

## 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

武田 侑季<sup>†1</sup> 積 佑典<sup>†2</sup> 三原 鉄也<sup>†3</sup> 永森 光晴<sup>†3</sup> 杉本 重雄<sup>†3</sup>

**概要:** 東日本大震災アーカイブの利活用性向上には、アーカイブされたリソースのメタデータを利用し、関連性の強いリソースでまとめることが有効である。本研究では、東日本大震災アーカイブにおいて特徴的な主題内容を表す用語に焦点を当て、オントロジーの開発を進めた。本報告では、東日本大震災アーカイブの一つである青森震災アーカイブのメタデータを対象として、施設・組織に特化して試作したオントロジーについて述べる。

**キーワード:** 東日本大震災, デジタルアーカイブ, メタデータ, 固有名詞, オントロジー

### Development of an ontology to aggregate metadata of resources in the Great East Japan Earthquake digital archives

YUKI TAKEDA<sup>†1</sup> SEKI YUSUKE<sup>†2</sup> TESTUYA MIHARA<sup>†3</sup>  
MITSU HARU NAGAMORI<sup>†3</sup> SUGIMOTO SHIGEO<sup>†3</sup>

**Abstract:** One effective way to improve the usability of the Great East Japan Earthquake Digital Archives is to aggregate archived resources that have strong semantic relation to each other, e.g. photographs about a particular events. This paper shows an ontology which is designed to aggregate metadata of those resources using semantic relationships among the resources. The ontology is designed to represent regional entities such as markets, buildings and events because the disaster digital archives primarily collect regional records. We collect phrases focusing on proper nouns from metadata of the Aomori archive, which is one of the Great East Japan Earthquake Digital Archives, because many of those proper nouns are names of the regional entities. We used a few authoritative vocabularies of public institutions and organizations to organize the ontology.

**Keywords:** The Great East Japan Earthquake, digital archives, metadata, proper noun, ontology

#### 1. はじめに

情報社会が発達し、個人の携帯端末によって、情報を手軽に取得、公開できる時代となった。そんな多くの情報が世に溢れている中で、重要な情報を整備し、必要に応じて提供するため、様々なものを対象としてデジタルアーカイブが構築されている。その中でも2011年3月11日に発生した東日本大震災を対象として文書資料、写真、音声、動画など多様なデジタル資料を扱うアーカイブは、災害に関わる地域の記録と記憶を蓄積し、防災や復興に重要な役割を持つものと認められている。国立国会図書館東日本大震災アーカイブひなぎく<sup>[1]</sup> (以下、ひなぎくと呼ぶ) は東日本大震災に関わる多数のデジタルアーカイブから、メタデータを収集し、資料に対して横断的なアクセスを可能

にしている。資料の種類、数が豊富に蓄積公開され、個々の資料を検索できる。一方、こうしたデジタルアーカイブに収録される多様な資料 (以下、リソースと呼ぶ) に関するメタデータは一点毎のリソースに与えられる。そのため、一か所で撮影された写真や相互に関連する写真と文書資料であっても個別のリソースとして扱われるため、結果的にリソース個々が独立して扱われることになり、検索結果表示の中で、リソース間の関係とは無関係に大量に並んでしまうという課題がある。当研究室では、このような課題を解決し、東日本大震災アーカイブの利活用性を向上するためには、関連するリソースを何らかの視点からまとめることが重要であると考え、様々な視点からメタデータを集約することによってリソースのまとまりを作り利用者に提供するというアプローチで研究を進めている。現在まで、

<sup>†1</sup> 筑波大学情報メディア創成学類  
College of Media Arts, Science and Technology, Univ. of Tsukuba.  
<sup>†2</sup> 筑波大学大学院図書館情報メディア研究科  
Graduate School of Library, Information and Media Studies, Univ. of

Tsukuba.  
<sup>†3</sup> 筑波大学図書館情報メディア系  
Faculty of Library, Information and Media Science, Univ. of Tsukuba.

