"2012-03-28T16:10:35"	5
	"2012-03-28T18:51:02"	4
	"2012-03-28T15:50:52"	4
	(以下 869 件省略)	
地理情報 (緯度, 経度)	"40.678882, 141.435056"	1060
	"40.683078, 141.369032"	1060
主題	"人"	991
	"祭り"	752
	"ホッキ貝"	379
	"紅白幕"	137
	(以下 50 件省略)	

## 5.2. 主題情報を用いた集約

同じ主題のリソースには、同じ主題語が共起するという特徴を基に、リソースのクラスタリングを行なった。主題語の共起ネットワークをクラスタリングして主題語のクラスタを作成し、各クラスタについて、クラスタ中の主題語が主題として付与されたリソース群を集約とした。

表 5-6 主題情報を用いた集約における集約対象リソース数と集約数

アーカイブ	全リソース数	集約対象 リソース数	集約数
青 森	68,297	68,032	3,596
久 慈	127,500	127,383	8,290
震録伝	125,414	124,552	7,910
河 北	115,867	115,684	3,988
郡 山	49,605	49,374	2,881

本手法で作成された主題語クラスタの例を図 5-3 に、図 5-3 の「竹ぼうき」「ごみ回収作業」「土嚢袋」から成るクラスタから作成された集約の例を図 5-4 に示す。この集約は、「竹ぼうき」「ごみ回収作業」「土嚢袋」のいずれかを主題語として与えられたリソースを集めたものである。集められたリソースは 25 件あり、23 件が写真、1 件が動画、1 件が行政文書であった。写真と動画はいずれも工事の様子を撮影したものであった。行政文書は「報道機関への投げ込みについて」というタイトルの電子文書であったが、文書にはアクセスできなかった。これらのリソースのうち、「土嚢袋」が主題語として与えられたリソースは 19 件、「竹ぼうき」が 7 件、「ごみ回収作業」が 1 件であった。その他には、「建物」「人」「トラック」が多く付与されていた。このリソース群全体としては、概ね「土嚢袋を用いた被災地の復旧」を表していると判断できる。全てのリソースが土嚢を用いた復旧作業に関するものとは限らないが、それに関連するリソースが集まった。

図 5-3、図 5-4 の例の他に、20 件の集約も同様の確認をしたところ、いずれも何らかの主題内容に関連したリソースが集まっていることがわかった。ただし、それぞれの集約リソース群に対して「土嚢袋を用いた被災地の復旧」のようなタイトルを与えることは、ヒューリスティックでも難しい場合が多い。

