

LOD を利用した放送コンテンツアーカイブ  
のためのメタアーカイブの構築

筑波大学  
図書館情報メディア研究科  
2015 年 3 月  
萩原 和樹

1. はじめに .....	1
2. 放送コンテンツのデジタルアーカイブとその連携 .....	2
2.1 放送コンテンツアーカイブ .....	2
2.2 放送コンテンツアーカイブの利用と連携 .....	5
3. Linked Open Data を利用したメタアーカイブの構築 .....	8
3.1 LOD を利用した放送コンテンツのメタアーカイブ .....	8
3.2 放送情報と主題情報のメタデータの整備 .....	10
3.3 関連研究 .....	12
4. 放送コンテンツアーカイブ連携のためのメタデータモデル .....	15
4.1 放送コンテンツのメタデータモデルの定義 .....	15
4.2 放送情報の構造と LOD 化 .....	18
4.3 放送コンテンツの主題情報の統一 .....	21
5. メタアーカイブ構築のためのメタデータ作成手法 .....	22
5.1 WWW 上にあるアーカイブからのメタデータ取得 .....	22
5.2 RDF ファイルの記述と外部 LOD へのリンク作成 .....	25
6. 適用事例 ～NHK デジタルアーカイブスのメタアーカイブ構築～ .....	28
6.1 対象とするアーカイブ ～NHK デジタルアーカイブス～ .....	28
6.2 テレビ番組情報のメタデータ構築 .....	31
6.3 NHK-DA メタアーカイブのメタデータ記述規則 .....	34
6.4 NHK-DA 統制語彙の作成 .....	40
6.5 メタアーカイブシステムの構築 .....	41
7. 考察と課題 .....	47
8. おわりに .....	50
謝辞 .....	51
参考文献 .....	52

## 目次

図 1	BBC Archives	3
図 2	ina.fr サイト	3
図 3	NHK デジタルアーカイブス トップページ	4
図 4	エコチャンネルの Web ページ ( <a href="http://cgi4.nhk.or.jp/eco-channel/jp/movie/play.cgi?did=D0013770773_00000">http://cgi4.nhk.or.jp/eco-channel/jp/movie/play.cgi?did=D0013770773_00000</a> )	6
図 5	NHK クリエイティブ・ライブラリーの Web ページ ( <a href="http://www1.nhk.or.jp/creative/material/02/D0002040075_00000.html">http://www1.nhk.or.jp/creative/material/02/D0002040075_00000.html</a> )	7
図 6	デジタルアーカイブの連携図	7
図 7	メタアーカイブシステム概要図	8
図 8	RDF グラフ 表現例	9
図 9	放送情報とアーカイブの放送コンテンツの関係	10
図 10	主題情報とアーカイブの放送コンテンツの関係	11
図 11	Europeana トップページ ( <a href="http://www.europeana.eu/">http://www.europeana.eu/</a> )	13
図 12	Europeana Data Model の RDF グラフ	13
図 13	国立国会図書館東日本大震災アーカイブ トップページ ( <a href="http://kn.ndl.go.jp/">http://kn.ndl.go.jp/</a> )	14
図 14	既存の放送コンテンツとアーカイブのメタデータの構造を表現した RDF グラフ	15
図 15	既存の放送コンテンツのウェブページとそのメタデータ ( <a href="http://cgi4.nhk.or.jp/eco-channel/jp/movie/play.cgi?did=D0013773004_00000">http://cgi4.nhk.or.jp/eco-channel/jp/movie/play.cgi?did=D0013773004_00000</a> )	15
図 16	本研究で提案するメタデータモデル	16
図 17	テレビ番組の構造を表した RDF グラフ	19
図 18	放送コンテンツアーカイブとテレビ番組情報の関係を記述した RDF グラフ	19
図 19	放送コンテンツと主題情報を記述した RDF グラフ	21
図 20	ウェブブラウザのデベロッパーツールの使用例	23
図 21	出力した CSV ファイル	23
図 22	アーカイブの動画一覧ページ	24
図 23	RDF/TURTLE で記述されたトリプル 例	25
図 24	キーワードをラベルに持つリソースが DBpedia に存在するか問い合わせるクエリ	26
図 25	NDL にキーワードをラベルに持つリソースが存在するか問い合わせるクエリとそのリ ソースの URI を取得するクエリ	26
図 26	RDF/TURTLE 形式で記述された「コウノトリがよみがえる里(1)」のメタデータ	27
図 27	戦争証言アーカイブス ( <a href="http://www.nhk.or.jp/shogenarchives/">http://www.nhk.or.jp/shogenarchives/</a> )	28
図 28	東日本大震災アーカイブス ( <a href="http://www9.nhk.or.jp/311shogen/">http://www9.nhk.or.jp/311shogen/</a> )	29
図 29	エコチャンネル ( <a href="http://www.nhk.or.jp/eco-channel/">http://www.nhk.or.jp/eco-channel/</a> )	29
図 30	NHK 映像マップ みちしる ( <a href="http://www.nhk.or.jp/michi/">http://www.nhk.or.jp/michi/</a> )	30
図 31	NHK クリエイティブ・ライブラリー ( <a href="http://www1.nhk.or.jp/creative/">http://www1.nhk.or.jp/creative/</a> )	30
図 32	みのがし なつかし ( <a href="http://www.nhk.or.jp/archives/search/onair/">http://www.nhk.or.jp/archives/search/onair/</a> )	31
図 33	みのがし なつかしが公開しているテレビ番組の情報 ( <a href="http://cgi2.nhk.or.jp/archives/tv60bin/detail/index.cgi?das_id=D0009050069_00000">http://cgi2.nhk.or.jp/archives/tv60bin/detail/index.cgi?das_id=D0009050069_00000</a> )	32
図 34	「連続テレビ小説 あまちゃん」のメタデータの RDF/TURTLE ファイル	33
図 35	システム概要図	41
図 36	メタアーカイブシステム遷移図	42
図 37	メタアーカイブ トップページ画面	42

図 38	全文検索「ラーメン」による検索結果の表示	43
図 39	メタアーカイブ キーワードページ画面	43
図 40	キーワード「兵庫県」を持つリソースの URI とタイトルを取得する SPARQL 構文	44
図 41	テレビ番組ジャンルページ画面	44
図 42	テレビ番組の個別ページ画面	45
図 43	「NHK ニュースおはよう日本」をシリーズに持つリソースの URI とタイトルを取得する SPARQL 構文	45
図 44	メタアーカイブ 放送コンテンツ画面	46
図 45	アーカイブ「エコチャンネル」の放送コンテンツのメタデータを ID から取得する SPARQL 構文	46

## 表目次

表 1	基準となる放送コンテンツのメタデータ語彙と記述規則	17
表 2	放送コンテンツのメタデータで使用する名前空間接頭辞	17
表 3	テレビ番組シリーズのメタデータ記述構造（簡易 DSP）	19
表 4	テレビ番組シリーズの出演者・制作者の構造化メタデータ（簡易 DSP）	20
表 5	テレビ番組で使用する名前空間接頭辞	20
表 6	5.1 節で出力したメタデータのプロパティと値	26
表 7	「連続テレビ小説 あまちゃん」のメタデータ詳細	32
表 8	戦争証言アーカイブスの放送コンテンツで定義する名前空間接頭辞	34
表 9	戦争証言アーカイブスの放送コンテンツのメタデータ記述規則	34
表 10	戦争証言アーカイブスの放送コンテンツのメタデータ記述規則 構造化メタデータ	35
表 11	東日本大震災アーカイブスの放送コンテンツで定義する名前空間接頭辞	35
表 12	東日本大震災アーカイブスの放送コンテンツのメタデータ記述規則	36
表 13	エコチャンネルの放送コンテンツで定義する名前空間接頭辞	37
表 14	エコチャンネルの放送コンテンツのメタデータ記述規則	37
表 15	NHK 映像マップ みちしるの放送コンテンツで定義する名前空間接頭辞	38
表 16	NHK 映像マップ みちしるの放送コンテンツのメタデータ記述規則	38
表 17	NHK クリエイティブ・ライブラリーの放送コンテンツで定義する名前空間接頭辞	39
表 18	NHK クリエイティブ・ライブラリーの放送コンテンツのメタデータ記述規則	39
表 19	3 つのアーカイブのキーワード数と他 LOD とのリンク可能な数	40
表 20	NHK-DA 全体での総キーワード数とリンク可能な数	40
表 21	NHK デジタルアーカイブス 記述項目対応表	47



## 1. はじめに

近年、テレビで放送された番組の映像・音声の情報を保存してアーカイブしていく試みがなされている。放送というメディアは電波を用いて一方的に不特定多数の人々に向けたもので、そのコンテンツを保存していく試みはあまりなされていなかった。しかし、映像や音声情報の保存技術の進歩につれ、放送番組は後世に保存していくべき文化遺産であると考えられるようになってきた。時事的なニュース映像も後の時代には学術研究や批評の対象になることも理由に挙げられる。海外での事例を述べると、英国放送協会（BBC）では 1936 年から全放送番組を対象にアーカイブが行われており、フランスでも国立視聴覚研究所（INA）がすべての放送番組を管理し、活発な研究利用が行われている。日本では日本放送協会が（NHK）が 2003 年から番組やアーカイブの整備に取り組んでいる。代表的なものを挙げると、World Wide Web（WWW）上で公開している「NHK デジタルアーカイブス」がある。このサービスは NHK が保存している映像や音声のある一定のテーマに沿ったアーカイブとして公開されており、自由に閲覧できるようになっている。

テレビで放送された映像や音声を WWW 上で公開するデジタルアーカイブが増えて、検索や閲覧などの利用性の向上が求められるようになってきた。アーカイブの利用性向上のための手法として、デジタルアーカイブ同士の連携がある。デジタルアーカイブの連携とは、複数のアーカイブから何らかの関連を持つ所蔵資料のメタデータを収集し、資料の相互運用性を向上させることで関連する資料の統合検索を可能にし、効率よく情報を入手できるようにする試みである。実際に複数のアーカイブからメタデータを収集し、それらを統合検索できるようにしたシステムをメタアーカイブと呼ぶ。複数のアーカイブを連携させる上で課題となるのは、記述規則の異なるメタデータの統一と関連するコンテンツを結び付ける共通語彙の整備である。既存のアーカイブの記述規則を共通化することは現実的ではなく、個別のアーカイブの要件に則した記述を無視することも適切ではない。

本研究では、既存の放送コンテンツアーカイブを連携して、統合的な検索を可能にしたメタアーカイブの構築を行う。そのために個別のアーカイブの記述規則を保持しつつ、コンテンツ同士の関連性を持たせるためにアーカイブ間で共通するメタデータである主題情報の統合及び放送コンテンツに共通するテレビ番組情報の Linked Open Data 化を行う。放送情報であるテレビ番組情報と WWW 上の情報資源であるアーカイブのコンテンツを LOD 化することで、異なるアーカイブのコンテンツであってもその関連している情報同士を結びつけることが可能となる。本研究では、放送情報の LOD 化手法及びそれらと結び付けるためのアーカイブのコンテンツのメタデータ記述規則の定義について述べる。さらに NHK デジタルアーカイブスを対象として、WWW 上の情報資源からメタデータモデルの適用と LOD データセットの作成を行いつつ、作成したメタデータを元に複数のアーカイブのコンテンツを統合的に検索することが出来るメタアーカイブの構築を行う。

以下、2 章では放送コンテンツを公開しているデジタルアーカイブの現状と連携する際の問題について説明する。3 章ではデジタルアーカイブを連携させるための手法を分析し、連携に必要なメタデータについて説明を行う。4 章では 3 章の手法に基づき、放送コンテンツのアーカイブに適したメタデータモデルの提案を行う。5 章では実際にメタデータモデルを NHK デジタルアーカイブスに適用した事例を述べつつ、作成したメタデータを利用したメタアーカイブの構築について紹介する。6 章ではメタデータモデルの提案とメタアーカイブの構築における考察と課題を検討する。7 章では本研究のまとめを述べる。

## 2. 放送コンテンツのデジタルアーカイブとその連携

本章では、本研究で対象とする放送コンテンツのデジタルアーカイブがどのようなものか説明を行い、それらを連携する上でどのような問題があるか述べる。

### 2.1 放送コンテンツアーカイブ

アーカイブとは、将来に向けて、文書あるいは記録などのコンテンツの収集・蓄積を行い、それらを長期に渡って提供するサービス、あるいはそれを行う組織のことである。古文書や公文書などの重要な文化遺産を後世に保存していくことは文化活動を更に発展させていくための基盤となる上、現代の文化や歴史を正しく認識することを可能にするなど重要な意味を持つ。アーカイブへの認識が高まる中、より長期的にコンテンツを保存していくアーカイブの形としてデジタル技術を用いたデジタルアーカイブを作成する事例が増えてきた。

デジタルアーカイブとは「図書・出版物、公文書、美術品・博物品、歴史資料等公共的な知的資産の総デジタル化を進め、インターネット上で電子情報として共有・利用できる仕組み」([1]より引用)である。また、デジタルアーカイブには、建造物や遺跡、美術品などの物体をデジタル化したものも含まれる。保存の難しい文書や文化財などのコンテンツの情報をデジタル化し、長期に渡って提供できることがデジタルアーカイブの利点である。

本研究におけるデジタルアーカイブは蓄積されたコンテンツを WWW 上でだれでも自由に検索・閲覧することができるサービスとして捉える。その中でもテレビやラジオで放送された映像や音声(放送コンテンツ)を公開しているものを放送コンテンツアーカイブとして定義する。このアーカイブは WWW 上にあるサイトでデータベースにアクセスすることで誰でも自由に映像・音声などの放送されたコンテンツそのものもしくは一部を検索・閲覧することが可能である。

放送コンテンツアーカイブの代表的な事例は、英国放送協会(BBC) [2]とフランスの国立視聴覚研究所(INA) [3]が挙げられる。イギリス BBC の BBC Archives (図 1 参照)では全放送番組を対象に保存活動を行っており、2008 年の時点でその量はテレビ番組が 65 万時間、ラジオ番組は 35 万時間にも上る[32]。INA (図 2 参照)では 2006 年から所有する映像資料、約 10 万番組、1 万時間に及ぶコンテンツがインターネットを通じて自由に閲覧することが可能となった[32]。

日本でも 2003 年から日本放送協会(NHK)が保存活動を始めており、NHK アーカイブスと呼ばれる組織で NHK のテレビ・ラジオ番組等の映像・音声の保存を行っている。NHK アーカイブスがインターネットを通じて公開している NHK デジタルアーカイブス[4]では約 1 万本の映像がテーマ別に配信されている(図 3 参照)。