リソースに付与された「時空間情報」を用いた集約<sup>[2]</sup>や、「キーワードの共起関係」を用いた集約<sup>[3]</sup>を作成してきた。これらの研究から、東日本大震災アーカイブの課題に対して、関連性の強いリソースの集約を利活用性向上に役立てるための知見を得た。そこで、本研究では、青森震災アーカイブ<sup>[4]</sup>および久慈・野田・普代震災アーカイブ<sup>[5]</sup>のメタデータから、特徴的な主題内容を表す用語として組織や施設等に関係する語句を収集し、震災アーカイブ向けのオントロジーの開発を進めている。本報告では、青森震災アーカイブを対象として試作したオントロジーについて述べる。

## 2. 東日本大震災アーカイブの連携と利活用

### 2.1 東日本大震災アーカイブ

ひなぎくは、56の震災アーカイブや関連リソースを収集したデータベースを検索対象とし、多様な震災関連リソースをアクセス可能にしている(2017年12月現在)。そして、ひなぎくは連携する多くの震災アーカイブが持つリソースの横断利用を可能にするために、Dublin Coreを基盤にしたシンプルな構造のメタデータスキーマ<sup>[6]</sup>と、OAI-PMH<sup>[7]</sup>に基づくメタデータハーベスティングを採用している。

本研究において主に利用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、それぞれ、青森県八戸市、三沢市、おいらせ町、階上町と岩手県久慈市、野田村、普代村が構築したものである。2017年12月時点における両震災アーカイブが持つリソース数は、青森震災アーカイブ 68287件、久慈・野田・普代震災アーカイブ 127500件である。両アーカイブともに、自治体の収集した多様な資料に対してメタデータを付与し、デジタルアーカイブ化を進めたものである。多様な資料の中でも特に写真のデータが大量にあるが、写真に映ったものがそのままタイトルやキーワードとしてメタデータで付与されている。これは、メタデータ付与は予算的、時間的制約もあって、震災アーカイブ向けに作られたガイドライン<sup>[8]</sup>に沿って、非メタデータ専門家によって行われたことが主因であると考えられる。このように、一点毎のリソースに対してメタデータが付与されるという特色を持つ一方、シンプルなスキーマとガイドラインに基づき作成されたメタデータ品質に起因するいろいろな課題を持っている。

### 2.2 震災アーカイブの問題点

問題点の一つに、震災アーカイブの連携のしづらさが挙げられる。東日本大震災のような県をまたいだ広域な災害である場合、リソースの利活用のためには、地域ごとに独立しているアーカイブを連携させることが必要になる。一方で、地域特有の固有名詞が付与されたリソースを扱う場合、タイトルやキーワードが、単純な文字列での識別にしか使われていなく、文字列が示す概念の持つ関係(例えば、同じ種類に属する言葉であること)が与えられていないため、関連性の高いリソース同士をまとめることができ

ず、バラバラに表示されてしまうこともあり得る。例えば、水産に関する資料を探したい場合、青森震災アーカイブ内の資料の1つである「八戸水産会館」と、久慈・野田・普代震災アーカイブ内の資料の1つである「マルサ嵯峨」は、どちらの資料にも、「水産に関する資料である」ことがデータとして付与されていないため、まとめて表示することができない。現状では、固有名詞が示す対象を利用者が認知していなければ、関連した資料を見つけることができない。

### 2.3 本研究の目的

そこで本研究では、タイトルやキーワードという主題情報(以下、主題語と呼ぶ)に含まれる語句を抽出し、それを意味にもとづいて階層化し、機械解釈が可能な形式で表現すること(すなわち、オントロジー化すること)で、主題を表す語句の意味によって、類似・関連する資料を集約し、検索・閲覧できるようにすることを目的としている。RDF化したメタデータにおいてリテラルとして扱われていた記述内容に対して、概念を与えることで、単純なキーワード検索でしか扱うことができなかった資料が、例えば、上位下位関係で集約し、まとめて表示できるようになる。さらには、表記や日付の詳細が異なっても、概念同士の関係に基づき結びつけることで、類似した資料をまとめることができる。

本研究では、主題語の中でも、「津波」「行政」といった普通名詞よりも、「小中野中学校」「川口コミュニティセンター」などといった、固有名詞の方が地域に依存するリソースの内容をより具体的に表していると思われるため、固有名詞でまとめることで、より地域に基づく関連性の高い集約ができると考え、固有名詞に注目して語句を抽出することとした。一方で、固有名詞を扱う際、地域に関する背景知識がないために、類似した内容のものを機械的にまとめて表示することが難しい。そこで、本研究では、津波被害や復旧・復興に関わるリソースに関連が強いと思われる施設や組織の観点から語句を組織化し、オントロジーを構築した。