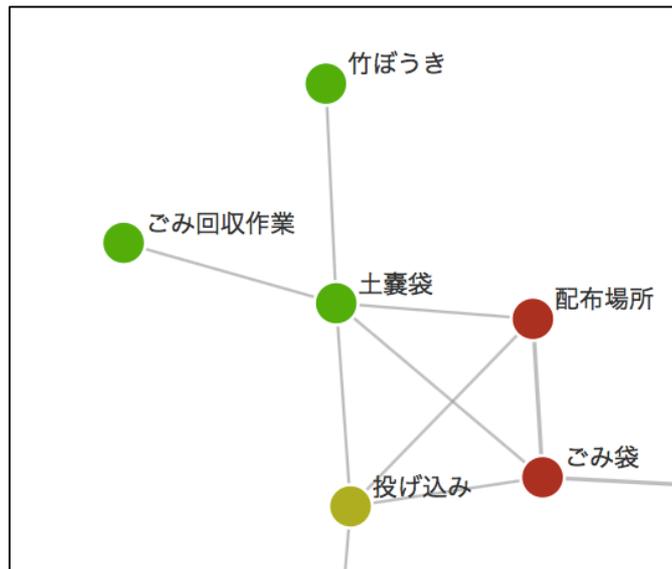


図 5-3 主題語クラスタの例

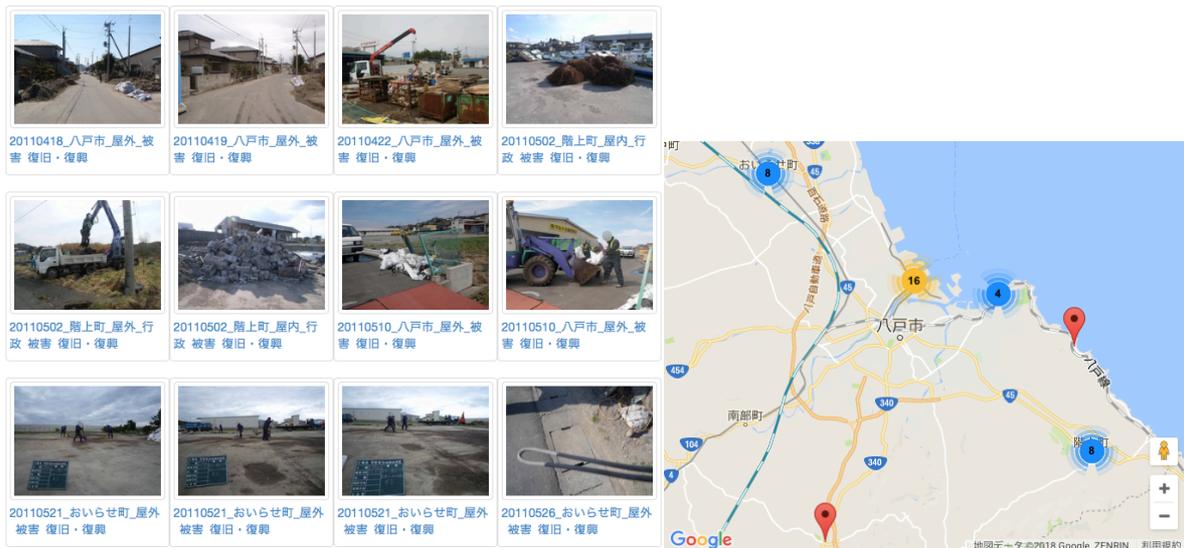


図 5-4 主題語を用いた集約の例

表 5-7 主題情報を用いて作成した集約の数とサイズ 1 の集約数

アーカイブ	集約数	サイズ 1 の集約数	サイズ 100 以上の集約数
青森	3,596	1,060	1,332
久慈	8,290	2,240	609
震録伝	7,910	2,047	4,612
河北	3,988	1,240	612
郡山	2,881	839	320

サイズ1の集約数は表5-7のようになった。あるリソースは、多くの場合、複数の語が主題語として与えられている。そのため、この集約の作成方法では、あるリソースは複数の集約に出現する可能性がある。そのため、サイズ1の集約に含まれるリソースは、必ずしも孤立したリソースとは限らない。孤立したリソース数を把握するため、サイズ1の集約に含まれるリソースのうち、他のサイズ2以上の集約に含まれるものを除外し、集計したものを表5-8に示す。

表 5-8 主題情報を用いて作成した集約の数と孤立したリソース数

アーカイブ	集約対象 リソース数	集約数	孤立した リソース数
青森	68,032	2,736	200
久慈	127,383	6,231	181
震録伝	124,552	5,874	11
河北	115,684	2,756	8
郡山	49,374	2,116	75

孤立したリソース数は集約対象のリソース数に対して十分に小さい。孤立したリソースのうち文書リソースの件数は、青森震災アーカイブは194件、久慈・野田・普代震災アーカイブは178件、郡山震災アーカイブは48件と、孤立したリソース数の多くを占めた。これらの文書リソースは、他のリソースには使用されなかった固有名詞が付与される傾向にあった。みちのく震録伝と河北新報震災アーカイブは0件であった。

サイズ100以上の集約数に着目すると、久慈・野田・普代震災アーカイブ、河北新報震災アーカイブ、郡山震災アーカイブでは少ない。その一方、青森震災アーカイブでは37.0%、みちのく震録伝では58.3%もの集約がサイズ100以上となっている。このようなサイズの大きい集約がどのような集約であるかを確認してみたところ、表4-5にあげられる、頻出する主題語を含んだクラスターから作成される集約であった。一般的に自然言語処理において、頻出語は意味を持たない語（ストップワード）として除外される場合がある。震災アーカイブの主題語においても、「建物」や「道路」といった、多くのリソースに主題として付与される語は、そのリソースを特徴付けるものではないと判断できる。

かさ じゅうぎ ない イカ釣り漁船 オブジェ ガソリン計量機 ガラス 顔後  
 ガードパイプ ゲート コンクリート壁 コンクリート柱 シャッター 説明版  
 ショベルカー セメントタンク テトラポット トイレタンク トラック石堀 門柱  
 ブレーカーボックス ブロック塀 プール ミラー リヤカー 事務機器 押入れ  
 人ガードレール 倒壊 冠水跡 冷凍ショーケース 凝着物 噴水 塗炭 墓標  
 小上がり 床タイル 建物 指 業務用エアコン 水産加工排水処理装置 一輪車  
 汚水蓋 海重機 消火栓 漁船用給油施設 漂着物 片づけ作業 牽引車 瓦礫  
 町並み 番号札 破損 綱 被災ゴミ瓦礫 被災跡 被災 f ゴミ 視線湯道標 鉄扉  
 証明写真撮影機 貯蔵タンク 車庫 農機 道路冠水 雨水缶 電気基盤 l 機械  
 高架水槽 魚屋 魚網 ゴール ビニールホース ローラーコンベア  
 安全第一 家畜小屋 漂着物川 河口付近 玄関照明 破損車両  
 ※下線：全リソースの 1% に対して共通に付与されている主題語

図 5-5 サイズが大きい集約の主題語クラスタの例

青森震災アーカイブから作成した集約のうち、集約サイズが最も大きい 33,279 の集約の主題語を図 5-5 に示す。「建物」や「被災跡」は、約 68,000 件のリソースがある青森震災アーカイブで 15,000 以上のリソースに付与されている主題語である。このような語が同じクラスタに分類されることは、主題語の関連性を機械的に判別するには有用であるが、本論文ではリソースを集約することに焦点を当てているため、これらの頻出語はストップワードとして扱い、除外すべきである。