図 1 BBC Archives

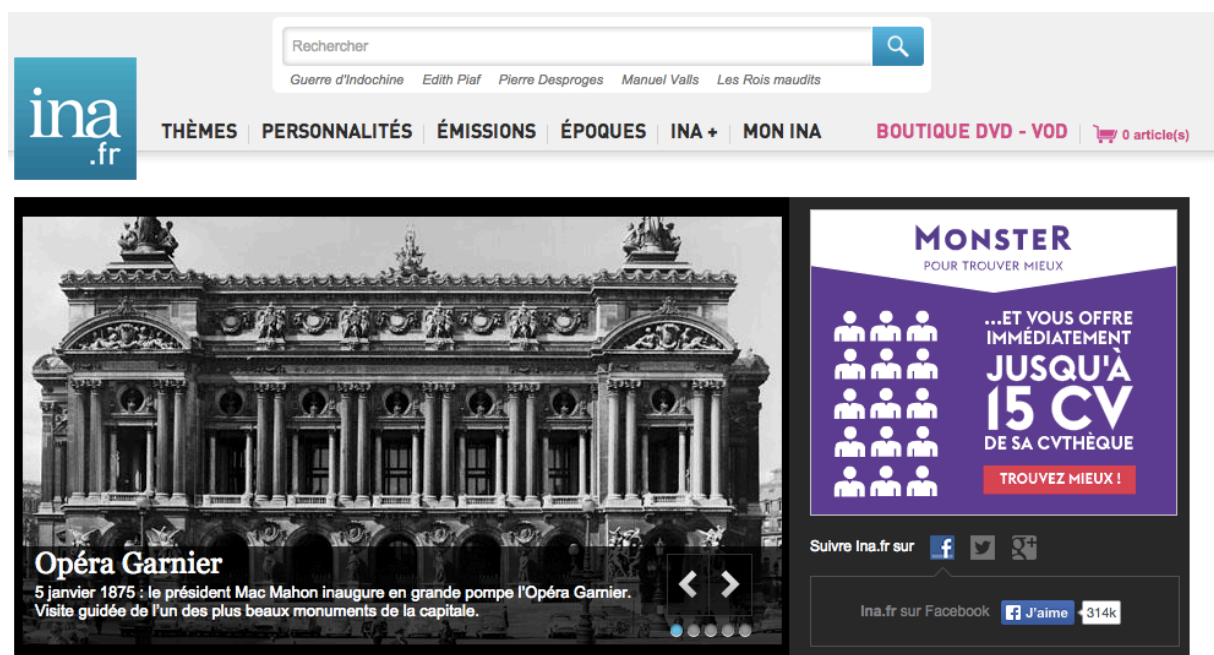


図 2 ina.fr サイト



図 3 NHK デジタルアーカイブス トップページ

## 2.2 放送コンテンツアーカイブの利用と連携

アーカイブは映像や音声などのコンテンツを保存する役割だけでなく、そのコンテンツが利用されることも期待されている。特に放送コンテンツアーカイブは映像に興味のある利用者が閲覧されたり学校教育の場で活用されたりするなど、コンテンツの利用の面からも重要な存在である。放送コンテンツアーカイブが増えていく一方、そのコンテンツを独自の方針で公開・運営しているものが多く、その検索機能が利用者の要求に応えていない場合がある。例えば、テレビで放送された番組を見て、番組の映像そのものや関連する映像を視聴したい利用者がいた場合、テレビ番組のタイトルからアーカイブの検索を行う。テレビ番組に関する情報が紐付けられている放送コンテンツアーカイブならば検索は容易だが、メタデータの不足から目的の映像を探すことが困難なアーカイブも存在する。特にテレビ番組の情報から検索するとき、ひとつのアーカイブに限らず、複数のアーカイブへアクセスして映像を探すことになる。しかし、同じ放送局で作られたアーカイブであってもそれらを横断して、検索することが難しいという問題がある。これは日本における放送されたコンテンツの保存に関する統一的な制度は決まっておらず、放送局の部署毎に Web サイトで放送コンテンツを公開するに留まっていることが原因である。

放送コンテンツアーカイブ検索の具体例として、NHK デジタルアーカイブスについて説明を行う。NHK デジタルアーカイブスではトップページから構成している 6 つのアーカイブをまとめて検索することが可能である。検索方法はフリーワード検索と都道府県名検索の 2 種類を選択することができる。一見、統合的に検索することが可能に見えるが、フリーワード検索の場合、各アーカイブの放送コンテンツのタイトルと概要から全文検索を行っているため、利用者が求める検索結果だけでなく、関心の薄いコンテンツも表示されてしまう。また、NHK のアーカイブの放送コンテンツにはメタデータとして収録されているテレビ番組情報が付与されているが、この放送情報から 6 つのアーカイブを一度にたどることは難しい。図 4 と図 5 は NHK デジタルアーカイブスを構成している「エコチャンネル[5]」と「NHK クリエイティブ・ライブラリー[6]」で公開されているコンテンツのページである。どちらも NHK が放送されたテレビ番組「NHK スペシャル」から抽出された映像であるが、どちらもサイトの構造やテレビ番組のメタデータの記述の仕方が異なるため、関連するコンテンツのページへのアクセスは出来ない上、トップページからの検索で両者を探すことも困難である。

デジタルアーカイブの検索の効率をより高める手法としてデジタルアーカイブの連携がある。デジタルアーカイブの連携とは、コンテンツの内容や管理方法、提供方法が異なるアーカイブ同士のメタデータを収集し、アーカイブ間を横断した資料の関連付けを行うことで関連資料の統合検索を可能にしたサービスとして利用できるようにすることである。アーカイブの連携によって、ひとつのサービスから複数のアーカイブに対して検索、利用できるようになることが求められる。2012 年に総務省が公開した「デジタルアーカイブの構築・連携のためのガイドライン[7]」によるとデジタルアーカイブを連携させることで、統合的な検索を実現し、デジタル資料の流通性を高めることで、知的資産の保存が図れるのみならず、新しい利用方法を生み出す効果があると述べている。

放送コンテンツアーカイブも関連するコンテンツ同士を関連付けさせることで利用者にとって使いやすいサービスとなることが求められる。しかし、部署毎に放送コンテンツを公開している現状では同じ放送局であっても連携することが難しい。デジタルアーカイブの連携を実現するには、デジタルアーカイブそのものの構築から運営方針までいくつかの異なる視点から問題を解決していく必要がある。その問題の 1 つはデジタルアーカイブのメタデータの相互運用性を高めることである。同じテレビ番組から作成されたコンテンツが異なるアーカイブで公開されていても、収録番組のメタデータの値が「NHK ニュース」や「NHK ニュース 2015 年 8 月放送」と異なっている場合が存在する。この場合、テレビ番組に関する語彙が統一されていないために関連するコンテンツであってもリンクすることができなくなっている。デジタルアーカイブ同士を連携させるためには、メタデータ

の記述規則を統一する必要がある。

図 6 はデジタルアーカイブの連携をイメージしたものである。デジタルアーカイブが公開しているコンテンツのメタデータを収集し、統一されたメタデータ記述規則に沿って記述することでシステムは 3 つのアーカイブ間を横断した関連資料の統合検索することが可能となる。メタデータを記述する技術として Linked Open Data (LOD) が用いられる。LOD は定められた定義に則って記述された様々な情報源のデータがリンクされたデータ集合もしくはそのための技術のことである。LOD として記述されたデータは相互運用性を持ち、他の LOD と結びつくことが容易となる。[8]

現状の放送コンテンツアーカイブは、詳しいメタデータが記述されているものの記述規則がアーカイブ毎に異なるため、アーカイブ同士の連携を行い、関連する資料を横断して検索することは難しい。共通する情報であるテレビ放送の情報は同じ放送局であっても、統一したメタデータは付与されておらず、文字列の記述形式の違いからそのままではリンクすることが困難である。本研究では複数のアーカイブのメタデータを蓄積し、統合的な検索を可能にするシステムをメタアーカイブシステムと定義する。既存のアーカイブが所蔵しているコンテンツのメタデータを LOD として記述することで、複数のアーカイブに対して統合的な検索を行えるようにする。LOD に則って記述され、統一されたメタデータは本来異なるアーカイブのコンテンツであっても、コンテンツ同士の関連性を生み出し、検索しやすくなる。また、外部 LOD とのリンクを行い、アーカイブのコンテンツと関連する情報を利用者に提示することも可能となる。メタアーカイブの構築のためには、複数のアーカイブを連携するためのメタデータモデルが必要となる。運営する機関や作成者が異なる以上、アーカイブ毎にメタデータの記述規則も異なっている場合が多いからである。本研究では、放送コンテンツアーカイブを対象としたメタアーカイブの構築を行うために、独自のメタデータモデルを提案する。

次章では既存の放送コンテンツアーカイブを対象にメタデータの収集と分析を行い、それらを連携する LOD を用いたメタデータモデルを定義した。



図 4 エコチャンネルの Web ページ

([http://cgi4.nhk.or.jp/eco-channel/jp/movie/play.cgi?did=D0013770773\\_00000](http://cgi4.nhk.or.jp/eco-channel/jp/movie/play.cgi?did=D0013770773_00000))





図 5 NHK クリエイティブ・ライブラリーの Web ページ  
([http://www1.nhk.or.jp/creative/material/02/D0002040075\\_00000.html](http://www1.nhk.or.jp/creative/material/02/D0002040075_00000.html))

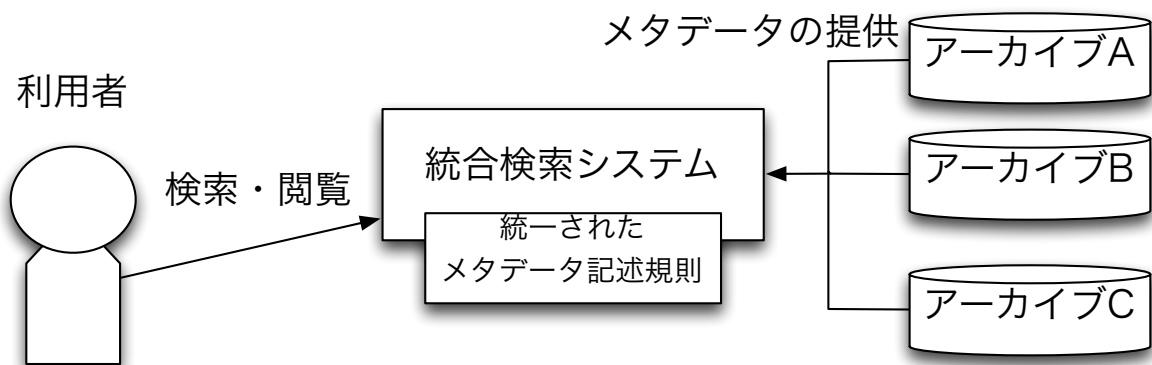


図 6 デジタルアーカイブの連携図

### 3. Linked Open Data を利用したメタアーカイブの構築

3 章では Linked Open Data (LOD) を用いた放送コンテンツアーカイブの連携手法としてメタアーカイブの構築と構築を行うための方法について述べる。

#### 3.1 LOD を利用した放送コンテンツのメタアーカイブ

本研究では、既存の放送コンテンツのメタデータを記述してアーカイブ同士の連携を行う。作成したメタデータを集積して、複数のアーカイブのコンテンツに対して検索とアクセスを提供するシステムをメタアーカイブと定義する。図 7 はメタアーカイブシステムのシステム概要図である。既存の放送コンテンツアーカイブからコンテンツのメタデータを収集し、それらを連携可能にするメタデータモデルの定義を行う。異なるアーカイブのコンテンツであっても、記述規則が統一されたメタデータによって関連するコンテンツの検索が容易となる。本節では、メタアーカイブの構築に必要なメタデータ記述のための技術と手法について述べる。

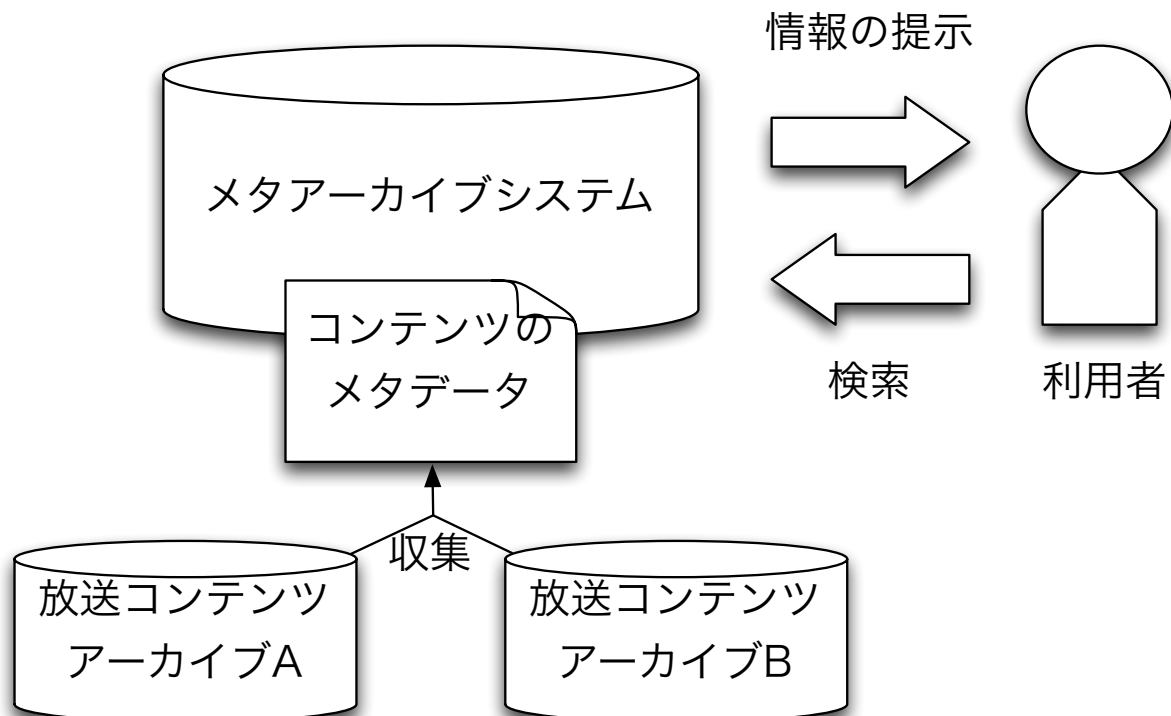


図 7 メタアーカイブシステム概要図

放送コンテンツを利用する際に、利用者が求めている情報とは関連の薄い映像が検索されてしまう、関連するコンテンツが表示されず検索しにくいといった問題を解決するためにメタデータの記述にセマンティックウェブの技術を利用する。セマンティックウェブでは、ウェブ上のデータをより意味的な情報として構築することで、コンピュータがデータの意味を理解し、人間が求める情報を探し出すことが容易となる。異なるアーカイブにある放送コンテンツを、セマンティックウェブの技術に沿って意味的にそのメタデータを記述することで、関連するコンテンツが探し出せるようになると思う。

初期のウェブは HTML 文書がハイパーリンクによって結び付けられている文書のウェブであった。それに対して、セマンティックウェブは WWW 上の情報の意味を機械が理解し、人間が情報を探索・利用することを支援することを目的としたウェブである。セマンティックウェブの実現に向けた試み



のひとつが Linked Data である。Linked Data は Tim Berners-Lee が 2006 年に提唱した、以下の原則を基に作成される[9]。

1. *Use URIs as names for things.*
2. *Use HTTP URIs so that people can look up those names.*
3. *When someone looks up a URL, provide useful information, using the standards (RDF, SPARQL)*
4. *Include links to other URIs, so that they can discover more things.*

「Linked Data – Design Issues」より引用

Linked Data はこの 4 つの原則に従って作成され、ある情報資源について記述されたメタデータの意味を機械が読み取り、他の情報資源との関連を理解することが可能となる。Linked Data は必ずしもウェブ上でデータが公開され、誰でも自由に使えるようにする必要はなく、メタデータ作成者が自ら、もしくは所属する組織内でのみ利用するために作成する場合がある。一方、誰でも自由に使えるように公開している Linked Data を Linked Open Data (LOD) と呼ぶ。LOD として公開されているデータセットは参照方法が標準化されており、誰でも自由に取得・編集することができる。分野の異なるデータセットでも LOD の技術を用いてリンクすることでより多くの情報を引き出せる。LOD として公開されているデータセットの例として、Wikipedia の情報を LOD 化した DBpedia[10]や国立国会図書館が作成した典拠データである国立国会図書館件名標目表 (NDLSH) [11]が挙げられる。

本研究では放送コンテンツのメタデータを LOD として作成するために Resource Description Framework (RDF) のグラフ構造で記述する。RDF は Web 上のリソースに関する情報を明瞭かつ論理的に表現するデータモデルであり、それを記述する言語体系である[12]。RDF での記述により、メタデータ内のリソースとリソースの関係を主語 (リソース)・述語 (プロパティ)・目的語 (プロパティの値) のトリプルで構成される 1 つの文で表現することができる。図 8 は「URI “http://example.com/archive1”で識別されるリソースのタイトルはデジタルアーカイブ A である」というメタデータを表現した RDF グラフである。図中の楕円はリソース、矢印はプロパティ、目的語となる長方形は値がリテラルであることを示している。