例えば、「公民館」と調べた時に、現状では、「公民館」のキーワードがついたリソースのみが検索結果として羅列されるが、オントロジーを使うことによって、「集会所」として、「集会所」や集会所の上の階層である「教育文化施設」のメタデータとともに繋げ、まとめることが可能となる。さらには、「川口コミュニティセンター」や「下長公民館」といった、固有名詞に対して、「公民館」である、という概念をリソースに付与することで、扱いづらかった固有名詞を一般化した語彙として扱うことが可能になる。

### 2.4 関連研究

2.1でも述べたように、震災アーカイブのリソース一点毎に主題語が付与されている。大量のデジタルデータ

キーワードによって集約する手法として、Folksonomy がある。Folksonomy とは、エンドユーザーが付与するキーワードのタグ付け（ユーザが対象資料の特徴を端的に表すキーワードを自身の主観によって選択する行為）によって、分類することを言う。Folksonomy に関係した研究として、小原<sup>[9]</sup>と丹羽<sup>[10]</sup>が挙げられる。小原では、Flickr 内の写真を対象にし、写真に付けられたタグの関係分析を行い、利活用性の向上を示した。この研究では、タグの形容詞と名詞の組み合わせの関係を分析してそのエントロピーを用いて、類似した資料をまとめている。丹羽では、Web の膨大なドキュメントを対象にし、Folksonomy のタグ間の共起率に基づいて、タグをクラスタリングし、類似したドキュメントをまとめ、階層化させる手法を論じている。

キーワードによる集約手法についていくつか述べたが、震災アーカイブのメタデータを集約する場合、同様の手法を適用することはできない。これは、メタデータの付与の仕方に違いがあるからである。震災アーカイブの写真のメタデータの場合、写真撮影者ではないメタデータ作成担当者がキーワードを付与する。そのため、撮影目的や対象に関する撮影者の意図を反映した語を入れることは困難であり、写真に映る内容をそのまま主題語として用いていることが多い。エンドユーザーがその写真の撮影意図や対象を理解しにくいだけでなく、キーワードの類似性によってまとまりを作ったとしても、その写真の示す対象が適切に定義されていないならば、まとまりとして写真にアクセスすることは難しい。また、この関連研究では、共起関係から階層構造を作成しているが、背景的な内容との関係性が階層構造に反映されていない。そこで、本研究では、震災や地域といったドメインに応じて作成したオントロジーを使い、概念に基づくまとまりの作成を目指す。

### 3. メタデータ集約のためのオントロジー

#### 3.1 本研究におけるオントロジーの役割

本研究におけるオントロジーの役割は、利用者が特定の地域に対する知識を持たなくとも、同様の役割や意味を持ったリソースが自動的に結びつけられ、まとめて扱うことを可能にすることである。オントロジーの多くは、ある分野において、言葉と概念を統一して扱いたい時に使われる。オントロジーを使うことにより、知識利用の際に、前提条件などの背景知識がオントロジーによって明確化されているため、他者が記述した知識であっても理解が容易となり、他のオントロジーとの対応を調べることで知識の変換も可能となる<sup>[11]</sup>このようにして、オントロジーは知識の共有・再利用性の向上に貢献し、あらゆる知識を扱う際の基盤としての役割を果たす。

#### 3.2 オントロジー作成の概要

本研究では、固有名詞に対応する上位概念を作成し、既存の分類項目を用いて階層化する、という手順でオントロ

ジーを作成する。その際に、震災アーカイブ内のリソースが持つ主題語を取り出し、その主題語に含まれた固有名詞と繋ぐことができるオントロジーを目指す。シンプルな階層構造を表現するために、SuperClass-SubClass の関係を用いる。例えば、「下長公民館」という主題語が付与されたリソースに対して、「公民館」という概念を付与する。SuperClass-SubClass の関係によって、「公民館」の上位概念である「教育文化施設」に分類される、「小中野中学校」とまとめることを目指す。(図 1) また、この SuperClass-SubClass の関係づけの際には、既存の分類項目を用いつつ、より人間の意図する概念や分類にするために、手作業でオントロジーの作成を行った。そして、震災アーカイブ内に現れた固有名詞をインスタンスとして扱い、オントロジーの下位クラスに結びつけを行うこととした。本研究で扱うオントロジーのモデルを図 2 に示す。