ストップワードの対象とする主題語は、各アーカイブにおいて 1% 以上のリソースに共通に付与されているものとした。付録 3 に添付するストップワードを除外し、同様の手法で集約を行なった。これにより、大きい集約が減少した。小さい集約は増加したものの、孤立したリソースに変化は見られなかった。

時空間情報による集約手法では、写真のみを対象としていた。時空間情報を用いて集約されたリソース群は、同じ被写体が写っていることが多い。連続写真によって出現数が増えすぎてしまう。時空間情報を用いた集約手法で生成された集約群に対して主題語のクラスタリングを行うと、増えすぎていた出現が抑えられるため、クラスタリングの結果及び集約結果が大きく異なると予想。比較するため、主題情報についても写真のみを抽出したデータセットの構築を行なった。

### 5.3. オントロジーを用いた集約

オントロジーを用いた集約手法では、語の正確な意味に基づいた関連によってリソースを集約することを目的とする。

震災アーカイブでは、被害の復旧・復興に当たった組織名や、建物の名前といった固有名詞が付与されていることがしばしばある。本研究では、震災アーカイブオントロジーを構築する上で、まずは組織・施設の関係性を記述することから始めることとした。また、固有名詞は青森震災アーカイブのメタデータから抽出し、構築を行なった。青森震災アーカイブを選択した理由は、2点ある。リソース数が郡山震災アーカイブについて少ないため、固有名詞抽出を網羅しやすい。また、メタデータの付与方法が久慈・野田・普代震災アーカイブと酷似しているため、応用が効きやすいと推測できるためである。

オントロジーを構築し、リソース集約をする手順は以下の通りである。

1. 固有名詞の抽出
2. 固有名詞の上位概念の構築
3. 固有名詞とリソースのリンク

以降、構造化対象の固有名詞を「インスタンス」、固有名詞の分類（上位概念）を「クラス」と呼ぶこととする。

「病院」「株式会社」といった施設名や組織名のプレフィクスやサフィクスに相当する文字列を使用し、リソースのタイトル・主題語に対して検索し、固有名詞を抽出してインスタンスとした。クラス構造は、既存の分類「施設用途分類」「国土数値情報」「Wikipedia」を参考に、震災アーカイブ独自の構造を構築した。インスタンスとリソースとの関連づけは、インスタンスのラベルとリソースのメタデータの文字列とのマッチングにより行う。使用するメタデータは、リソースの内容を記述していると考えられる主題、タイトル、シリーズタイトル、アブストラクトとした。インスタンスは青森震災アーカイブから抽出したものであるが、他のアーカイブでもマッチングを試みた。

青森震災アーカイブから抽出した固有名詞をインスタンスとしているので、青森震災アーカイブのリソースと数多くマッチしているのは自然である。他でアーカイブでマッチした固有名詞は、「自衛隊」や「東北地方整備」といった、青森県に関わらず復旧・復興事業に関わった組織名や、「中央公民館」といった、複数の組織や施設で同一の名称を持つような語であった（表 5-10）。また、久慈・野田・普代震災アーカイブ、みちのく震録伝、河北新報震災アーカイブは東北地方を中心として地理的に幅広い領域を対象としたリソースを収集しているため、青森県固有の名詞も数多くマッチした。

しかし、集約対象とできるリソース数は青森震災アーカイブでは3割程度となり、他のアーカイブでは5%を下回った。

表 5-9 オントロジーを用いた集約における集約対象リソース数と集約数

アーカイブ	全リソース数	マッチしたリソース数	カバー率
青森	68,297	22,940	33.59%
久慈	127,383	4,608	3.62%
震録伝	124,552	1,935	1.55%
河北	115,684	2,117	1.83%
郡山	49,374	723	1.46%

表 5-10 オントロジーとマッチした固有名詞とリソース数（青森震災アーカイブを除く）

固有名詞	マッチしたリソース数
久慈港	2,148
自衛隊	1,978
仙台港	882
中央公民館	636
漁業協同組合	610
仮設店舗	599
東北地方整備局	309
郵便局	283
仙台駅	235
陸上自衛隊	225

オントロジーを構築したことによって語と語の関係を機械的に判断可能となったため、推論を用いてリソース検索することができるようになった。例えば、「マルサ嵯峨」というテキストがメタデータとして与えられていた場合、従来は単純なテキスト検索しかできなかったのに対し、「マルサ嵯峨」が「水産加工場」であるという事実に基づいてリソースを検索することが可能である。これにより、リソースに「水産加工場」というテキストがメタデータとして付与されていない場合でも、「水産加工場」と検索したとき「マルサ嵯峨」に関するリソースを取得することができる。

## 5.4. 複数手法の併用

### 5.4.1. 時空間情報と主題情報の併用

時空間情報を用いた集約手法では、一連に撮影された写真のリソース群がまとめられた。そこで作成された集約群に対して、主題情報に基づいた集約手法を適用した。

集約前のアーカイブリソースに対して主題情報を用いた手法を適用した場合、集約のサイズが大きいものが多数作成されてしまった。本手法を用いることで、類似する内容のリソースを、一連に撮影された写真群単位でまとめることができる。