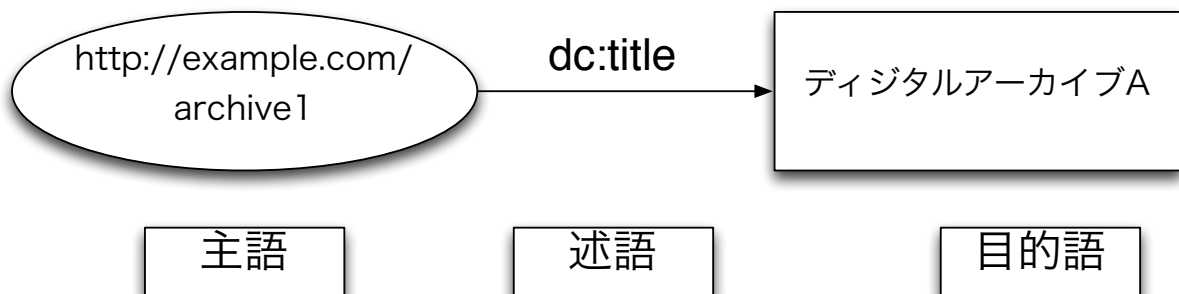


図 8 RDF グラフ 表現例

### 3.2 放送情報と主題情報のメタデータの整備

異なるアーカイブのメタデータの記述規則を統一する場合、放送コンテンツがどのような映像であるかを示す放送情報や主題情報を整備する必要がある。放送情報とは、映像や音声といった放送コンテンツがどのテレビ・ラジオ番組から収録されたメタデータである。同じテレビ番組で収録されたコンテンツでも、放送年月日毎に番組内容が異なるため同じリソースのメタデータとして扱うことができない。特に既存の放送コンテンツアーカイブでは番組情報は基本的に文字列であり、同じテレビ番組であっても全角半角などの表記揺れがあるためそのままの構造では統一することが難しい。図9は放送情報と放送コンテンツアーカイブの関係を図で示したものである[13]。本来、同じテレビ番組シリーズ（ここでは放送年月日は異なるが、放送番組名が同じものを指す）で収録されたコンテンツであるが、異なるアーカイブで公開されているので、放送番組名の表記が異なるため、そのままでは同じテレビ番組シリーズをメタデータに持っていない、その関連性を示すことができない。本研究ではこの放送に関する情報を整備して、LODとして利用しやすい形にすることで複数のアーカイブに対して放送情報から関連するコンテンツの検索を可能にする。例えば、「似通ったジャンルの番組の映像」や「同時期に放送された番組の映像」といった検索を行うことができれば、利用者は放送情報から関連するコンテンツへのアクセスが容易になる。

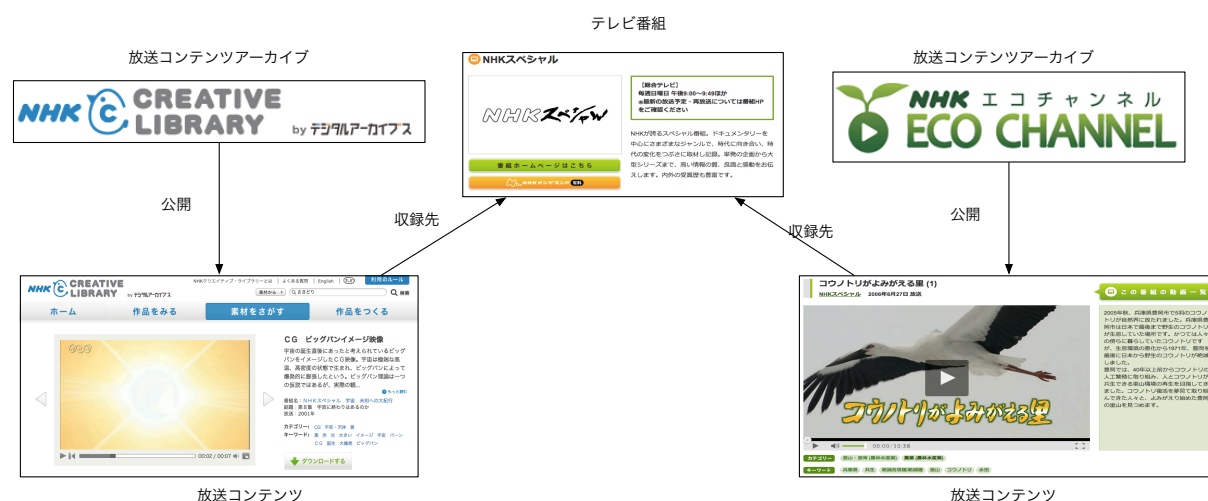


図9 放送情報とアーカイブの放送コンテンツの関係

主題情報とは、放送コンテンツに付与されている、検索や利用のために付与されたメタデータである。アーカイブによって項目名は異なるが、キーワードやジャンル、テーマといったメタデータがそれにあたる（図10参照）。アーカイブ同士を連携するメタアーカイブを構築する際、キーワードやジャンルなどアーカイブ毎に項目名や値が異なるため、メタアーカイブに共通する主題情報を整理する必要がある。

本研究では放送コンテンツアーカイブから収集したメタデータから主題情報を示すものを集積し、分類を行うことでメタアーカイブ統制語彙としてリソースを作成する。

作成する放送情報と主題情報はアーカイブ間での意味を統一するために一般的な単語とその意味を持つ外部 LOD リソースを用いる。本研究では、作成するメタデータと結び付けるリソースとして日本語版 Wikipedia を LOD 化した DBpedia Japanese[14]と国立国会図書館が公開している件名標目表を LOD 化した NDLSH を選択した。DBpedia は単語そのもののページを1つのリソースとして

LOD 化しており、その単語と本研究の放送情報や主題情報を結び付けることで、関連情報が辿りやすくなる。放送情報の場合、DBpedia へのリンクを付与できるメタデータとして、テレビ番組名・出演者・制作者が挙げられる。放送コンテンツから番組情報と関連する情報を取得することが可能になる。主題情報には DBpedia と NDL SH へのリンクを付与する。この手法を用いて、アーカイブ毎に異なる主題情報と言葉の意味が一般化されている LOD と紐付けることでメタアーカイブ内での主題情報の意味を統一する。

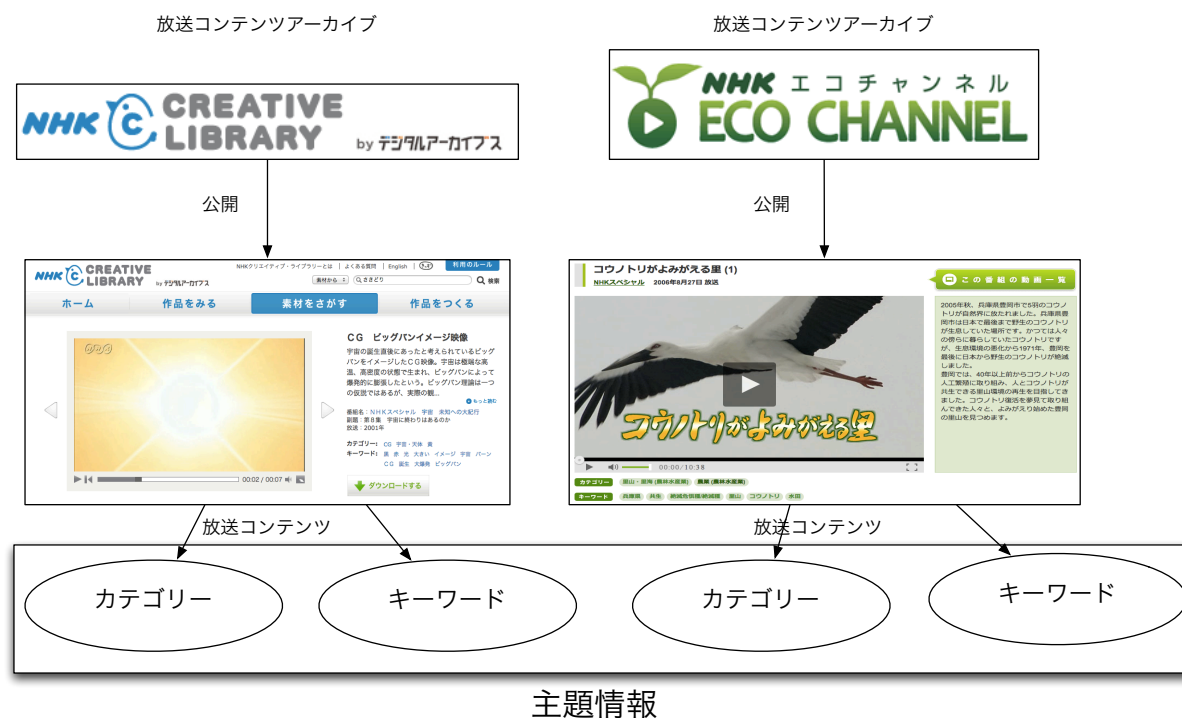


図 10 主題情報とアーカイブの放送コンテンツの関係

### 3.3 関連研究

デジタルアーカイブにおける統合検索の先行事例として、ヨーロッパの図書館、博物館、美術館、文書館等が持つデジタル資料を連携して提供する Europeana[15]がある（図 11 参照）。Europeana では絵画や書籍、映画といった異なるメディアのコンテンツを独自のメタデータモデルである Europeana Data Model (EDM) で管理・統一している。Europeana では各所蔵機関からメタデータの提供を受けて、EDM に従ってメタデータを変換し、関連する資料の紐付けを行っている。図 12 は EDM の規則に基づいた RDF グラフで、ふたつの所蔵機関が持っている「モナ・リザ」に関する所蔵品の関係を記述したものである[16]。EDM では所蔵品そのものに URI を付与し、`edm:ProvidedCHO` クラスとして扱う。「フランス美術総局」と「ルーヴル美術館」では「モナ・リザ」に関する資料を所蔵しており、そのメタデータの集合体を `ore:Aggregation` クラスとし、所蔵機関毎に異なるメタデータを記述している。集合体のクラスは `edm:ProvidedCHO` クラスに対して、プロパティ `edm:aggregatedCHO` でその関連性を表している。その結果、異なる機関で作成された「モナ・リザ」に関する所蔵品はメタデータによってその関連性が記述されることで、コンテンツの統合検索を容易にしている。

日本におけるデジタルアーカイブ連携の事例として、2013 年に公開された国立国会図書館東日本大震災アーカイブ（通称：ひなぎく）[17]（図 13 参照）がある。ひなぎくは東日本大震災に関する音声・動画・写真・ウェブ情報等のデジタルデータや、関連する文献情報を統合的に検索できるポータルサイトである。2014 年 10 月 15 日時点で 39 のデータベースを対象とした横断検索が可能であり、各デジタル資料のメタデータは国立国会図書館ダブリンコアメタデータ記述[18]を元に作成され、統一されている。

このようにデジタルアーカイブの連携は海外で積極的に行われており、日本もまた総務省がガイドラインを公開して連携を行うよう呼びかけている。しかしながら、日本の放送コンテンツアーカイブは他のアーカイブや情報資源と連携されておらず、独自にコンテンツを収集し、公開している現状がある。上記の事例でも述べたように、デジタルアーカイブの連携のためにはコンテンツのメタデータの収集を行い、記述規則を統一するためのメタデータモデルの定義が必要だと考えられる。

EDM の場合、ひとつの美術品の概念を URI として付与することで異なる所蔵機関のメタデータを結びつけた。これにより、現実世界の美術品とインターネット上で公開されているデジタル化された同じ美術品を関連する資料として表現することができる。一方、放送コンテンツアーカイブで公開されている映像や音声はテレビで放送された映像の一部であり、映像ひとつひとつはアーカイブ毎に独自に抽出されたものである。アーカイブ毎に映像には識別子が付与されているものの同一の映像は存在しないため、あるコンテンツを概念として捉え、その関連する資料同士を集合体として、その関係を記述する EDM では放送コンテンツに対して適切ではないと考察する。アーカイブが所蔵している放送コンテンツの元の情報源はテレビ番組の映像である。そのため、アーカイブにある放送コンテンツとテレビ番組の関係を記述するメタデータモデルが必要である。本研究では Europeana と同様に異なるアーカイブのコンテンツを統合検索できるようにすることを目的としているが、メタアーカイブの構築に用いるメタデータモデルは放送コンテンツアーカイブに適した形を提案する。EDM では `Aggregation` クラスを用いて、美術品とそれに関連する情報をリンクしているが、本研究で提案するメタデータモデルでは共通する情報であるテレビ番組情報と主題情報の整理を行うことで関連資料の結びつけを行っている点が異なる。

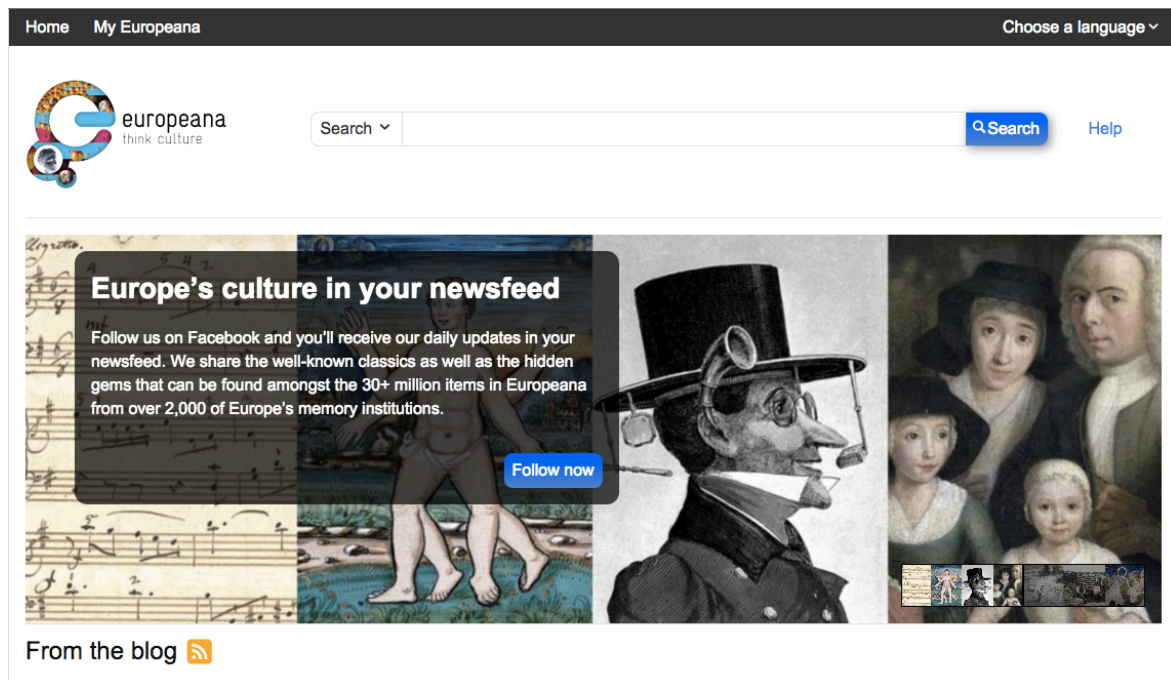


図 11 Europeana トップページ (<http://www.europeana.eu/>)