### 4. 震災アーカイブのためのオントロジーの作成

以下では、ひなぎくと連携している震災アーカイブの 1 つである青森震災アーカイブのメタデータを利用して作成したオントロジーについて述べる。

#### 4.1 震災アーカイブのメタデータ調査

青森震災アーカイブの主題語を対象として、主題語に使われている固有名詞を分類の観点から調査した。

主題語に使われている多くの固有名詞は、「小中野中学校」や「おいらせ漁業協同組合」といった、施設・組織名が高い割合を占めていることが判明した。また、人の手でメタ

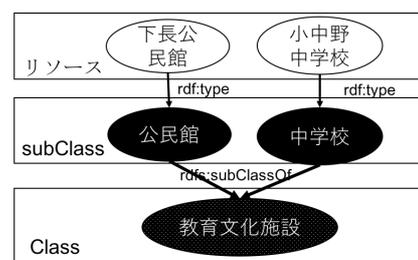


図 1 SuperClass-SubClass を用いた集約例

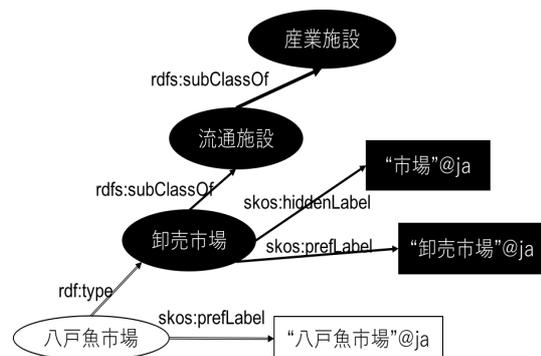


図 2 オントロジーのクラス・インスタンスのモデル

データを作成されているため、「川口コミュニティセンター」と「川口コミュニティーセンター」など、表記揺れがいくつか存在していることが確認できた。

調査の結果を踏まえ、主題語に現れる施設・組織名を持つ資料を集約対象とし、震災と地域に根付いたオントロジーの作成を行った。確認できた表記揺れは、SKOS<sup>[12]</sup>のhiddenLabelで付与し、表記揺れの改善を行った。本研究では、オントロジーの作成にはオントロジエディタ protégé<sup>[13]</sup>を用いた。

#### 4.2 本オントロジーで利用する語彙

本オントロジーでは、主として施設・組織に関する固有名詞に関係する「施設用途分類i」<sup>[14]</sup>と「国土数値情報ii」<sup>[15]</sup>に掲載されている語彙をクラスとして利用した。一方で、震災アーカイブに特化したオントロジーの作成を目指しているため、「防災施設」、「インフラ施設」は重要であると考え、例外としてこの二つの分類に関しては普通名詞を含めて扱うこととした。また、この2つの施設分類の階層は、Wikipediaの「Category:防災施設」<sup>[16]</sup>と「Category:インフラストラクチャー」<sup>[17]</sup>を参考にした。

それぞれ、(1)既存のデータを扱う、(2)「施設用途分類」と「国土数値情報」を使って固有名詞を扱う、(3)Wikipediaのカテゴリーを利用する、という(1)～(3)の段階ごとにオントロジーを作成した。

#### 4.3 オントロジーの具体的な作成手順

**第1段階:**当研究室において、昨年、震災アーカイブのメタデータ内に頻繁に現れる固有名詞を対象にし、固有名詞ごとにリソースを集約する研究<sup>[18]</sup>を行った。そこで作成された、主題語に現れる固有名詞のリストを用いて、その固有名詞を、オントロジー作成におけるインスタンスとして扱うことにした。そして、その固有名詞からその分類となる概念を推定し、インスタンスのクラスを定義した。例えば、インスタンス「川口コミュニティセンター」に対して、「公民館」をクラスとする。(図3)