本手法で作成された集約数を表 5-11 に示す。

表 5-11 主題情報を用いた集約における集約対象リソース数と集約数

アーカイブ	時空間情報を用いた集約	集約数
青森	6,545	526
久慈	8,511	806
震録伝	8,173	2,714
郡山	1,749	641

### 5.4.2. 主題情報とオントロジーの併用

オントロジーを用いた集約では、語の意味に基づいた関連のあるリソースを集約することができた。主題語を用いた集約では、統計的に関連の強い語をまとめ、大まかに関連のある集約を構築した。主題語を用いた集約に対して、オントロジーを用いた手法でリソース集約をすれば、大まかな関連のあるリソース群単位で、意味に基づいた関連のあるものをまとめることが可能である。

オントロジーを用いてリソースを集約すると、オントロジーの語彙とリソースのマッチ件数は高くなく、多くのリソースをカバーできなかった。主題語を用いた集約を対象としてオントロジーで集約することで、大まかに関連のあるリソース群によってカバー率を広げることがもできる。

表 5-12 固有名詞とリソースとのリンキング

固有名詞	個別リソースで マッチング	主題語による集約との マッチング	集約に含まれる リソース数の和
小中野小学校	1,792	71	51,646
三沢市漁業協同組合	91	41	52,192
株式会社松本鐵工所 八戸事務所	71	31	51,535
北日本鍍金株式会社	59	32	50,343
八戸東高等学校	58	29	49,752
青森工場	54	50	50,886
八戸工場付近	44	40	52,576
以下略			

## 6. リソース集約システム

5章では、時空間情報と作成者情報を用いた手法、主題を用いた集約手法、オントロジーを用いた集約手法を震災アーカイブのリソースに適用し、データセットを構築した。この段階では、集約結果の分析として統計的な分析を行った。しかし、利活用性の高い集約が作成されているかどうかを評価するには、実際に利用可能な形でのアウトプットが必要である。そのため、5章で構築したデータセットにメタデータを付与し、RDFで記述した。このRDFはSPARQLエンドポイントへ格納し、httpリクエストによってメタデータを参照できるようにした。また、このメタデータを用いてリソースを集約し閲覧することのできるインターフェースを開発した。

時空間情報を用いた集約には、メタデータとして集まったリソースの中心時刻（平均を使用した）と中心地理座標、作成者を付与した。

主題情報を用いた集約には、集約に用いた主題語クラスに含まれる主題語を付与した。

データセットの公開後、外部のリソースとリンクして様々なリソースを発見できるよう、メタデータはRDFで記述しRDFストアへ格納、SPARQLで問い合わせできるようにした。

本データセットを用い、集約閲覧システムを開発した。集約のメタデータ、及び集約されたリソースのメタデータを閲覧することができる。集約されたリソースの地理情報を地図上にマッピングし、リソースの位置を確認できる。オントロジーを利用し、概念の上位下位概念をたどりながら、特定の概念に関するリソース、及び集約を閲覧することができる。

The screenshot displays two main sections: 'Ontology Structure' and 'Related Resources'.  
**Ontology Structure:**

- URI:** [http://mdlab.slis.tsukuba.ac.jp/disaster/tmp\\_vocabulary/1.1.171204/class/ショッピングセンター](http://mdlab.slis.tsukuba.ac.jp/disaster/tmp_vocabulary/1.1.171204/class/ショッピングセンター)
- Pref Label:** ショッピングセンター
- Super Class:** 商業施設
- Instances (1):** ショッピングセンター駐車場

**Related Resources:**

- Original Resources (3):** Three image thumbnails showing a street scene with a blue fence and a paved area.
- Aggregations (8):** Three aggregation cards, each containing an image thumbnail and the following metadata:
  - archive: aomori
  - title: 20110312\_八戸市\_ショッピングセンター

図 6-1 リソース集約システムを用いた集約の閲覧（オントロジーを用いた集約例）

## 7. 考察

5章では、時空間情報を用いた手法、主題情報を用いた手法、オントロジーを用いた手法でリソースを集約した。

時空間情報を用いた手法では、撮影機会の観点からリソースを集約した。この手法で作成された集約には、道路沿いに移動しながら撮影した写真群、ある復興イベントの会場内を歩きながら撮影した写真群といったものがあった。これらの集約は、ある時点での、ある地点の周辺情報を知るために有用である。また、本手法で用いた GeoHexV3 の機能を用いれば、地理的隣接関係に基づいて集約間をリンクすることができ、移動しながら撮影した写真群をたどりながら閲覧することが可能となる。

時空間情報を用いて、他の観点からの集約手法も考えられる。ある地点に関して、時間において撮影した写真は、定点観測写真として利用価値がある。

主題情報を用いた集約手法では、大まかに内容が関連するリソースが集められた。従来の、アーカイブリソースが並列にリストとして取得されるシステムと比較すると、リソース間の関連を発見しやすい。この集約を用いることで、アーカイブリソースを用いて被災地のストーリーを記述するキュレーションタスクの支援ができるのではないかと考えられる。