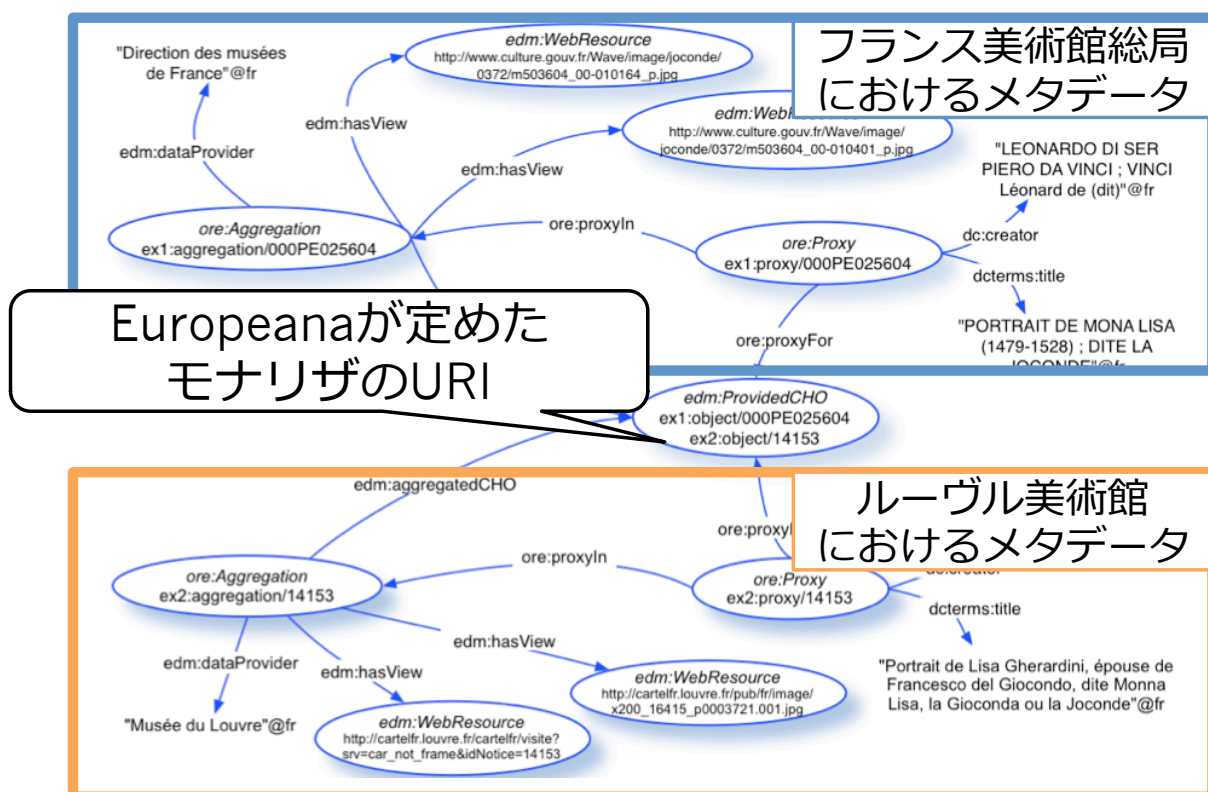


図 12 Europeana Data Model の RDF グラフ



図 13 国立国会図書館東日本大震災アーカイブ トップページ (<http://kn.ndl.go.jp/>)



## 4. 放送コンテンツアーカイブ連携のためのメタデータモデル

3 章では放送コンテンツのメタデータを用いて、複数のアーカイブを連携したメタアーカイブシステムについて述べた。本章では連携のための RDF に基づいたメタデータモデルを説明する。

### 4.1 放送コンテンツのメタデータモデルの定義

提案するメタデータモデルでは、放送コンテンツアーカイブとその放送コンテンツをリソースとしてメタデータを記述する。アーカイブとコンテンツの関係を RDF グラフで記述したものが図 14 である。放送コンテンツアーカイブは複数の放送コンテンツから構成されているリソース、放送コンテンツ自体もリソースであるとしてメタデータを記述する。ウェブ上にある放送コンテンツのメタデータは基本的に文字列で表記されている（図 15 参照）。そのため、既存のメタデータの構造では関連するコンテンツ同士を結び付けることが難しい。

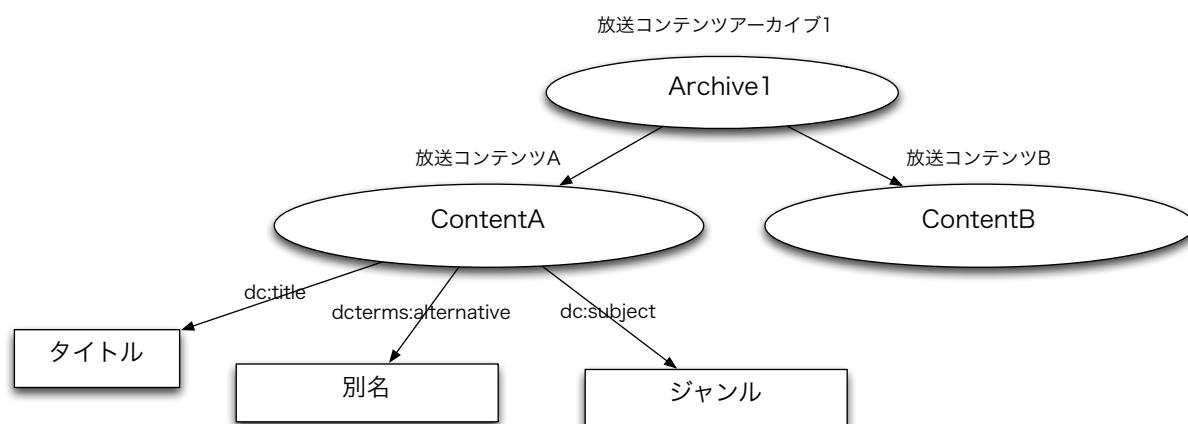


図 14 既存の放送コンテンツとアーカイブのメタデータの構造を表現した RDF グラフ

タイトル・テレビ番組名・放送年月日

コウノトリがよみがえる里 (1)  
NHKスペシャル 2006年8月27日 放送

概要

2005年秋、兵庫県豊岡市で5羽のコウノトリが自然界に放たれました。兵庫県豊岡市は日本で最後まで野生のコウノトリが生息していた場所です。かつては人々の傍らに暮らしていたコウノトリですが、生息環境の悪化から1971年、豊岡を最後に日本から野生のコウノトリが絶滅しました。豊岡では、40年以上前からコウノトリの人工繁殖に取り組み、人とコウノトリが共生できる里山環境の再生を目指してきました。コウノトリ復活を夢見て取り組んできた人々と、よみがえり始めた豊岡の里山を見つめます。

カテゴリー・キーワード

カテゴリー 里山・里海(農林水産省) 農業(農林水産省)

キーワード 兵庫県 共生 絶滅危惧種/絶滅種 里山 コウノトリ 水田

図 15 既存の放送コンテンツのウェブページとそのメタデータ  
([http://cgi4.nhk.or.jp/eco-channel/jp/movie/play.cgi?did=D0013773004\\_00000](http://cgi4.nhk.or.jp/eco-channel/jp/movie/play.cgi?did=D0013773004_00000))

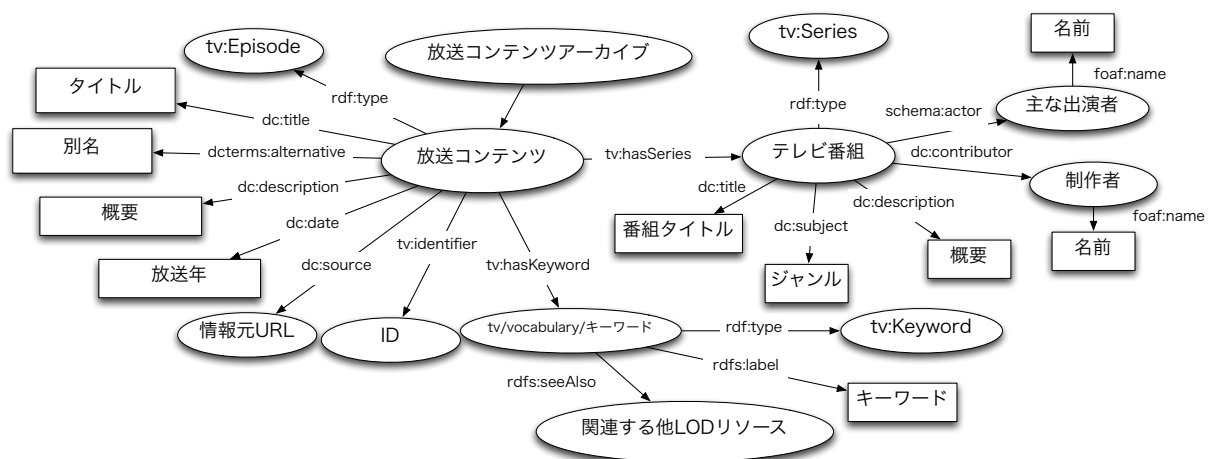


図 16 本研究で提案するメタデータモデル

図 16 は本研究で定義するメタデータの RDF モデルの全体像である。メタデータモデルを定義する際の要件として、異なる放送コンテンツアーカイブのメタデータを統合することが挙げられる。2 章で述べたように、放送コンテンツのアーカイブはメタデータ項目が明確に定められておらず、同じ放送局であっても記述項目や項目値が異なっている。そのため、放送コンテンツアーカイブで定義されているメタデータを分析し、どのアーカイブにも存在する必須の項目が何であるか調査する必要がある。このとき、メタデータの記述項目は代表的な放送コンテンツアーカイブである NHK デジタルアーカイブスで定義されているメタデータに対して、調査を行った。その結果、放送コンテンツに共通する必須の項目は、タイトルとサブタイトル（タイトルの別名）、概要、放送年、コンテンツの取得元 URL、コンテンツの ID、キーワード、テレビ番組であると分析した。この項目は NHK デジタルアーカイブスとして構成されている放送コンテンツアーカイブにもそれに値するもの、もしくは類似するものがある。このメタデータモデルに基づいたメタデータを記述することで、異なるアーカイブのコンテンツ同士の関連性を生み出し、リンクすることが可能となる。また、放送情報と主題情報はアーカイブ間で記述形式や項目値が異なっているが、図のように本研究で定義する語彙を共通のものとすることで同じテレビ番組やキーワードから統合的に検索することが可能となると考える。

メタデータモデルで定義しているメタデータのプロパティは基本的に Dublin Core (DC) [19]等の LOD のためのメタデータ語彙を用いる。メタデータ語彙とは、メタデータ記述に用いる項目と項目値を明確に定めたものである。図 16 のように放送コンテンツのタイトルは `dc: title` で表すなど、メタデータ語彙の中でも広く使用されているものを優先する。広く使用されているメタデータ語彙を用いることで、LOD として使用されやすくすることも目的としている。放送コンテンツのクラスを指すためのメタデータ語彙は本研究独自の語彙として `tv` (名前空間:[http://purl.org/net/tv\\_schema/ns#](http://purl.org/net/tv_schema/ns#))を設定し、表現する。表 1 は放送コンテンツが所持する主なメタデータとそれに適したプロパティとその値について定義したものである (使用するメタデータ語彙は表 2 参照)。メタデータを記述する際、表 1 に該当するメタデータがある場合、その定義に沿うこととする。放送コンテンツのクラスは `tv: Episode` として表現するが、放送情報と放送コンテンツの関係に関する説明は次節で述べる。

本研究で提案するメタアーカイブでは表 1 の定義をもとにメタデータを記述していくが、アーカイブ毎にメタデータの記述項目や構造が異なる場合が多い。アーカイブを管理している団体が異なるため、そのメタデータの定義や項目数も一律ではないからである。例えば、あるアーカイブでは放送年月日などの時間情報に関するメタデータを具体的に記述しているが、別のアーカイブでは地名などの



地理情報のメタデータが具体的に記述されている場合もある。これはアーカイブがある一定のテーマ（地理や郷土に関する映像や年代別の映像）に沿って作成されている関係上、適切なメタデータが異なることが理由である。異なるアーカイブを連携するために、表 1 で定義されているメタデータのみを記述してしまうと、元のデータと差異を生み出してしまい、本来のアーカイブで付与されたメタデータの欠落や項目の意味の解釈が変わる恐れが出てくる。本研究で定義するメタデータモデルでは、主要なメタデータに関しては表 2 に準じた語彙や値制約を適用するが、アーカイブ毎に定義されているメタデータ項目をそのまま保持するために、足りない項目に関しては新たにメタデータ語彙を使用して記述することとする。

表 1 基準となる放送コンテンツのメタデータ語彙と記述規則

メタデータ	プロパティ	値タイプ	値制約
URI	tv:Episode	ID	<a href="http://purl.org/net/tv/episode/">http://purl.org/net/tv/episode/</a>
URL(情報元)	dc:identifier	参照値	
タイトル	dc:title	文字列	
別名	dcterms:alternative	文字列	
概要	dc:description	文字列	
放送年	dc:date	文字列	YYYY-MM-DD
ID	tv:identifier	文字列	
カテゴリー	tv:hasCategory	構造化	#構造化カテゴリー
キーワード	tv:hasKeyword	構造化	#構造化キーワード
テレビ番組	tv:hasSeries	参照値	<a href="http://purl.org/net/tv/series/">http://purl.org/net/tv/series/</a>

表 2 放送コンテンツのメタデータで使用する名前空間接頭辞

[@NS]	
dc	<a href="http://purl.org/dc/elements/1.1/">http://purl.org/dc/elements/1.1/</a>
dcterms	<a href="http://purl.org/dc/terms/">http://purl.org/dc/terms/</a>
tv	<a href="http://purl.org/net/tv/ns#">http://purl.org/net/tv/ns#</a>

## 4.2 放送情報の構造と LOD 化

本節では、放送コンテンツ共通の情報となる放送情報のメタデータモデルについて説明を行う。放送情報の中でもテレビ番組の情報を基準にメタデータを記述する。ここでいうテレビ番組とは、テレビで放送された番組のことであり、そのメタデータは番組名、放送時間、放送期間、ジャンル、出演者、制作者などが挙げられる。

テレビ番組はその番組ごとにリソースが異なるが、同じテレビ番組であってもリソースが同じではない。テレビ放送は一定の時間毎にテレビで放送されるが、その番組内容は放送されて年月日によって異なる。例えば、同じニュース番組であっても 2015 年 1 月 1 日に放送された内容と 2015 年 1 月 2 日に放送された内容は異なっている。番組の構成や番組名、ニュースキャスターといった出演者などの基本的な情報は放送日時によらず同じであるものの、ニュースの内容や出演ゲストは違うため同じリソースとして表現することは難しい。

本研究で提案する番組情報のメタデータモデルは、番組名や放送期間、ジャンル、メインキャストなどのテレビ番組の基本的な情報をシリーズクラスと定義し、放送内容や放送日時など放送毎に異なる映像情報をエピソードクラスと定義する。図 17 は NHK のテレビ番組「連続テレビ小説 あまちゃん」のメタデータを RDF グラフで表現したものである。番組シリーズを表すリソースとそのタイトル (dc:title)、別名 (dcterms:alternative)、ジャンル (dc:subject)、概要 (dc:description) のリソースが RDF トリプルで記述されている。シリーズクラスとエピソードクラスは 1 対 N の関係を持ち、その関係は tv:hasSeries のプロパティで表現する。

テレビ番組のシリーズは表 3・表 4 のメタデータ記述規則を基に記述する。表 3・4 は簡易記述セットプロファイル (DSP) と呼ばれるメタデータ記述基盤構築事業が提案している書式で、メタデータの記述対象と記述項目を定めたものである。表 3 はテレビ番組のシリーズのメタデータ記述項目と規則を表したもので、表 4 にはテレビ番組シリーズの出演者と制作者のメタデータを構造化メタデータで表したものである。表 5 はメタデータ記述規則に使用されているメタデータ語彙の接頭辞である。基本的に既存のメタデータ語彙を使用しているが、シリーズとその ID には本研究で設定したメタデータ語彙 tv を使用する。