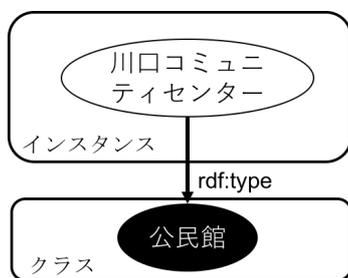


図3 クラスとインスタンスの繋げ方  
 (例:川口コミュニティセンター)

**第2段階:**第1段階では、インスタンスに直接繋がるクラスのみを作成しており、そのクラスの上位下位概念や、同じ親を持つことになるという概念が足りなかったため、「施設用途分類」と「国土数値情報」を用いて、青森震災アーカイブのメタデータに含まれる固有名詞の中で対応する施設等を探した。具体的には、「施設用途分類」の分類項目を最上位クラスとし、インスタンスに直接繋がる最下位クラスは、「国土数値情報」の施設の分類を用いた。そして、このクラスに対して、国土数値情報の施設分類別データから抽出した施設名リストを使い、そのリスト内によく現れた普通名詞をラベル (skos:hiddenLabel) として付与した。これらのラベルの語を主題語にもつりソースを調べ、その主題語を確認し、ラベルを含んだ主題語のうち、固有名詞と思われる語を取り出し、インスタンスとしてオントロジーに付与した。(図4)

ここで、インスタンスとして扱うことができなかつたリソースの主題語を分析したところ、「港」や「漁業」というラベルによって、施設とは関係のないリソースが取り出されていることがあった。例えば、港湾一般を示す「港」を主題語に含むリソースの中には、「港町」という固有の地名があるために、その「港町」の場所に関連するリソースも「港」として取り出されてしまっていた。こうした必ずしも施設・組織を示す語が含まれるケースについては、本オントロジーで対象とする「港」と異なるため、ここで扱うべきではないと考えた。そこで、インスタンスのラベルを持つリソースのみを扱うこととした。

**第3段階:**このオントロジーは、震災アーカイブに特化したオントロジーであるため、震災に関連した施設を関連づけるための主題語が必要である。震災時に重要となる施設として、防災施設やインフラ施設が挙げられる。しかし、第2段階において基にした施設用途分類や国土数値情報は通常の施設の用途の観点でのみ分類されており、防災やインフラに関する分類がない。そこで、Wikipediaの「Category:

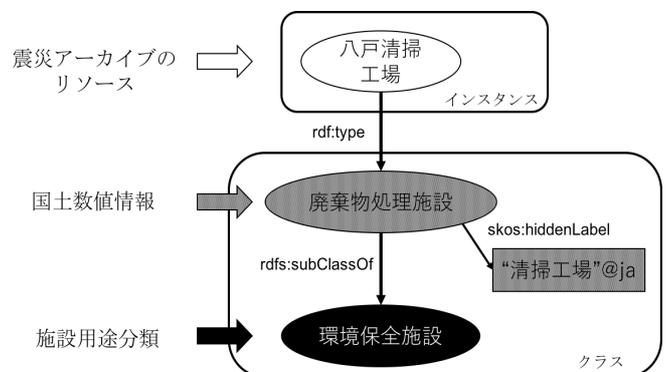


図4 オントロジーと元にしたデータの繋がり

i 一般社団法人公共建築協会が運営している公共建築設計者情報システムで用いられている施設の用途別分類表  
 ii 全国各地にある公共施設の名称や場所のデータを施設の分類別に

提供している、国土交通省が提供している公共施設など国土に関する基礎的な空間情報のデータベース集

防災施設」と「Category:インフラストラクチャー」を参考にし、新たにクラスを作り、オントロジーを拡張した。

#### 4.4 作成したオントロジーについて

前節で記した工程を経て、震災に特化したオントロジーを作成したオントロジーのクラス階層を表1に示す。

また、表2~4に前節の各項ごとのリソース数を示す。

表2 1段階目のインスタンスとその語彙を持つリソース数 (上位5件)

主題語	主題語を持つリソース数
小中野小学校	1792
三沢市漁業協同組合	91
株式会社松本鐵工所八戸事務所	71
北日本鍍金株式会社	59
八戸東高等学校	58