また、時空間情報を用いた集約手法で作成されたデータセットに対して主題情報を用いて集約すると、一連に作成された写真群に対して、何らかの主題に関連しているものが集められる。主題情報を用いた集約手法によって作成された集約と同様、キュレーションタスクの支援として有用と考えられる。前者と後者とで、どのように用途が変化しうるかは、別途分析が必要である。

オントロジーを用いた集約手法では、構造化された語の意味に基づいて、関連するリソースを集めることができた。震災アーカイブは、防災・減災、復旧・復興に活用されることを目的としている。震災において、避難施設の確認や、用途別の施設の被災状況の確認、災害支援組織に関する情報収集は必須である。これまで震災アーカイブでは、施設や組織の関係を用いた検索はできなかった。本研究で構築したオントロジー、およびオントロジーを用いたリソースの集約のデータセットは、災害・減災対策、復旧・復興事業において有用であろうと考えられる。

## 8. おわりに

本研究プロジェクトは、2015年より始動した。始動した当初は、「利活用性の向上」という抽象的な目的のために、どのようなアプローチを取るかをディスカッションしていた。始めに挙げたのは、キュレーション支援であった。Japan Digital Archiveでは、ユーザが独自の観点からリソースをキュレートし、共有できたため、同様のタスクを震災アーカイブでも実現するインタフェースの構築を目的とした。しかし、横断検索サービスであるひなぎくでリソースの検索をしても、被災や復旧・復興の様子を淡々と記録したリソースが並び、また被写体に関する主題情報しか得られなかったため、何をどう収集したてまとめたら良いのかわからず、震災アーカイブの利活用性の低さを実感した。そして、利活用性の向上を図るためには、インタフェースの構築以前に、データセットの改善を行うべきであろうと判断した。

複数の情報を与えられた際、人はその関連性を想起できるため、何らかの観点からリソースを集約することで、地域の被災や復興のストーリーを発見できるのではないかと考えた。多様な観点から収集されたリソースを閲覧することで、発見できるストーリーも異なってくる。

本研究では、データセットの構築にとどまり、その評価まで行うことができなかった。今後、集約システムを用いて集約の評価を行なっていきたい。

現在はアーカイブごとに、アーカイブ内のリソースを用いて集約を作成している。アーカイブを超えたリソースから集約を作成したい。

それぞれの手法で作成した集約のメタデータは、集約されているリソースのユニオンとしている。メタデータの付与は、それぞれの集約が何であるかを定義することに相当するため、どのようなメタデータを付与するのが妥当であるかを検討する必要がある。また、機械的に付与できるデータ、機械的には付与できないデータに分けて検討する必要がある。

本研究で作成したオントロジーは、施設と組織の用途の観点から分類したものである。今後、外部語彙 (DBpedia, WordNet) とリンクさせ、語の意味に基づいた検索による震災アーカイブの利活用支援に取り組みたい。

構築したデータセットが実際に利活用されるべく、リリースしたい。現在は、筑波大学内からのみアクセス可能な形での公開となっている。

総務省の DigitalArchiveNetwork ワークショップにて、本データセットのアーカイブ構築者視点からの利用像が提案された。東北大学災害科学国際研究所の柴山は、付与されたメタデータのバリデーションに使えるのではないかと提案する。メタデータ分析で得られた知見から、ひなぎくでのメタデータ連携時の項目のマッピングの見直し、スキーマ定義の改善の必要性も指摘された。

本研究で構築したデータセットを用いれば、キュレーションが容易になるのではないかと考えている。今後、キュレーションタスクによってユーザ評価を行なっていきたい。

## 謝辞

主指導教員の杉本重雄先生、副指導教員の永森光晴先生、三原鉄也先生には、ご指導いただきましたことに厚く感謝を申し上げます。メタデータ研究室では良き先輩方や同期、後輩達に恵まれ、研究や私生活において多々助けられました。

また、Digital Archive Network(DAN)ワークショップに参加させていただいた折は、共通の対象への問題意識を持った方々と議論することができ、本研究の有用性を確認することができ、研究の助けとなりました。