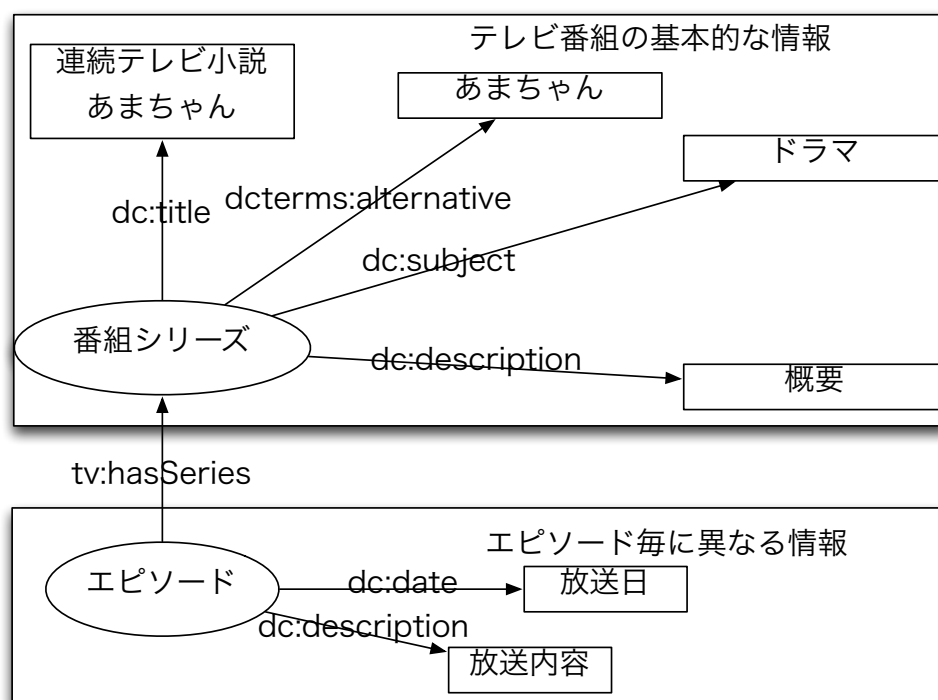


図 17 テレビ番組の構造を表した RDF グラフ

エピソードクラスにはアーカイブにおける放送コンテンツのリソースが当てはまる。放送コンテンツは放送されたテレビ番組の映像から収録されたものから構成されており、ある時点で放送されたエピソードと同一であると考えられる。図 18 は放送コンテンツアーカイブとテレビ番組情報の関係を RDF グラフで記述したものである。放送コンテンツ Content A はテレビ番組シリーズ A のエピソードであるので、`rdf:type` のプロパティでエピソードクラスとして表現する。放送コンテンツをエピソードとみなすことで、テレビ番組シリーズを通じて、他のアーカイブの放送コンテンツと連携することが可能となる。

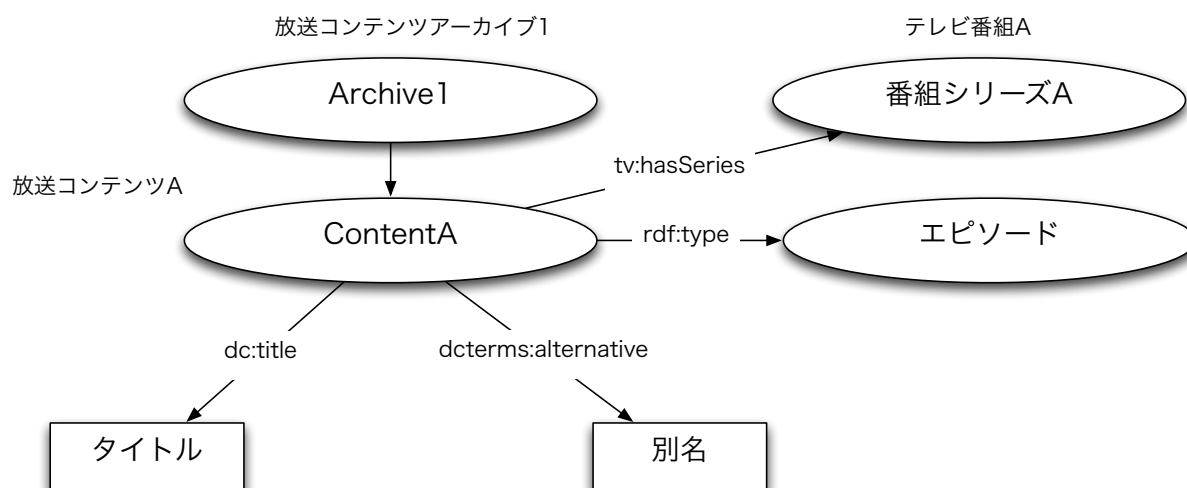


図 18 放送コンテンツアーカイブとテレビ番組情報の関係を記述した RDF グラフ

表 3 テレビ番組シリーズのメタデータ記述構造（簡易 DSP）

[MAIN]						
#記述項目名	プロパティ	最小	最大	値タイプ	値制約	説明
シリーズ	tv:Series	1	1	ID	<http://purl.org/net/tv/series/>	テレビ番組のシリーズクラス
タイトル	dc:title	1	1	文字列		シリーズタイトル
タイトルの一般名	dcterms:alternative	0	1	文字列		タイトルの別名・略名
概要	dc:description	1	1	文字列		テレビ番組の説明文や概要を記述
シリーズID	tv:identifier	1	1	ID		テレビ番組シリーズのIDを設定
情報源	dc:identifier	1	1	参照値		情報の取得元
放送開始年	dc:date	1	1	文字列	YYYY	テレビ番組が放送された年
放送期間	dcterms:temporal	1	1	文字列	YYYY/YYYY	テレビ番組が放送された期間
ジャンル	dc:subject	1	1	文字列		テレビ番組のジャンル
外部リンク	rdfs:seeAlso	0	1	参照値	<http://ja.dbpedia.org/resource/>	他のLODへのリンク
出演者	schema:actor	0	1	構造化	#構造化出演者	テレビ番組の出演者
制作者	dc:contributor	0	1	構造化	#構造化制作者	テレビ番組の制作者
制作元	dc:creator	1	1	文字列		番組を放送したテレビ局

表 4 テレビ番組シリーズの出演者・制作者の構造化メタデータ（簡易 DSP）

[構造出演者]						
#記述項目名	プロパティ	最小	最大	値タイプ	値制約	説明
出演者	foaf:Person	1	1	ID	<http://purl.org/net/tv/person/>	
名前	foaf:name	1	1	文字列		出演者の名前
外部リンク	rdfs:seeAlso	0	1	参照値	<http://ja.dbpedia.org/resource/>	出演者の情報を提供する他LODリソース
[構造化制作者]						
#記述項目名	プロパティ	最小	最大	値タイプ	値制約	説明
制作者	foaf:Person	1	1	ID	<http://purl.org/net/tv/person/>	
名前	foaf:name	1	1	文字列		制作者の名前
外部リンク	rdfs:seeAlso	0	1	参照値	<http://ja.dbpedia.org/resource/>	制作者の情報を提供する他LODリソース

表 5 テレビ番組で使用する名前空間接頭辞

[@NS]	
rdfs	http://www.w3.org/2000/01/rdf-schema#
dc	http://purl.org/dc/elements/1.1/
dcterms	http://purl.org/dc/terms/
foaf	http://xmlns.com/foaf/0.1/
schema	http://schema.org/
tv	http://purl.org/net/tv/ns#

### 4.3 放送コンテンツの主題情報の統一

3.2 節では、アーカイブ毎にキーワードやジャンルといった放送コンテンツの主題情報が異なるため、メタアーカイブを構築する際に共通する主題情報を整理する必要があると述べた。本節では収集した主題情報を統合し、アーカイブ間で連携がしやすいようにメタデータを構造化した主題情報のモデルについて述べる。

放送コンテンツアーカイブに付与されている主題情報は基本的に「観光」や「世界遺産」などの文字列である。提案するメタデータモデルは主題情報をリソースとして URI を与えてその構造化を行い、他 LOD リソースへのリンクを付与する。図 19 は放送コンテンツとその主題情報を RDF グラフで示したものである。放送コンテンツ A のキーワード「群馬県」はプロパティ `tv:hasKeyword` で結び付け、リソースは `tv:Keyword` クラスとする。キーワードのリソースには `dc:title` でキーワードの名前を表現する。また、キーワードに URI を付与することでキーワードと関連を持つ他の実体との関係を記述することができる。`rdfs:seeAlso` は主語であるリソースについての情報を提供するリソースを表現するプロパティであり、DBpedia や NDLSH のリソースへのリンクを記述する。

主題情報の対象となるメタデータは放送コンテンツのキーワードやジャンル、カテゴリーなどが挙げられる。この主題情報となるメタデータはメタアーカイブの対象となる複数のアーカイブから取得し、重複や表記揺れの修正を行い、メタデータモデルに当てはめる。4.1 節でアーカイブのメタデータ項目はそのまま保持すると述べたが、主題情報を表すメタデータは本節で提案するモデルを適用し、主題情報の値は元のまま構造化を行う。

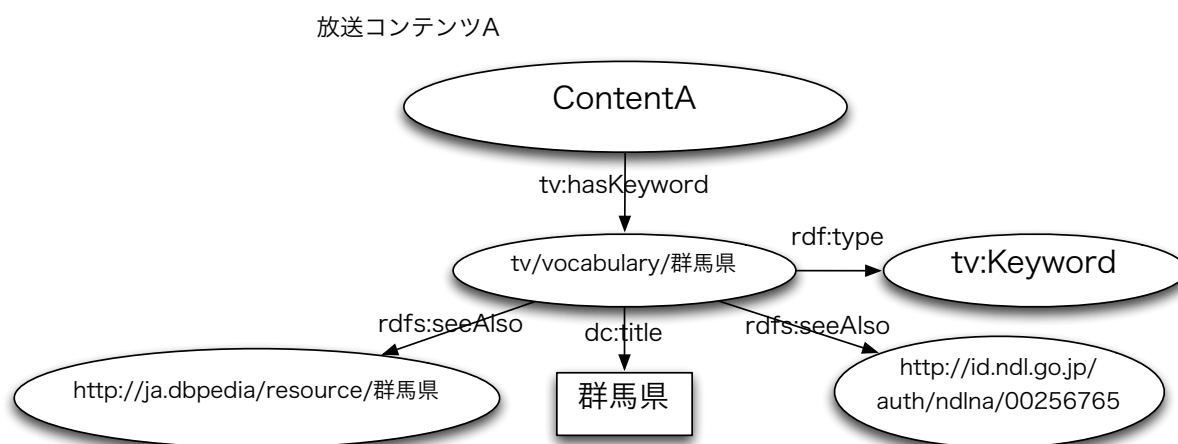


図 19 放送コンテンツと主題情報を記述した RDF グラフ

## 5. メタアーカイブ構築のためのメタデータ作成手法

本章では、Web 上にあるデジタルアーカイブのメタデータを収集し、RDF ファイルへと変換するための手法について述べる。

### 5.1 放送コンテンツのメタデータ取得と変換手法

4 章で述べたメタデータモデルから実際にメタデータの作成を行うとき、以下の手順を行う。

1. デジタルアーカイブから放送コンテンツのメタデータを取得する
2. 取得したメタデータを RDF ファイルに変換する

1 の手順を実行する際に、幾つかの問題を解決する必要がある。1 つ目はデジタルアーカイブのコンテンツのデータを API で取得できるものが少ないという問題である。そのため、メタデータを Web ページから取得しなければならないが、デジタルアーカイブのコンテンツ数は数百件から数千件にも上る。Web ページを閲覧して、データを入力する手作業は時間と労力の関係から現実的ではない。この問題を解決するために、本研究では Web ページからメタデータを自動的に取得するプログラムの作成を行う。その具体的な手法は 5.2 節で説明する。

2 の手順では 1 の手順で取得したデータを RDF ファイルに変換を行うプログラムの作成を行う。このプログラムはメタデータモデルで定義されているメタデータ語彙を選択し、RDF ファイルとして出力する機能を持たせる。このとき、主題情報と放送情報のメタデータには外部 LOD へのリンクを付与する。プログラムと出力する RDF ファイルの詳細については 5.3 節で述べる。

### 5.2 WWW 上にあるアーカイブからのメタデータ取得

放送コンテンツアーカイブのコンテンツのメタデータを取得するために、ウェブスクレイピングを用いる。ウェブスクレイピングはウェブサイトから情報を抽出するコンピュータソフトウェア技術のことである。本節では、WWW 上にあるアーカイブから自動的にメタデータを抽出し、Comma-Separated Value (CSV) フォーマットのファイルとして出力する方法を説明する。

放送コンテンツアーカイブは WWW 上のウェブページとして公開されているため、そのファイルフォーマットは基本的にマークアップ言語である HTML である。そこでウェブスクレイピングのツールとして、Ruby のライブラリーでマークアップ言語を解析できる Nokogiri[20]を用いた。HTML 要素の抽出を可能にする Nokogiri を用いてメタデータを取得する。また、アーカイブは数百や数千にも上るコンテンツを公開しているものもあるため、自動でブラウザを操作しウェブページにアクセスする必要がある。本研究ではブラウザをプログラムで動かすフレームワークである Selenium WebDriver[21]を用いた。必要なメタデータを取得するための手順を以下に示す。

1. 放送コンテンツのページの HTML 構造を理解する
2. 取得したいメタデータが含まれている要素を抽出する XPath 構文を決める
3. 放送コンテンツのページへアクセスしてメタデータの抽出を行い、CSV ファイルとして出力する
4. ブラウザの自動操作を行い、3 の手順を全てのコンテンツに対して行う

1・2 の手順では、実際にウェブブラウザから放送コンテンツアーカイブのウェブページにアクセスを行うことから始める。そして、ウェブページ上に記述されているタイトルや説明文などのメタデ

ータの確認を行う。取得するメタデータが決定したあと、メタデータが含まれている要素の XPath 構文をブラウザのデベロッパーツールから取得する。図 20 は実際にウェブブラウザからエコチャンネルのページ ([http://cgi4.nhk.or.jp/eco-channel/jp/movie/play.cgi?did=D0013773004\\_00000](http://cgi4.nhk.or.jp/eco-channel/jp/movie/play.cgi?did=D0013773004_00000)) からタイトルを含む要素の XPath 構文の取得を行ったときのスクリーンショットである。

3 の手順ではプログラムを用いて、ブラウザを起動して自動で XPath 構文を用いたメタデータの抽出を行う。XPath 構文を用いて取得した情報は空白除去などの文字列整形を行い、メタデータとして使用できる要素にする。取得したメタデータは図 21 のように CSV ファイルとして出力する。図 21 は CSV ファイルで出力しているフィールドは「タイトル,収録番組,放送年月日,概要,カテゴリー,キーワード,URL」である。このフィールドの値は対象とするアーカイブ毎に異なるため、適宜フィールドの要素を変更する必要がある。

4 の手順では 3 の手順をアーカイブの全ての放送コンテンツのページに適用する。アーカイブには動画一覧ページ (図 22 参照) があるため、Nokogiri を用いてコンテンツの URL を取得することが可能となる。Selenium WebDriver を用いて取得した URL から各ウェブページにアクセスし、メタデータを取得する処理を繰り返す。