表3 上位階層とその下位に分類されるクラスのインスタンスの語彙をもつリソース数

インスタンスの上位クラス	リソース数
住宅	5
教育文化施設	2775
産業施設	8440
福祉医療施設	102
行政施設	1043
金融機関	145

表4 3段階目に追加した語の上位階層とその下位に分類される語彙をもつリソース数

上位クラス	リソース数
インフラストラクチャー	4542
防災施設	4700

以上の結果を踏まえ、オントロジー作成の各段階のクラスにおけるカバー率の比較を表5に示す。カバー率とは、アーカイブ内の全リソースの中で、「施設・組織に関する固有名詞」、または、「震災時に重要となる施設の語彙」を主題語にもつリソース (DA リソースと呼ぶ) の割合のことをいう。具体的には、1段階目では、全リソースに対する、既存のデータの主題語を持つリソースの割合、2段階目では、全リソースに対する、固有名詞の主題語を持つリソースの割合、3段階目では、固有名詞、または防災、インフラ施設の主題語を持つリソースの割合のことを指す。

段階ごとに比較したところ、段階を追ってカバー率が良くなってきており、本研究では、最もカバー率の高い3段階目をオントロジーの完成形とした。

表1 完成したクラス階層

"イベント"@ja	"まつり"@ja			
"組織・施設"@ja	"インフラストラクチャー"@ja	"水道施設"@ja		
		"交通施設"@ja	"ヘリポート"@ja	
			"港湾"@ja	
			"空港"@ja	
			"鉄道"@ja	
			"駐車場"@ja	
		"情報通信施設"@ja		
		"水道施設"@ja		
		"燃料供給所"@ja		
		"石油関連インフラ"@ja		
		"電力インフラ"@ja		
"住宅"@ja		"個人住宅"@ja		
		"寄宿舍"@ja		
		"集合住宅"@ja		
"教育文化施設"@ja		"体育施設"@ja	"公園"@ja	
			"屋内体育施設"@ja	"体育館"@ja
			"屋外競技施設"@ja	"野球場"@ja
		"宗教施設"@ja	"修道院"@ja	
			"墓苑"@ja	
			"寺"@ja	
			"教会"@ja	
			"神社"@ja	
		"教育施設"@ja	"保育所"@ja	
			"学校"@ja	"中学校"@ja
				"大学"@ja
				"小学校"@ja
				"特別支援学校"@ja
				"短期大学"@ja
				"高等学校"@ja
				"高等専門学校"@ja
			"幼稚園"@ja	
		"文化施設"@ja	"博物館"@ja	
			"図書館"@ja	
			"植物園"@ja	
			"水族館"@ja	
			"美術館"@ja	
		"集会施設"@ja	"公民館"@ja	
			"商工会議所"@ja	
			"教育センター"@ja	
			"研修センター"@ja	
			"集会所"@ja	
"産業施設"@ja		"事務施設"@ja	"事務所"@ja	
			"株式会社"@ja	
		"商業施設"@ja	"ショッピングセンター"@ja	
			"ホームセンター"@ja	
			"商店街"@ja	
			"観光施設"@ja	
			"遊戯施設"@ja	
			"集客施設"@ja	
		"宿泊施設"@ja		
		"流通施設"@ja	"コンテナターミナル"@ja	
			"トラックターミナル"@ja	
			"倉庫"@ja	
			"卸売市場"@ja	
			"工場"@ja	
			"航空貨物ターミナル"@ja	
			"鉄道貨物駅"@ja	
		"農林水産施設"@ja	"水産施設"@ja	"水産加工施設"@ja
				"水産物加工施設"@ja
				"海産物"@ja
				"漁業"@ja
				"漁業施設"@ja
				"畜産施設"@ja
"研究機関"@ja				
"福祉医療施設"@ja		"医療施設"@ja	"病院"@ja	"労災病院"@ja
			"診療所"@ja	
		"福祉施設"@ja	"児童福祉施設"@ja	
			"老人福祉施設"@ja	
			"障がい者福祉施設"@ja	
"行政施設"@ja		"一般行政施設"@ja	"市区町村役場"@ja	"本庁"@ja
			"郵便局"@ja	
		"特殊行政施設"@ja	"保安防災施設"@ja	"ボランティアセンター"@ja
				"対策本部"@ja
				"消防施設"@ja
				"自衛隊"@ja
				"警察施設"@ja
			"環境保全施設"@ja	"上水道関連施設"@ja
				"下水道関連施設"@ja
				"廃棄物処理施設"@ja
"金融機関"@ja		"保険"@ja		
		"協同組織金融機関"@ja		
		"証券関連"@ja		
		"銀行"@ja		
"防災施設"@ja		"ダム"@ja		
		"保安林"@ja		
		"堤防"@ja		
		"防波堤"@ja		
		"防潮堤"@ja		
		"防潮林"@ja		
		"防災倉庫"@ja		
		"防砂林"@ja		
		"防風林"@ja		