改めて、私の研究活動に関わった皆さまに感謝を申し上げます。

## 文献リスト

- [1] The Library of Congress. "Home Page". American Memory from the Library of Congress. 1994. <https://memory.loc.gov/ammem/index.html>, (参照 2018-01-10).
- [2] 国立国会図書館. "国立国会図書館デジタルコレクション". 2011. <http://dl.ndl.go.jp>, (参照 2018-01-10).
- [3] 文化庁. "文化遺産オンライン". 2008. <http://bunka.nii.ac.jp>, (参照 2018-01-10).
- [4] アジア歴史資料センター. "アジア歴史資料センター". 2001. <https://www.jacar.go.jp>, (参照 2018-01-10).
- [5] Europeana. "Europeana Collections". 2017-10-31. <https://www.europeana.eu>, (参照 2018-01-10).
- [6] Digital Public Library of America. "Digital Public Library of America". 2018-01-3. <https://dp.la>, (参照 2018-01-10).
- [7] 国立国会図書館. "国立国会図書館サーチ". 2018-01-09. <http://iss.ndl.go.jp>, (参照 2018-01-10).
- [8] 国立国会図書館. "検索対象データベース". 国立国会図書館サーチ. 2018-01-5. <http://iss.ndl.go.jp/information/target/>, (参照 2018-01-10).
- [9] 知的財産戦略本部. "知的財産推進計画 2017". 2018-5-16, p. 81. <https://www.kantei.go.jp/jp/singi/titeki2/kettei/chizaikeikaku20170516.pdf>, (参照 2018-01-10).
- [10] 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". 2017-12-26. <http://kn.ndl.go.jp/>, (参照 2018-01-10).
- [11] Europeana. "Europeana Data Model Documentation". Europeana. 2014-11-18. <https://pro.europeana.eu/resources/standardization-tools/edm-documentation>, (参照 2018-01-10).
- [12] 積佑典, 本間維, 三原鉄也, 永森光晴, 杉本重雄. "アーカイブ横断型メタデータ連携による東日本大震災アーカイブ群からのコンテンツ集約手法". 情報処理学会全国大会講演論文集. 2016, vol. 78, no. 1, p. 557-558.
- [13] 横山雄哉, 積佑典, 三原鉄也, 永森光晴, 杉本重雄. "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". 情報処理学会全国大会講演論文集. 2017, vol. 79, no. 4, p. 939-940.
- [14] 八戸市, 三沢市, おいらせ町, 階上町. "青森震災アーカイブ". 2016-01-08. <http://archive.city.hachinohe.aomori.jp>, (参照 2018-01-10).
- [15] 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 (2016年10月版)". 国立国会図書館東日本大震災アーカイブ. 2016.

- [http://kn.ndl.go.jp/static/files/ndlkn\\_schema\\_Ja201610.pdf](http://kn.ndl.go.jp/static/files/ndlkn_schema_Ja201610.pdf), (参照 2018-01-10)
- [16] 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)". 総務省. 2013-03.  
[http://www.soumu.go.jp/menu\\_seisaku/ictseisaku/ictriyou/02ryutsu02\\_03000114.html](http://www.soumu.go.jp/menu_seisaku/ictseisaku/ictriyou/02ryutsu02_03000114.html), (参照 2018-01-10)
- [17] Tim Berners-Lee. "Linked Data". W3C. 2006-07-27.  
<https://www.w3.org/DesignIssues/LinkedData.html>, (参照 2018-01-10)
- [18] 神崎正英. "リンクするデータ". The Web KANZAKI. 2011-03-07.  
<https://www.kanzaki.com/works/2011/pub/0307wi2.html#s2>, (参照 2018-01-10)
- [19] 青森震災アーカイブ. "H23年度 生活・経済活動復興支援助成金受付集計表". 2014.  
<http://archive.city.hachinohe.aomori.jp/archive/item/?id=M2014031017200092167>, (参照 2018-01-10)
- [20] Europeana. "Portraitbüste des Agrippa Postumus (?)". Europeana. 2016-04-26.  
[http://data.europeana.eu/aggregation/provider/2058811/DAI\\_500f9d9e605706a652fcd195d7946374\\_\\_artifact\\_\\_cho](http://data.europeana.eu/aggregation/provider/2058811/DAI_500f9d9e605706a652fcd195d7946374__artifact__cho), (参照 2018-01-11)
- [21] 佐藤翔輔. "視点 東日本大震災アーカイブを使ってみた". 情報管理. 2017, vol. 59, no. 10, p. 690-694. [https://www.jstage.jst.go.jp/article/johokanri/59/10/59\\_690/\\_pdf/-char/ja](https://www.jstage.jst.go.jp/article/johokanri/59/10/59_690/_pdf/-char/ja), (参照 2018-01-10)
- [22] みちのく震録伝. "みちのく・いまをつたえ隊(07 利府・多賀城)". 2015-01-07.  
<http://search.shinrokuden.irides.tohoku.ac.jp/shinrokuden/uuid/28aabb60-9608-11e4-9f70-000c2923bf22>, (参照 2018-01-10)

# 付録

## 付録 1 Europeana で 2 件以上のリソースから成る集約リソースを取得する SPARQL 文

Endpoint の URL: <http://sparql.europeana.eu>

クエリ入力ボックスに以下を入力して[Run Query]をクリックする

```
select ?aggregation (count(?view) as ?views_num)
where {
  ?aggregation a ore:Aggregation ;
    edm:hasView ?view .
}
group by ?aggregation
having(count(?view)>1)
limit 100
```