上記の手順を放送コンテンツアーカイブ毎に適用することで必要なメタデータの取得することができる。

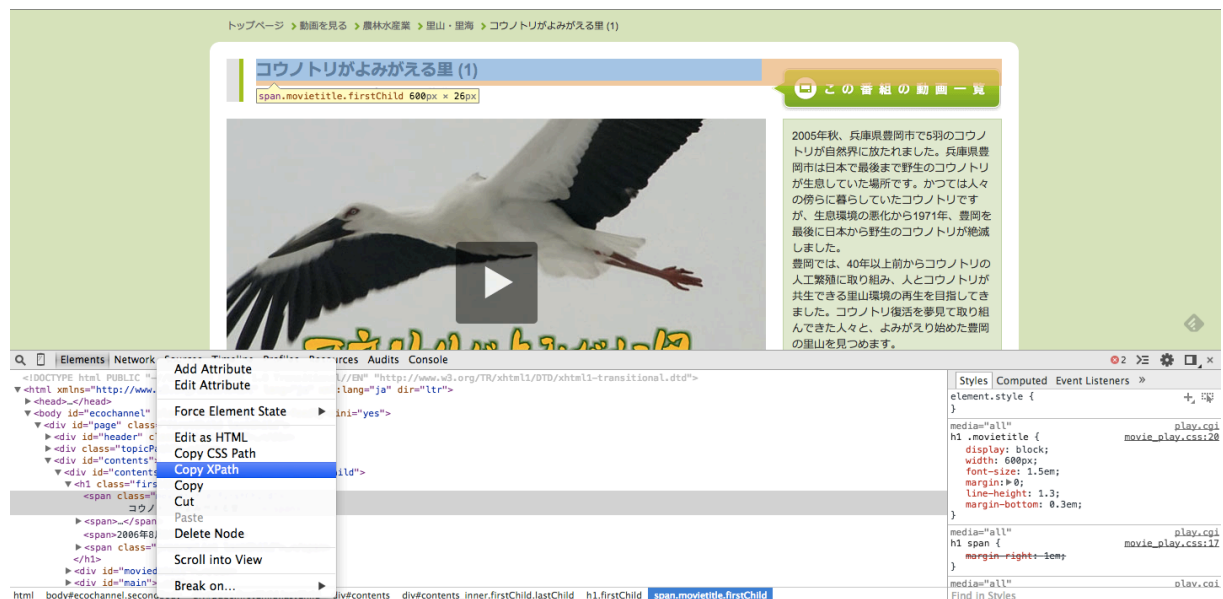


図 20 ウェブブラウザのデベロッパーツールの使用例

コウノトリがよみがえる里 (1),NHKスペシャル,2006年8月27日 放送,2005年秋、兵庫県豊岡市で5羽のコウノトリが自然界に放たれました。兵庫県豊岡市は日本で最後まで野生のコウノトリが生息していた場所です。かつては人々の傍らに暮らしていたコウノトリですが、生息環境の悪化から1971年、豊岡を最後に日本から野生のコウノトリが絶滅しました。豊岡では、40年以上前からコウノトリの人工繁殖に取り組み、人とコウノトリが共生できる里山環境の再生を目指してきました。コウノトリ復活を夢見て取り組んできた人々と、よみがえり始めた豊岡の里山を見つめます。 ,里山・里海 (農林水産業):農業 (農林水産業),兵庫県:共生:絶滅危惧種/絶滅種:里山:コウノトリ:水田,[http://cgi4.nhk.or.jp/eco-channel/jp/movie/play.cgi?did=D0013773004\\_00000](http://cgi4.nhk.or.jp/eco-channel/jp/movie/play.cgi?did=D0013773004_00000)

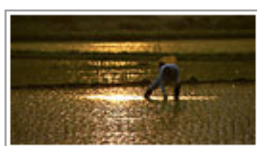
図 21 出力した CSV ファイル

## 関連動画一覧



15:46

コウノトリがよみがえる  
里 (4) (NHKスペシャル)



11:10

コウノトリがよみがえる  
里 (3) (NHKスペシャル)



7:45

コウノトリがよみがえる  
里 (2) (NHKスペシャル)



10:38

コウノトリがよみがえる  
里 (1) (NHKスペシャル)



14:41

熱帯雨林消滅 -問われる  
日本の開発・援助-(2) (NHKスペシャル)



14:03

熱帯雨林消滅 -問われる  
日本の開発・援助-(1) (NHKスペシャル)



4:58

2011 ニッポンの生きる  
道 ノーベル賞 根岸博士  
「ニッポンの英知を結集  
せよ！」 (NHKスペシャル)



3:45

「ホットスポット 最後の  
楽園」 予告 (NHKスペシ  
ヤル)



4:47

自動車革命 次世代カー  
電池をめぐる闘い オープ  
ニング (NHKスペシャル)



3:02

映像詩 里山 ~覚えてい  
ますか ふるさとの風景  
~ オープニング (NHK  
スペシャル)



2:29

映像詩 里山 ~命めぐる  
水辺~ オープニング (NHK  
スペシャル)



3:13

映像詩 里山 ~森と人  
響きあう命~ オープニ  
ング (NHKスペシャル)

図 22 アーカイブの動画一覧ページ



### 5.3 RDF ファイルの記述と外部 LOD へのリンク作成

本節では、5.2 節で出力した CSV ファイルを LOD として利用するために RDF ファイルへと変換する手法を説明する。

本研究では RDF 記述形式として RDF/TURTLE 形式を選択した[22]。TURTLE とは RDF の標準フォーマットのひとつであり、その標準仕様は World Wide Web Consortium (W3C) の勧告となっている。図 23 は RDF/TURTLE で記述された RDF トリプルであり、「URI `http://example.com/arhive1`」で識別されるリソースのタイトルはアーカイブ A である」ことを示している。このフォーマットに則って、放送コンテンツや放送番組のメタデータの RDF ファイルを作成していく。

<code>&lt;http://example.com/arhiveA&gt; &lt;http://purl.org/dc/elements/1.1/title&gt; “アーカイブ A”.</code>
--

図 23 RDF/TURTLE で記述されたトリプル 例

5.1 節で出力したメタデータを具体例として RDF ファイル作成手順の説明を行う。RDF ファイルの作成にあたって、以下のような手順を行った。

まず取得したメタデータをモデルに当てはめ、記述項目と規則を定義する。今回出力したメタデータは 4.1 で定義したメタデータモデルの項目に充足するので、新たにメタデータ語彙を選択し、項目名の追加を行う必要はない（表 7 参照）。

次に CSV ファイルを読み込み、RDF/TURTLE ファイルに変換を行うプログラムを記述する。RDF ファイルを記述するとき、名前空間接頭辞の宣言後、表 6 のフォーマットに沿ったトリプルの記述を行う。プログラム中の処理でメタデータ「キーワード」と「カテゴリー」は主題情報として、他 LOD である DBpedia と NDL SH へのリンクを作成する。

リンクを作成するために RDF クエリ言語 SPARQL[23]を用いてウェブにある LOD データセットへアクセスを行う。SPARQL はグラフパターンをクエリとして記述しグラフパターンに一致するデータの検索を行う。図 24 は DBpedia のデータセットに指定するキーワードをラベルに持つリソースが存在するか問い合わせ、その値を boolean 値で取得するクエリである。DBpedia のリソースの URI は `http://ja.dbpedia.org/resource/〇〇` で表記されているので、true が返ってきた場合は 〇〇 にキーワード名を入れて URI として表現することとする。NDL もまた図 24 のクエリを用いて、boolean 値を取得するが、NDL はラベルが同じリソースを複数もつ場合がある。これは件名標目表の中で標目名と書名が一致しているリソースがあるから・・・である。そのため、DBpedia で用いたような boolean 値を取得するだけではリソースの URI を表現することが出来ない。NDL のリソースの URI は `http://id.ndl.go.jp/auth/ndlsh/00571539` のように ID が付与されているので、URI を取得する必要がある。図 25 は NDL のデータセットから boolean 値を取得するクエリと URI を取得するクエリである。まず boolean 値を取得するクエリを問い合わせたあと、true が返ってきた場合 URI を取得するクエリを問い合わせることで NDL とのリンクを作成する。

図 26 は記述したプログラムを実行して出力したファイルの一部である。キーワードの構造は 4.3 節の主題情報のデータモデルを適用している。本研究では、出力したファイルのフォーマットを基準として放送コンテンツのメタデータの RDF ファイルを記述していくこととする。

表 6 5.1 節で出力したメタデータのプロパティと値

メタデータ	プロパティ	値
URI	tv:Episode	http://purl.org/net/tv/episode/D0013773004_00000
URL	dc:identifier	http://cgi4.nhk.or.jp/eco-channel/jp/movie/play.cgi?did=D0013773004_00000
タイトル	dc:title	コウノトリがよみがえる里 (1)
別名	dcterms:alternative	
概要	dc:description	2005 年秋…を見つめます。
放送年	dc:date	2006-08-27
ID	tv:identifier	D0013773004_00000
カテゴリー	tv:hasCategory	里山・里海 (農林水産業), 農業 (農林水産業)
キーワード	tv:hasKeyword	兵庫県, 共生, 絶滅危惧種/絶滅種, 里山, コウノトリ, 水田
テレビ番組	tv:hasSeries	NHK スペシャル

```
PREFIX dbp: http://ja.dbpedia.org/resource/
PREFIX rdfs: http://www.w3.org/2000/01/rdf-schema#
ASK WHERE{ dbp:#{キーワード} rdfs:label ?o }
```

図 24 キーワードをラベルに持つリソースが DBpedia に存在するか問い合わせるクエリ

```
# キーワードが存在するか問い合わせるクエリ
PREFIX rdfs: http://www.w3.org/2000/01/rdf-schema#
ASK WHERE{ ?o rdfs:label "#{キーワード}" }

#キーワードをラベルに持つリソースの URI を取得するクエリ
PREFIX rdfs: http://www.w3.org/2000/01/rdf-schema#
SELECT ?o WHERE{ ?o rdfs:label "#{キーワード}" }
```

図 25 NDL にキーワードをラベルに持つリソースが存在するか問い合わせるクエリとそのリソースの URI を取得するクエリ

```

@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix dterms: <http://purl.org/dc/terms/> .
@prefix tv: <http://purl.org/net/tv/ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

<http://purl.org/net/tv/episode/D0013773004_00000> rdf:type tv:Episode.
<http://purl.org/net/tv/episode/D0013773004_00000> tv:identifier "D0013773004_00000".
<http://purl.org/net/tv/episode/D0013773004_00000> dc:title "コウノトリがよみがえる里 (1)"@ja.
<http://purl.org/net/tv/episode/D0013773004_00000> dc:description "2005 年秋、兵庫県豊岡市で 5 羽のコウノトリが自然界に放たれました。兵庫県豊岡市は日本で最後まで野生のコウノトリが生息していた場所です。かつては人々の傍らに暮らしていたコウノトリですが、生息環境の悪化から 1971 年、豊岡は最後に日本から野生のコウノトリが絶滅しました。豊岡では、40 年以上前からコウノトリの人工繁殖に取り組み、人とコウノトリが共生できる里山環境の再生を目指してきました。コウノトリ復活を夢見て取り組んできた人々と、よみがえり始めた豊岡の里山を見つめます。"@ja.
<http://purl.org/net/tv/episode/D0013773004_00000> dc:date "2006/08/27".
<http://purl.org/net/tv/episode/D0013773004_00000> dc:source <http://cgi4.nhk.or.jp/eco-channel/jp/movie/play.cgi?did=D0013773004_00000>.
<http://purl.org/net/tv/episode/D0013773004_00000> tv:hasCategory <http://www.nhk.or.jp/tv/vocabulary/里山・里海>.
<http://www.nhk.or.jp/tv/vocabulary/里山・里海> rdfs:label "里山・里海".
<http://www.nhk.or.jp/tv/vocabulary/里山・里海> rdfs:seeAlso <http://ja.dbpedia.org/resource/里山>.
<http://www.nhk.or.jp/tv/vocabulary/里山・里海> rdfs:seeAlso <http://id.ndl.go.jp/auth/ndlsh/01072661>.
<http://purl.org/net/tv/episode/D0013773004_00000> tv:hasCategory <http://www.nhk.or.jp/tv/vocabulary/農業>.
<http://www.nhk.or.jp/tv/vocabulary/農業> rdfs:label "農業".
<http://www.nhk.or.jp/tv/vocabulary/農業> rdfs:seeAlso <http://ja.dbpedia.org/resource/農業>.
<http://www.nhk.or.jp/tv/vocabulary/農業> rdfs:seeAlso <http://id.ndl.go.jp/auth/ndlsh/00568603>.
<http://purl.org/net/tv/episode/D0013773004_00000> tv:hasKeyword <http://www.nhk.or.jp/tv/vocabulary/兵庫県>.
<http://www.nhk.or.jp/tv/vocabulary/兵庫県> rdfs:label "兵庫県".
<http://www.nhk.or.jp/tv/vocabulary/兵庫県> rdfs:seeAlso <http://ja.dbpedia.org/resource/兵庫県>.
<http://www.nhk.or.jp/tv/vocabulary/兵庫県> rdfs:seeAlso <http://id.ndl.go.jp/auth/ndlsh/00259339>.
<http://purl.org/net/tv/episode/D0013773004_00000> tv:hasKeyword <http://www.nhk.or.jp/tv/vocabulary/共生>.
<http://www.nhk.or.jp/tv/vocabulary/共生> rdfs:label "共生".
<http://www.nhk.or.jp/tv/vocabulary/共生> rdfs:seeAlso <http://ja.dbpedia.org/resource/共生>.
<http://purl.org/net/tv/episode/D0013773004_00000> tv:hasKeyword <http://www.nhk.or.jp/tv/vocabulary/絶滅危惧種/絶滅種>.
<http://www.nhk.or.jp/tv/vocabulary/絶滅危惧種/絶滅種> rdfs:label "絶滅危惧種/絶滅種".
<http://purl.org/net/tv/episode/D0013773004_00000> tv:hasKeyword <http://www.nhk.or.jp/tv/vocabulary/里山>.
<http://www.nhk.or.jp/tv/vocabulary/里山> rdfs:label "里山".
<http://www.nhk.or.jp/tv/vocabulary/里山> rdfs:seeAlso <http://ja.dbpedia.org/resource/里山>.
<http://www.nhk.or.jp/tv/vocabulary/里山> rdfs:seeAlso <http://id.ndl.go.jp/auth/ndlsh/01072661>.
<http://purl.org/net/tv/episode/D0013773004_00000> tv:hasKeyword <http://www.nhk.or.jp/tv/vocabulary/コウノトリ>.
<http://www.nhk.or.jp/tv/vocabulary/コウノトリ> rdfs:label "コウノトリ".
<http://www.nhk.or.jp/tv/vocabulary/コウノトリ> rdfs:seeAlso <http://ja.dbpedia.org/resource/コウノトリ>.
<http://www.nhk.or.jp/tv/vocabulary/コウノトリ> rdfs:seeAlso <http://id.ndl.go.jp/auth/ndlsh/00566590>.
<http://purl.org/net/tv/episode/D0013773004_00000> tv:hasKeyword <http://www.nhk.or.jp/tv/vocabulary/水田>.
<http://www.nhk.or.jp/tv/vocabulary/水田> rdfs:label "水田".
<http://www.nhk.or.jp/tv/vocabulary/水田> rdfs:seeAlso <http://ja.dbpedia.org/resource/水田>.
<http://www.nhk.or.jp/tv/vocabulary/水田> rdfs:seeAlso <http://id.ndl.go.jp/auth/ndlsh/00571539>.