表 5 段階ごとのクラス数、インスタンス数、青森震災アーカイブのリソースにおけるカバー率

	クラス数	インスタンス数	DA リソース数	カバー率 (DA リソース数/全リソース数)
1段階目	75	82	2614	4% (2614/68297)
2段階目	115	379	9425	14% (9425/68297)
3段階目	133	379	18667	27% (18667/68297)

## 5. 外部データへの適用によるオントロジーの評価

今回作成したオントロジーを他のデータに適用することでオントロジーの有用性を調べる。

### 5.1 既存の集約に対するオントロジーの適用

我々の研究室では、共起関係に基づくキーワードのクラスタリングによるリソース集約作成をすすめている。これは、キーワードの共起関係によるものであるため、似た内容のリソースを集約することはできるが、キーワードの意味内容によって他の集約を含むリソースに結び付けることはできない。先に述べたように既存のメタデータの意味的記述品質は高くない。そのため、背景知識等を利用せずに統計的性質のみを用いて機械的に作成した集約では、概念の関係性で類似したリソースは集約することができない。よって、他の資料にはない特徴的なキーワードが付与されている場合、集約のサイズが1となるまとまり（以下、サイズ1の集約と呼ぶ）、つまり、まとめられていない集約が多数できている。そこで、このオントロジーを使い、サイズ1の集約をまとめることができたか有効であるかを示す。サイズ1の集約をまとめる方法として、オントロジーのインスタンスを使い、そのインスタンスのキーワードを含むサイズ1の集約を対象にし、それをインスタンスの上位概念でまとめた。対象となるサイズ1の集約のうち、上位概念ごとにまとめた結果を表6に、サイズ1の集約数の変化を表7に示す。

この結果から、オントロジーによって、サイズ1の集約数を減らすことができ、特徴的な語彙を持っていたためにまとめることができなかったリソースの集約が確認できた。

### 5.2 他のアーカイブへのオントロジーの適用

アーカイブ同士を連携する際に、本オントロジーが有効であるかを確かめるため、久慈・野田・普代震災アーカイブ

表 6 まとめられたサイズ1の集約数

上位概念	まとめられたサイズ1の集約数
イベント	7
インフラストラクチャー	2
住宅	0
教育文化施設	10
産業施設	32
福祉医療施設	5
行政施設	11
金融機関	3
防災施設	0

表 7 オントロジー適用前後のサイズ1の集約数の変化

オントロジー適用前のサイズ1の集約数	オントロジーによってまとめられた集約数	オントロジー適用後のサイズ1の集約数
1060	70	990

ブに対して、本オントロジーを適用して、それによってまとめられた結果を調べた。

この適用の結果として、表8に上位クラスごとに集約されたリソース数とそのカバー率を示す。カバー率は49%となり、久慈・野田・普代震災アーカイブのリソースにも固有名詞が多く付与されていることがわかった。これより、このオントロジーは青森震災アーカイブだけでなく、久慈・野田・普代震災アーカイブでも使うことができ、複数のアーカイブ連携に有効である可能性があることを示すことができた。

## 6. おわりに

本報告では震災に関連するリソースに特化したオントロジーの作成手法について述べた。オントロジーの作成段階ごとに震災アーカイブのリソースのカバー率を比較し、最後に完成したオントロジーを用いて、オントロジー活用の可能性を述べた。現時点では、プロパティに関して、SuperClass-SubClassの関係を使うことにより、固有名詞を上位の概念で扱うことができるようになり、固有名詞を主題としたリソースも、類似した内容で集約を行うことができた。このことによって、固有名詞を普通名詞的に扱いたいという要求を満たせるようになったが、まだ、上位概念間の関係を示したいという要求が考えられる。そのために、クラス間に関係を示すプロパティ（例えば「株式会社」が「中学校」に「支援をする」という関係）を付与することで、より語彙同士の関係によるつながりや集約を作成することが可能になると想定する。また、今後、この作成