これによって 2 列の表が結果として返ってくる。「aggregation」列が集約の URI、「views\_num」列が集約されているリソースの数である。

## 付録 2 リソースのメタデータ例

見やすさのため、インデント・改行を施した。

```
<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:ada="http://archive.city.hachinohe.aomori.jp/terms/"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:dcndl="http://ndl.go.jp/dcndl/terms/"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#"
  xmlns:v="http://www.w3.org/2006/vcard/ns#"
  xmlns:cc="http://creativecommons.org/ns#"
  xmlns:ma="http://www.w3.org/ns/ma-ont#"
  xmlns:premis="http://multimedialab.elis.ugent.be/users/samcoppe/
    ontologies/Premis/premis.owl#"
  xmlns:dcat="http://www.w3.org/ns/dcat#"
  xmlns:ndlkn="http://kn.ndl.go.jp/terms/"
  xmlns:jmp20="http://zgate.gsi.go.jp/ch/jmp/"
  xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <ndlkn:MetaResource
    rdf:about="http://archive.city.hachinohe.aomori.jp/archive/item/
      ?id=M2014031017200092167">
    <dcterms:created rdf:datatype="http://purl.org/dc/terms/W3CDTF">
      2014-03-10T17:20:00
    </dcterms:created>
    <dcterms:modified rdf:datatype="http://purl.org/dc/terms/W3CDTF">
      2014-12-24T10:56:37
    </dcterms:modified>
    <dcterms:conformsTo rdfs:label="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 2013 年 3 月版"
```

```

    rdf:resource="http://kn.ndl.go.jp/schema/2013/03/ndlkn.pdf"/>
<dcterms:creator>
  <foaf:Agent rdf:about="http://id.ndl.go.jp/auth/entity/00279933">
    <foaf:name>
      八戸市
    </foaf:name>
    <dcndl:transcription>
      ハチノヘシ
    </dcndl:transcription>
  </foaf:Agent>
</dcterms:creator>
</ndlkn:MetaResource>
<ndlkn:Resource
  rdf:about="http://archive.city.hachinohe.aomori.jp/archive/item/
    ?id=M_2014031017200092167#entity">
<dcterms:title>
  H23年度 生活・経済活動復興支援助成金受付集計表
</dcterms:title>
<dc:title>
  <rdf:Description>
    <rdf:value xml:lang="ja">
      H23年度 生活・経済活動復興支援助成金受付集計表
    </rdf:value>
    <dcndl:transcription xml:lang="ja">
      H23 ネット セイカツ・ケイザイカツドウフッコウシエンジョセイキン
      ウケツケシュウケイヒョウ
    </dcndl:transcription>
  </rdf:Description>
</dc:title>
<dcndl:seriesTitle>
  <rdf:Description>
    <rdf:value>
      平成 23 年度生活経済活動復興支援助成金①
    </rdf:value>
  </rdf:Description>
</dcndl:seriesTitle>
<dcterms:creator>
  <foaf:Agent>
    <foaf:name xml:lang="ja">
      おいらせ町
    </foaf:name>
    <dcndl:transcription xml:lang="ja">
      オイラセチヨウ
    </dcndl:transcription>
  </foaf:Agent>
</dcterms:creator>
<dcterms:contributor>
  <foaf:Agent>
    <foaf:name xml:lang="ja">
      おいらせ町
    </foaf:name>
    <dcndl:transcription xml:lang="ja">
      オイラセチヨウ
    </dcndl:transcription>
  </foaf:Agent>
</dcterms:contributor>
<dcterms:subject>
  <rdf:Description>

```

```

    <rdf:value>
      行政
    </rdf:value>
  </rdf:Description>
</dcterms:subject>
<dcterms:subject>
  <rdf:Description>
    <rdf:value>
      行政文書
    </rdf:value>
  </rdf:Description>
</dcterms:subject>
<dcterms:subject>
  <rdf:Description>
    <rdf:value xml:lang="ja">
      復興, 支援, 助成金, 集計表
    </rdf:value>
  </rdf:Description>
</dcterms:subject>
<dcterms:language rdf:datatype="http://purl.org/dc/terms/
  ISO639-2">
  jpn
</dcterms:language>
<dcterms:extent xml:lang="ja">
  1 枚
</dcterms:extent>
<dcndl:materialType rdfs:label="文書データ"
  rdf:resource="http://ndl.go.jp/ndltype/Document"/>
<dcterms:identifier
  rdf:datatype="http://www.w3.org/2000/01/rdf-schema#Literal">
  コンテンツ番号 : OIJ00000001
</dcterms:identifier>
<dcterms:spatial>
  <rdf:Description>
    <v:label>
      青森県上北郡おいらせ町中下田 135-2
    </v:label>
    <v:region>
      青森県
    </v:region>
    <v:locality>
      上北郡おいらせ町
    </v:locality>
    <geo:lat>
      40.599087
    </geo:lat>
    <geo:long>
      141.397772
    </geo:long>
  </rdf:Description>
</dcterms:spatial>
<dcterms:temporal>
  <rdf:Description>
    <rdf:value xml:lang="ja">
      2011年7月~2012年3月
    </rdf:value>
  </rdf:Description>
</dcterms:temporal>
<dcterms:accessRights>

```

```

S01P99U99
</dcterms:accessRights>
<dcterms:rights>
  <rdf:Description>
    <dcterms:description>
      保存、公開とも無期限
    </dcterms:description>
  </rdf:Description>
</dcterms:rights>
</ndlnk:Resource>
</rdf:RDF>