```

図 26 RDF/TURTLE 形式で記述された「コウノトリがよみがえる里(1)」のメタデータ

## 6. 適用事例 ～NHK デジタルアーカイブスのメタアーカイブ構築～

本研究では、提案するメタデータモデルをもとに複数のアーカイブを連携したメタアーカイブの構築を目指す。その適用事例として、NHK デジタルアーカイブスが保有する放送コンテンツに対してメタデータの LOD 化を行った。

### 6.1 対象とするアーカイブ ～NHK デジタルアーカイブス～

本研究では NHK が保有する放送コンテンツを公開している NHK デジタルアーカイブス (NHK-DA) を対象にメタデータの作成を行う。2.1 節でも説明を行ったが、NHK-DA では約 1 万本の映像をテーマ別に下記 6 つのアーカイブで配信している。

- 戦争証言アーカイブス
- 東日本大震災アーカイブス
- エコチャンネル
- NHK 映像マップ みちしる
- NHK クリエイティブ・ライブラリー
- みのがし なつかし

戦争証言アーカイブスでは、戦争体験者の証言動画や関連する「番組」、「日本ニュース」、「戦時録音資料」を視聴することができる[24]。放送コンテンツにあたるのは証言動画、日本ニュース、戦時録音資料である。これらのメタデータをモデルに適用して RDF ファイルとして記述する。



図 27 戦争証言アーカイブス (<http://www.nhk.or.jp/shogenarchives/>)

東日本大震災アーカイブスは2013年3月11日に起こった東日本大震災に関する証言やニュースなどの動画を地図上から選んで見ることができるサイトである[25]。主な放送コンテンツとして証言映像と復興映像があり、それらのメタデータを記述する。



図 28 東日本大震災アーカイブス (<http://www9.nhk.or.jp/311shogen/>)

エコチャンネルはNHKの環境情報専門の動画サイトである[5]。最新のエコビジネス情報や持続可能なエコライフの魅力、大自然の恵みを実感できる生き物たちの映像などのコンテンツを楽しむことができる。

2014年12月18日時点で、41本の番組から収録された動画は370本存在しており、それらのメタデータの収集を行った。



図 29 エコチャンネル (<http://www.nhk.or.jp/eco-channel/>)



NHK 映像マップ みちしるは「新日本風土記」などのテレビ番組や NHK が保存している歴史的映像から厳選した日本の映像を視聴できる動画配信サービスである[26]。コンテンツを地図とマッピングしており、地図から検索することやジャンルやキーワードから動画を検索することも可能となっている。2013 年 12 月 25 日時点での動画の本数は 2495 本あり、スクレイピングを行いそのメタデータの収集を行った。



図 30 NHK 映像マップ みちしる (<http://www.nhk.or.jp/michi/>)

NHK クリエイティブ・ライブラリーは NHK アーカイブスの番組や番組素材から切り出した映像や音声を視聴者の創作活動や表現に利用できるように「創作用素材」として提供するサービスである[6]。2014 年 11 月 26 日時点での動画の本数は 5097 本あり、それらのメタデータの収集を行った。



図 31 NHK クリエイティブ・ライブラリー (<http://www1.nhk.or.jp/creative/>)

みのがし なつかしは NHK が放送した番組やニュースを年代、ジャンルなどから検索・閲覧することができるサービスである。

2014 年 12 月 30 日時点でのコンテンツの本数は 1933 件あり、そのコンテンツの種類はニュース映像とテレビ番組があり、それら 2 つのメタデータの収集を行った。

The screenshot shows the NHK 'Mino Gashi Natsumashi' website. At the top, there's a header with the logo and navigation links like '使い方' (How to use), '登録するとこんな特典が!' (Benefits of registration), and 'ログイン' (Login). Below the header is a search bar with 'キーワードで探す' (Search by keyword) and a '検索' (Search) button. There are also filters for 'オンエアから探す' (Search by broadcast), '放送年から探す' (Search by broadcast year), 'ジャンルから探す' (Search by genre), '地域から探す' (Search by region), '特集テーマから探す' (Search by featured theme), 'ランキングから探す' (Search by ranking), and 'イメージから探す' (Search by image). A message states: '左列の時刻表から番組を選べると、関連する過去の番組や、出演者が登場した過去の番組が表示されます。' (Selecting a program from the left column's schedule will display related past programs and past programs featuring the cast members). The main content area is divided into two columns. The left column, titled '時刻表' (Schedule) and '東京' (Tokyo), shows a schedule for January 9th (Friday) from 12:00 PM onwards. The right column, titled 'ニュース・気象情報' (News and Weather Information), shows related past programs and cast members. The related past programs include 'テレビ気象台' (TV Weather Station), '金曜お天気博士' (Friday Weather Doctor), 'イブニングネットワーク' (Evening Network), and 'にっぽん列島ただいま6時' (Nippon Ryushima Tadayima 6 PM). The cast members section is empty, displaying '出演者: -' (Cast: -). A large message '情報はありません' (No information) is displayed in the center of the right column.

図 32 みのがし なつかし (<http://www.nhk.or.jp/archives/search/onair/>)

## 6.2 テレビ番組情報のメタデータ構築

テレビ番組情報のメタデータを作成するために「みのがし なつかし」が公開しているテレビ番組情報を用いる。

図 33 はみのがし なつかしテレビ番組のウェブページ「連続テレビ小説 あまちゃん」に記載されたメタデータの一部である。5.1 節で説明を行ったスクレイピング手法を用いて取得したメタデータが表 7 となる。4.2 節で定義したメタデータ記述構造をもとに表 7 のメタデータを適用する。シリーズ ID となる値は URL の値「das\_id=」の後方につく値とする。この値は NHK-DA 共通の ID となっており、リソースの URI にも使用する。「連続テレビ小説 あまちゃん」の場合は URI に付与されている値「D0009050069\_00000」として定義するとともに、放送コンテンツのリソースの URI は「[http://purl.org/net/tv/series/D0009050069\\_00000](http://purl.org/net/tv/series/D0009050069_00000)」となる。この記述構造を元に 5.2 節で述べた RDF 化手法を用いて作成した RDF/TURLTE ファイルが図 34 である。実際に作成したテレビ番組のリソース数は 1552 個となり、このデータを用いて放送コンテンツ同士の連携を行う。

■ 放送期間

2013年

■ この動画・静止画の番組放送年

2013年

■ 主な出演者

能年玲奈、小泉今日子、宮本信子、杉本哲太、渡辺えり

■ 番組詳細

夏休み。母・春子（小泉今日子）に連れられ、東京から初めて北三陸にやってきたヒロイン・天野アキ（能年玲奈）は、祖母・夏（宮本信子）と出会う。現役の海女を続ける祖母は、人生で初めて出会った「カッコいい！」と思える女性だった。「私、海女になりたいかも...」。祖母に憧れて海女を目指すアキは、やがて地元アイドルに。ヒロインの笑顔が元気を届ける人情喜劇。脚本家・宮藤官九郎が故郷の東北を舞台に描くオリジナル作品。

原作：宮藤官九郎

図 33 みのがし なつかしが公開しているテレビ番組の情報

([http://cgi2.nhk.or.jp/archives/tv60bin/detail/index.cgi?das\\_id=D0009050069\\_00000](http://cgi2.nhk.or.jp/archives/tv60bin/detail/index.cgi?das_id=D0009050069_00000))

表 7 「連続テレビ小説 あまちゃん」のメタデータ詳細

URL	<a href="http://cgi2.nhk.or.jp/archives/tv60bin/detail/index.cgi?das_id=D0009050069_00000">http://cgi2.nhk.or.jp/archives/tv60bin/detail/index.cgi?das_id=D0009050069_00000</a>
タイトル	連続テレビ小説 あまちゃん
ジャンル	連続テレビ小説
放送期間	2013 年
番組放送年	2013 年
番組詳細	夏休み。母・春子（小泉今日子）に連れられ、東京から初めて北三陸にやってきたヒロイン・天野アキ（能年玲奈）は、祖母・夏（宮本信子）と出会う。現役の海女を続ける祖母は、人生で初めて出会った「カッコいい！」と思える女性だった。「私、海女になりたいかも...」。祖母に憧れて海女を目指すアキは、やがて地元アイドルに。ヒロインの笑顔が元気を届ける人情喜劇。脚本家・宮藤官九郎が故郷の東北を舞台に描くオリジナル作品。
主な出演者	能年玲奈、小泉今日子、宮本信子、杉本哲太、渡辺えり
原作	宮藤官九郎



```

<http://purl.org/net/tv/series/D0009050069_00000> rdf:type tv:Series.
<http://purl.org/net/tv/series/D0009050069_00000> tv:identifier "D0009050069_00000".
<http://purl.org/net/tv/series/D0009050069_00000> dc:identifier <http://cgi2.nhk.or.jp/archives/tv60bin/detail/index.cgi?das_id=D0009050069_00000>.
<http://purl.org/net/tv/series/D0009050069_00000> dc:title "連続テレビ小説 あまちゃん"@ja.
<http://purl.org/net/tv/series/D0009050069_00000> dcterms:alternative "あまちゃん"@ja.
<http://purl.org/net/tv/series/D0009050069_00000> dc:description "夏休み。母・春子（小泉今日子）に連れられ、東京から初めて北三陸にやってきたヒロイン・天野アキ（能年玲奈）は、祖母・夏（宮本信子）と出会う。現役の海女を続ける祖母は、人生で初めて出会った「カッコいい！」と思える女性だった。「私、海女になりたいかも…」。祖母に憧れて海女を目指すアキは、やがて地元アイドルに。ヒロインの笑顔が元気を届ける人情喜劇。脚本家・宮藤官九郎が故郷の東北を舞台に描くオリジナル作品。"@ja.
<http://purl.org/net/tv/series/D0009050069_00000> dc:date "2013".
<http://purl.org/net/tv/series/D0009050069_00000> dcterms:temporal "2013".
<http://purl.org/net/tv/series/D0009050069_00000> rdfs:seeAlso <http://ja.dbpedia.org/resource/あまちゃん>.
<http://purl.org/net/tv/series/D0009050069_00000> dc:subject "連続テレビ小説".
<http://purl.org/net/tv/series/D0009050069_00000> schema:actor <http://purl.org/net/tv/person/能年玲奈>.
<http://purl.org/net/tv/person/能年玲奈> rdf:type foaf:Person.
<http://purl.org/net/tv/person/能年玲奈> foaf:name "能年玲奈".
<http://purl.org/net/tv/person/能年玲奈> rdfs:seeAlso <http://ja.dbpedia.org/resource/能年玲奈>.
<http://purl.org/net/tv/series/D0009050069_00000> schema:actor <http://purl.org/net/tv/person/小泉今日子>.
<http://purl.org/net/tv/person/小泉今日子> rdf:type foaf:Person.
<http://purl.org/net/tv/person/小泉今日子> foaf:name "小泉今日子".
<http://purl.org/net/tv/person/小泉今日子> rdfs:seeAlso <http://ja.dbpedia.org/resource/小泉今日子>.
<http://purl.org/net/tv/series/D0009050069_00000> schema:actor <http://purl.org/net/tv/person/宮本信子>.
<http://purl.org/net/tv/person/宮本信子> rdf:type foaf:Person.
<http://purl.org/net/tv/person/宮本信子> foaf:name "宮本信子".
<http://purl.org/net/tv/person/宮本信子> rdfs:seeAlso <http://ja.dbpedia.org/resource/宮本信子>.
<http://purl.org/net/tv/series/D0009050069_00000> schema:actor <http://purl.org/net/tv/person/杉本哲太>.
<http://purl.org/net/tv/person/杉本哲太> rdf:type foaf:Person.
<http://purl.org/net/tv/person/杉本哲太> foaf:name "杉本哲太".
<http://purl.org/net/tv/person/杉本哲太> rdfs:seeAlso <http://ja.dbpedia.org/resource/杉本哲太>.
<http://purl.org/net/tv/series/D0009050069_00000> schema:actor <http://purl.org/net/tv/person/渡辺えり>.
<http://purl.org/net/tv/person/渡辺えり> rdf:type foaf:Person.
<http://purl.org/net/tv/person/渡辺えり> foaf:name "渡辺えり".
<http://purl.org/net/tv/person/渡辺えり> rdfs:seeAlso <http://ja.dbpedia.org/resource/渡辺えり>.
<http://purl.org/net/tv/series/D0009050069_00000> dc:contributor <http://purl.org/net/tv/person/宮藤官九郎>.
<http://purl.org/net/tv/person/宮藤官九郎> rdf:type foaf:Person.
<http://purl.org/net/tv/person/宮藤官九郎> foaf:name "宮藤官九郎".

```

図 34 「連続テレビ小説 あまちゃん」のメタデータの RDF/TURTLE ファイル

### 6.3 NHK-DA メタアーカイブのメタデータ記述規則

本研究では、4.1 節で述べたメタデータモデルをもとに NHK-DA の各アーカイブの放送コンテンツの記述規則を定義して、RDF ファイルの作成を行った。結果から述べると、各アーカイブの記述項目こそ差異はあるものの、本研究で定義した主要なメタデータ項目は各アーカイブに適用できたので、メタデータの統合ができたといえる。本節では定義した各アーカイブのメタデータ記述規則について説明を行う。

戦争証言アーカイブスの放送コンテンツのメタデータには多数のメタデータが記述されているため、提案したメタデータモデルの項目だけでは RDF ファイルを作成することが難しい。特に映像に対して、セグメント毎に再生テキストが付与されているため、セグメントをクラスとして宣言し、構造化を行う。また、足りない項目はメタデータ語彙を新たに宣言し、構造化を行う必要がある。

表 8 はメタデータの記述構造に使用する名前空間接頭辞、表 9・表 10 は戦争証言アーカイブスの放送コンテンツメタデータの記述規則である。地理情報や収録年月日など足りない項目はメタデータ語彙 vCard Ontology[27]と FOAF[28]を利用した。セグメントを記述する際、tv:Segment クラスを新たに宣言し、メタデータの構造化を行った。

表 8 戦争証言アーカイブスの放送コンテンツで定義する名前空間接頭辞

[@NS]	
rdfs	http://www.w3.org/2000/01/rdf-schema#
dc	http://purl.org/dc/elements/1.1/
dcterms	http://purl.org/dc/terms/
foaf	http://xmlns.com/foaf/0.1/
v	http://www.w3c.org/2006/vcard/ns#
tv	http://purl.org/net/tv/ns#
schema	http://schema.org/

表 9 戦争証言アーカイブスの放送コンテンツのメタデータ記述規則

[MAIN]						
#記述項目名	プロパティ	最小	最大	値タイプ	値制約	説明
コンテンツ	tv:Episode	1	1	ID	<http://purl.org/net/tv/episode/>	
タイトル	dc:title	1	1	文字列		
概要	dc:description	1	1	文字列		
ID	tv:identifier	1	1	文字列		
情報源	dc:identifier	1	1	参照値		テレビ番組の取得元
登場人物	schema:actor	0	1	構造化	#構造化登場人物	
番組	tv:hasSeries	0	1	参照値	<http://purl.org/net/tv/series/>	
放送年月日	dc:date	0	1	文字列	YYYY-MM-DD	
収録年月日	schema:dateCreated	0	1	文字列	YYYY-MM-DD	
戦地	v:location	0	1	文字列		
キーワード	tv:hasKeyword	0		構造化	#構造化キーワード	
出来事の背景	schema:about	0	1	文字列		テレビ番組の主題の背景
関連証言	dc:relation	0		参照値		関連する他のテレビ番組
カテゴリ	tv:hasCategory	1	1	文字列		テレビ番組のカテゴリ
セグメント	tv:hasSegment	0		構造化	#構造化セグメント	

表 10 戦争証言アーカイブスの放送コンテンツのメタデータ記述規則 構造化メタデータ

[構造化登場人物]						
#記述項目名	プロパティ	最小	最大	値タイプ	値制約	説明
登場人物	foaf:Person	1	1		<http://purl.org/net/tv/person/>	
氏名	foaf:name	0	1	文字列		
所属組織	schema:jobTitle	0	1	構造化	#構造化所属組織	
プロフィール	schema:description	0	1	文字列		
[構造化所属組織]						
#記述項目名	プロパティ	最小	最大	値タイプ	値制約	説明
所属組織	foaf:Organization	1	1	ID	<http://www.nhk.or.jp/tv/organization/>	
ラベル	rdfs:label	1	1	文字列		
外部リンク	rdfs:seeAlso	0	1	参照値	<http://ja.dbpedia.org/resource/>	
[構造化キーワード]						
#記述項目名	プロパティ	最小	最大	値タイプ	値制約	説明
キーワード	tv:Keyword	1	1	ID	<http://purl.org/net/tv/vocabulary/>	
ラベル	rdfs:label	1	1	文字列		
外部リンク	rdfs:seeAlso	0	1	参照値	<http://ja.dbpedia.org/resource/>	
[構造化セグメント]						
#記述項目名	プロパティ	最小	最大	値タイプ	値制約	説明
セグメント	tv:Segment	1	1	ID	<http://purl.org/net/tv/segment/>	
タイトル	dc:title	1	1	文字列		セグメントのタイトル
再生時間	schema:duration	1	1	文字列		セグメントの再生時間
概要	dc:description	1	1	文字列	<http://ja.dbpedia.org/resource/>	セグメントの概要
キーワード	tv:hasKeyword	0		参照値	<http://purl.org/net/tv/vocabulary/>	