表 8 上位階層とその下位に分類されるクラスの語彙を  
 持つリソース数

上位クラス名	リソース数
イベント	23960
教育文化施設	12613
金融機関	321
行政機関	4865
産業機関	12570
住宅	5400
福祉医療施設	248
防災施設	165
インフラストラクチャー	2615
合計	<b>62757</b>
全リソース数	127500
カバー率	<b>49%</b>

したオントロジーに対して、他のオントロジー（NDL Authorities<sup>[19]</sup> や 日本語 WordNet<sup>[20]</sup>）を繋げることにより、オントロジーによるリソースのカバー範囲を広げ、集約することが出来ると考え、オントロジーを応用していくことで利活用性が向上していくと考えている。

#### 謝辞

資料を提供していただいた中山留菜氏に感謝する。本研究の一部は JSPS 科研費 JP16H01754 の助成による。

#### 参考文献

[1] “NDL 東日本大震災アーカイブひなぎく”.  
<http://kn.ndl.go.jp/>,(参照 2017-12-20).

[2] 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. 情報処理学会第 79 回全国大会, 2017.

[3] 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. p.557-558. 情報処理学会第 78 回全国大会.

[4] “青森震災アーカイブ”. <http://archive.city.hachinohe.aomori.jp/>,(参照 2017-12-20).

[5] “久慈・野田・普代震災アーカイブ”. <http://knf-archive.city.kuji.iwate.jp/>,(参照 2017-12-20).

[6] “NDL 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)”. [http://kn.ndl.go.jp/static/files/ndlkn\\_schema\\_Ja201610.pdf](http://kn.ndl.go.jp/static/files/ndlkn_schema_Ja201610.pdf), (参照 2017-12-20).

[7] “国立国会図書館サーチが提供する OAI-PMH”.  
[http://iss.ndl.go.jp/information/api/oai-pmh\\_info/](http://iss.ndl.go.jp/information/api/oai-pmh_info/), (参照 2017-12-25).

[8] “総務省「デジタルアーカイブ」の普及促進|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](http://www.soumu.go.jp/menu_seisaku/ictseisaku/ictriyou/02ryutsu02_03000114.html), (参照 2017-12-25).

[9] 小原侑也, 柳井啓司. Web 上の大量画像を用いた名詞と形容詞の関係分析. 研究報告コンピュータビジョンとイメージメディア, 2012, vol. 182, no.13, p.1-8.

[10] 丹羽智史, 土肥拓生, 本位田真一. Folksonomy の 3 部グラフ構造を利用したタグクラスタリング. 2006, JAWS2006.

[11] 古崎晃司. ドメインオントロジーの構築と利用. 2010, vol. 19, no.4, p.296-305.

[12] “SKOS”. <http://www.asahi-net.or.jp/~ax2s-kmtnt/internet/skos/note-skos-primer-20090818.html>, (参照 2017-12-20).

[13] “protégé”. <https://protege.stanford.edu/>, (参照 2017-12-20).

[14] “施設用途分類”. <http://www.pbaweb.jp/files/施設用途分類201701.pdf>, (参照 2017-12-20).

[15] “国土数値情報 ダウンロードサービス”.  
<http://nlftp.mlit.go.jp/ksj/>, (参照 2017-12-20).

[16] “Category:防災施設”. <https://ja.wikipedia.org/wiki/Category:防災施設>, (参照 2017-12-25).

[17] “Category:インフラストラクチャー”.  
<https://ja.wikipedia.org/wiki/Category:インフラストラクチャー>, (参照 2017-12-25).

[18] 中山留菜 (2016). 「主題語彙を利用した東日本大震災アーカイブのコンテンツ集約」筑波大学情報メディア創成学類平成 28 年度卒業論文

[19] “NDL Authorities”. <https://id.ndl.go.jp/auth/ndla/>, (参照 2017-12-20).

[20] “日本語 WordNet”. <http://compling.hss.ntu.edu.sg/wnja/>, (参照 2017-12-20).