```

### 付録 3 主題語を用いた集約手法におけるストップワード

アーカイブ名	ストップワード	語数
青森	建物, 被災跡, 人, 重機, 瓦礫, 車, 破損, 漂着物, トラック, 海, 被災ゴミ, 道路, 損傷, 木, 小学校, 防災教育, 被災船, 電柱, コンテナバッグ, 電信柱, 地面, 防潮堤, 子供, 災害ゴミ, 看板, 仮置き場, コンクリート, フェンス, 船, ガードレール, ブルーシート, 防波堤, コンテナバック, 木材, 雪, 黒板, ビニールシート, 被災車両, 津波, 家, 浸水跡, ロープ, 損壊, 車両, 壁, 瓦礫集積, 法面, タイヤ, 学校・教育機関, カラーコーン, 消波ブロック, 森林, 倒壊, 机, 泥, 窓, イス, 段ボール, カゴ, 護岸, 浸水, 道路標識, 車止め, 祭り, 作業員, 漁船, ダンプ, 漁網, 機械, 亀裂	70
久慈	人, 建物, 瓦礫, 木, 海, 車, 山, 子供, 被災跡, 道路, 学校・教育機関, 被災, 家, 重機, 船, 大人, 防波堤, 漁港, 椅子, 電柱, テーブル, マイク, 祭り, 消波ブロック, がれき, 津波, テント, ガードレール, 復興, 踊り, 海岸, 災害, トラック, ゴミ, 体育館, 被災後, 車両, 川, ステージ, 森林, 水門, 泥, 破損, 消防, 花, 漂着物, 防潮堤, 橋, 村長, 住宅, 岩, 太鼓, クレーン, 工事, 施設, 大漁旗, 漁船, まつり, ボランティア, 津波後, 歌手, 支援, 機械, 式典, 屋内, 線路, 鳥居, イス, ヘルメット	69
震録伝	建物, 道路, 人, 瓦礫, 津波被害, 車, 樹木, 看板, 地面, 浸水域, 木, 山, 電柱, 工事, 草, 自動車, 損壊, 被害, damage, 草地, 海, 住宅, ショベルカー, 林, 沿岸部, 砂利, 草木, トラック, 歩道, 土地, 人物, 掲示物, 損傷, 松林, 電線, ガードレール, 作業車, 船, イベント, 重機, 雪, 堤防, 農地, フェンス, 建物損傷, 土, 泥, 柵, カラーコーン, 海岸, 民家, 植木, 盛土, 建物被害, 住居, 堆積物, 家, ブルーシート, 損傷物, 建物跡地, 雑草, 駐車場, 漁船, 枯草,	101

	校舎, 流出物, テント, 土嚢, 漁港, 土砂, 水, 建物跡, マツの木, 建物内部, 群集, ポスター, ビニールハウス, のぼり, 曇り空, 室内, 川, クレーン車, バックフォー, 乗用車, 橋, 線路, クレーン, 仮設住宅, コーン, 浸水, バリケード, 被害跡, 更地, 廃棄物, 旗, 青空, 掲示板, 砂浜, 鉄塔, メッセージ, 店舗	
河北	東日本大震災, 震災, 復興, 津波, 被災地, 被害, 被災者, 被災する, 津波被害, がれき, 支援, 地震, 被災, 事故, 損壊, 福島第1原発, おとな, 仮設住宅, 災害, ボランティア, 避難, 復旧, 避難所, 電柱, 避難する, 車, 阪神・淡路大震災, 支援する, 男性, 津波被災地, 戸建, 道路, 再開する, 復興支援, 再生, 女性, 建物, 見通し, 希望, 笑顔, 沿岸部, 樹木, 仮設, 不安, 原発事故, 提供・がんばろう気仙沼, 岩手・宮城内陸地震, 宮城県沖地震, 応援, 原発, 再開, 避難者, 義援金, 行方不明者, 大震災, 流木, 晴れ, 雑草, 放射性物質, 提供・みやぎ生協, 災害公営住宅, ボランティア活動, 屋内, 森林, 屋外, 大津波, 死者, 室内, 基礎, 復旧する, 避難生活, 応援する, 余震, 復興計画, 希望する, 災害対策本部, tsunami, 絆, ベクレル, セシウム, 冠水, テーブル, 電気, 重機, 泥, 提供・東北学院大学災害ボランティアステーション, 宮城県連続地震, damage, 地震被害, 遺体, トラック, 暮らし, 物資, 子ども, 新潟県中越地震, 復興する, 海, 住宅, 救援・救助	99
郡山	地震, イベント, 亀裂, 学校だより, 建物, 被害調査, 体験談, 放射線量, 除染, 工事, 子ども, 川内村, 崩落, 復旧, 震災語り人, 視察, 避難所, 道路, 小学校, 倒壊, 表土除去, ボランティア, 破損, 外壁, 防災訓練, 避難, 仮設住宅, 住宅, 線量計測, 水道, 給水, 壁, 慰問, 地面, 会議, 天井, テレビ, 中学校, 富岡町, 通知, 支援物資, 支援活動, 復興, 物産展, 市長, 広報誌, 校庭	47