東日本大震災アーカイブスの放送コンテンツもまたセグメント構造となっているため、セグメントメタデータの構造化を行う必要がある。表 11 はメタデータの記述構造に使用する名前空間接頭辞、表 12 は東日本大震災アーカイブスの放送コンテンツメタデータの記述規則である。市町村・登場人物のメタデータを記述するために vCard Ontology と FOAF を利用する。登場人物のメタデータはテレビ番組の出演者・制作者のメタデータの構造化と同様の構造とする。

表 11 東日本大震災アーカイブスの放送コンテンツで定義する名前空間接頭辞

[@NS]	
rdfs	http://www.w3.org/2000/01/rdf-schema#
dc	http://purl.org/dc/elements/1.1/
dcterms	http://purl.org/dc/terms/
foaf	http://xmlns.com/foaf/0.1/
v	http://www.w3c.org/2006/vcard/ns#
da	http://purl.org/net/da/ns#
schema	http://schema.org/

表 12 東日本大震災アーカイブスの放送コンテンツのメタデータ記述規則

[MAIN]						
#記述項目名	プロパティ	最小	最大	値タイプ	値制約	説明
コンテンツ	tv:Episode	1	1	ID	<http://purl.org/net/hda/content/>	
タイトル	dc:title	1	1	文字列		
概要	dc:description	1	1	文字列		
ID	tv:identifier	1	1	文字列		
情報源	dc:identifier	1	1	参照値	<http://www9.nhk.or.jp/311shogen/map/#/>	テレビ番組の取得元
登場人物	schema:actor	0	1	文字列	#構造化登場人物	
番組	tv:hasSeries	0	1	参照値	http://purl.org/net/tv/series/	
放送年月日	dc:date	0	1	文字列	YYYY-MM-DD	
市町村名	v:region	1	1	文字列		
カテゴリ	tv:hasCategory	1	1	参照値		
セグメント	tv:hasSegment	0		構造化	#構造化セグメント	
[構造化セグメント]						
#記述項目名	プロパティ	最小	最大	値タイプ	値制約	説明
セグメント	tv:Segment	1	1	ID	<http://purl.org/net/tv/segment/>	
タイトル	dc:title	1	1	文字列		
再生時間	schema:duration	1	1	文字列		セグメントの再生時間
概要	dc:description	1	1	文字列		セグメントの概要
キーワード	tv:hasKeyword	0		参照値	<http://www.nhk.or.jp/tv/vocabulary/>	
[構造化登場人物]						
#記述項目名	プロパティ	最小	最大	値タイプ	値制約	説明
登場人物	foaf:Person	1	1	ID		
氏名	foaf:name	0	1	文字列		登場人物の氏名

エコチャンネルの放送コンテンツのメタデータは 5.2 で実際にメタデータを作成したように項目が充足しているので、メタデータモデルをそのまま適用する。表 13 はメタデータの記述構造に使用する名前空間接頭辞、表 14 はエコチャンネルの放送コンテンツメタデータの記述規則である。

表 13 エコチャンネルの放送コンテンツで定義する名前空間接頭辞

rdfs	<a href="http://www.w3.org/2000/01/rdf-schema#">http://www.w3.org/2000/01/rdf-schema#</a>
dc	<a href="http://purl.org/dc/elements/1.1/">http://purl.org/dc/elements/1.1/</a>
dcterms	<a href="http://purl.org/dc/terms/">http://purl.org/dc/terms/</a>
tv	<a href="http://purl.org/net/tv/ns#">http://purl.org/net/tv/ns#</a>

表 14 エコチャンネルの放送コンテンツのメタデータ記述規則

[MAIN]						
#記述項目名	プロパティ	最小	最大	値タイプ	値制約	説明
コンテンツ	tv:Episode	1	1	ID	<a href="http://purl.org/net/tv/episode/">http://purl.org/net/tv/episode/</a>	エコチャンネルで公開されているコンテンツ
タイトル	dc:title	1	1	文字列		
概要	dc:description	1	1	文字列		
ID	tv:identifier	1	1	文字列		コンテンツページのURLに含まれるdasID
情報源	dc:identifier	1	1	参照値	<a href="http://cgi4.nhk.or.jp/eco-channel/jp/movie/play.cgi?did=">http://cgi4.nhk.or.jp/eco-channel/jp/movie/play.cgi?did=</a>	エコチャンネル上の各コンテンツページ
放送日	dc:date	1	1	文字列	YYYY-MM-DD	
カテゴリー	tv:hasCategory	0		構造化	#構造化カテゴリー	
キーワード	tv:hasKeyword	1		構造化	#構造化キーワード	
テレビ番組	tv:hasSeries	1	1	参照値	<a href="http://purl.org/net/tv/series/">http://purl.org/net/tv/series/</a>	
[構造化カテゴリー]						
#記述項目名	プロパティ	最小	最大	値タイプ	値制約	説明
カテゴリー	tv:Keyword	1	1	ID	<a href="http://www.nhk.or.jp/tv/vocabulary/">http://www.nhk.or.jp/tv/vocabulary/</a>	統合語彙のURI
ラベル	rdfs:label	1	1	文字列		
外部リンク	rdfs:seeAlso	0	1	参照値	<a href="http://ja.dbpedia.org/resource/">http://ja.dbpedia.org/resource/</a> <a href="http://id.ndl.go.jp/auth/ndlsh/">http://id.ndl.go.jp/auth/ndlsh/</a>	DBpedia,NDLのリソース
[構造化キーワード]						
#記述項目名	プロパティ	最小	最大	値タイプ	値制約	説明
キーワード	tv:Keyword	1	1	ID	<a href="http://www.nhk.or.jp/tv/vocabulary/">http://www.nhk.or.jp/tv/vocabulary/</a>	統合語彙のURI
ラベル	rdfs:label	1	1	文字列		
外部リンク	rdfs:seeAlso	0	1	参照値	<a href="http://ja.dbpedia.org/resource/">http://ja.dbpedia.org/resource/</a> <a href="http://id.ndl.go.jp/auth/ndlsh/">http://id.ndl.go.jp/auth/ndlsh/</a>	DBpedia,NDLのリソース

NHK 映像マップ みちしる（みちしる）の放送コンテンツには多数のメタデータが付与されている。そのため、本研究で定義するメタデータモデルの語彙では記述できないプロパティが存在する。表 15 はみちしるのメタデータを記述する際に用いる名前空間接頭辞で表 16 はメタデータ記述規則である。みちしるのコンテンツには、動画 URL・サムネイル・撮影地・都道府県・緯度・経度といったメタデータ項目が存在しており、これらを記述する際にメタデータ語彙 tv 及び vCard Ontology と Basic Geo(WGS84 lat/long) Vocabulary[29]、FOAF を利用する。

表 15 NHK 映像マップ みちしるの放送コンテンツで定義する名前空間接頭辞

[@NS]	
rdfs	<a href="http://www.w3.org/2000/01/rdf-schema#">http://www.w3.org/2000/01/rdf-schema#</a>
dc	<a href="http://purl.org/dc/elements/1.1/">http://purl.org/dc/elements/1.1/</a>
dcterms	<a href="http://purl.org/dc/terms/">http://purl.org/dc/terms/</a>
foaf	<a href="http://xmlns.com/foaf/0.1/">http://xmlns.com/foaf/0.1/</a>
v	<a href="http://www.w3c.org/2006/vcard/ns#">http://www.w3c.org/2006/vcard/ns#</a>
wgs84_pos	<a href="http://www.w3.org/2003/01/geo/wgs84_pos#">http://www.w3.org/2003/01/geo/wgs84_pos#</a>
tv	<a href="http://purl.org/net/tv/ns#">http://purl.org/net/tv/ns#</a>

表 16 NHK 映像マップ みちしるの放送コンテンツのメタデータ記述規則

[MAIN]						
#記述項目名	プロパティ	最小	最大	値タイプ	値制約	説明
エピソード	tv:Episode	1	1	ID	<a href="http://purl.org/net/tv/episode/">http://purl.org/net/tv/episode/</a>	みちしるで公開されているコンテンツ
タイトル	dc:title	1	1	文字列		
サブタイトル	dcterms:alternative	1	1	文字列		
概要	dc:description	1	1	文字列		
みちしるID	tv:identifier	1	1	文字列		コンテンツページのURLに含まれるdasID
情報源	dc:identifier	1	1	参照値	<a href="http://cgi2.nhk.or.jp/michi/cgi/detail.cgi?dasID=&gt;">http://cgi2.nhk.or.jp/michi/cgi/detail.cgi?dasID=&gt;</a>	みちしる上の各コンテンツページ
動画URL	tv:movie_url	1	1	参照値		
サムネイル	foaf:depiction	1	1	参照値		
放送年	dc:date	0	1	文字列		
撮影地	dcterms:spatial	0	1	文字列		撮影者が居た場所。複数の場合はカンマ区切り
都道府県	v:region	1	1	文字列		
緯度	wgs84_pos:lat	1	1	文字列		
経度	wgs84_pos:long	1	1	文字列		
テーマ	tv:hasCategory	0		構造化	#構造化カテゴリー	
キーワード	tv:hasKeyword	1		構造化	#構造化キーワード	
[構造化テーマ]						
#記述項目名	プロパティ	最小	最大	値タイプ	値制約	説明
テーマ	tv:Category	1	1	ID	<a href="http://www.nhk.or.jp/tv/vocabulary/">http://www.nhk.or.jp/tv/vocabulary/</a>	統合語彙のURI
ラベル	rdfs:label	1	1	文字列		
外部リンク	rdfs:seeAlso	0	1	参照値	<a href="http://ja.dbpedia.org/resource/">http://ja.dbpedia.org/resource/</a> <a href="http://id.ndl.go.jp/auth/ndlsh/">http://id.ndl.go.jp/auth/ndlsh/</a>	外部LODへのリンク
[構造化キーワード]						
#記述項目名	プロパティ	最小	最大	値タイプ	値制約	説明
キーワード	tv:Keyword	1	1	ID	<a href="http://www.nhk.or.jp/tv/vocabulary/">http://www.nhk.or.jp/tv/vocabulary/</a>	統合語彙のURI
ラベル	rdfs:label	1	1	文字列		
外部リンク	rdfs:seeAlso	0	1	参照値	<a href="http://ja.dbpedia.org/resource/">http://ja.dbpedia.org/resource/</a> <a href="http://id.ndl.go.jp/auth/ndlsh/">http://id.ndl.go.jp/auth/ndlsh/</a>	外部LODへのリンク

NHK クリエイティブ・ライブラリー（クリエイティブ・ライブラリー）の放送コンテンツのメタデータをデータモデルに適用する。表 17 はクリエイティブ・ライブラリーのメタデータを定義するときに使用する名前空間接頭辞、表 18 はメタデータの記述規則である。この記述規則を用いて、メタデータの RDF ファイルを作成する。クリエイティブ・ライブラリーの放送コンテンツは提案したメタデータモデルと適応するものが多いが、異なる項目として副題・関連する素材が挙げられる。副題は DC タームズのプロパティの `dcterms:alternative` を用いる。関連する素材つまり関連コンテンツは「`tv:isRelatedTo`」プロパティを定義してリソース同士の関係を表現する。

表 17 NHK クリエイティブ・ライブラリーの放送コンテンツで定義する名前空間接頭辞

[@NS]	
<code>rdfs</code>	<code>http://www.w3.org/2000/01/rdf-schema#</code>
<code>dc</code>	<code>http://purl.org/dc/elements/1.1/</code>
<code>dcterms</code>	<code>http://purl.org/dc/terms/</code>
<code>tv</code>	<code>http://purl.org/net/tv/ns#</code>

表 18 NHK クリエイティブ・ライブラリーの放送コンテンツのメタデータ記述規則

[MAIN]						
#記述項目名	プロパティ	最小	最大	値タイプ	値制約	説明
コンテンツ	<code>tv:Episode</code>	1	1	ID	<code>http://purl.org/net/tv/episode/</code>	公開されているコンテンツ
タイトル	<code>dc:title</code>	1	1	文字列		
副題	<code>dcterms:alternative</code>	1	1	文字列		タイトルの別名・略名
概要	<code>dc:description</code>	1	1	文字列		
ID	<code>tv:identifier</code>	1	1	文字列		コンテンツページのURLに含まれるdasID
情報源	<code>dc:identifier</code>	1	1	参照値	<code>http://www1.nhk.or.jp/creative/material/</code>	各コンテンツページのURL
放送年	<code>dc:date</code>	0	1	文字列	YYYY	テレビ番組が放送された年
カテゴリー	<code>tv:hasCategory</code>	0		構造化	#構造化カテゴリー	
キーワード	<code>tv:hasKeyword</code>	0		構造化	#構造化キーワード	
テレビ番組	<code>tv:hasSeries</code>	0	1	参照値	<code>http://purl.org/net/tv/series/</code>	
関連する素材	<code>tv:isRelatedTo</code>	0		参照値		
[構造化カテゴリー]						
#記述項目名	プロパティ	最小	最大	値タイプ	値制約	説明
カテゴリー	<code>tv:Keyword</code>	1	1	ID	<code>http://www.nhk.or.jp/tv/vocabulary/</code>	統合語彙のURI
ラベル	<code>rdfs:label</code>	1	1	文字列		
外部リンク	<code>rdfs:seeAlso</code>	0	1	参照値	<code>&lt;http://ja.dbpedia.org/resource/&gt;</code> <code>&lt;http://id.ndl.go.jp/auth/ndlsh/&gt;</code>	DBpedia,NDLのリソース
[構造化キーワード]						
#記述項目名	プロパティ	最小	最大	値タイプ	値制約	説明
キーワード	<code>tv:Keyword</code>	1	1	ID	<code>http://www.nhk.or.jp/tv/vocabulary/</code>	統合語彙のURI
ラベル	<code>rdfs:label</code>	1	1	文字列		
外部リンク	<code>rdfs:seeAlso</code>	0	1	参照値	<code>&lt;http://ja.dbpedia.org/resource/&gt;</code> <code>&lt;http://id.ndl.go.jp/auth/ndlsh/&gt;</code>	DBpedia,NDLのリソース

## 6.4 NHK-DA 統制語彙の作成

NHK-DA のメタアーカイブを構築する際に、4.3 で説明を行った主題情報のデータモデルを用いて共通の主題情報を定義する。

NHK-DA の放送コンテンツには主題情報としてカテゴリー、キーワード、テーマなどのメタデータが付与されている。しかし、その記述規則はアーカイブ毎に異なり、同じ意味の単語でも関連付けられていない。そこで NHK-DA の主題情報となるメタデータを収集し、URI を付与してリソースとして作成する。その主題情報を持つ語彙を NHK-DA 統制語彙として定義し、その作成手法を述べる。そのために NHK-DA の主題情報のメタデータを全て収集する。3 つのアーカイブからメタデータを収集し、そのキーワード数と他の LOD とリンクすることが出来るキーワード数を集計し、表 19 にまとめた。

表 19 3 つのアーカイブのキーワード数と他 LOD とのリンク可能な数

アーカイブ名	エコチャンネル	みちしる	クリエイティブ・ライブラリー
総キーワード数	335	255	7326
DBpedia にある単語数	312	205	5121
NDL にある単語数	212	131	1649
両方と合致する単語数	212	131	1619

上記 3 つのアーカイブに戦争証言アーカイブスのキーワードを加え、集計した結果、総キーワード数は 8007 個となった。これに対し、重複除去すると 7957 個となる。全角英数字と半角英数字が混在しているので、半角英数字に統一・重複除去を行った数は 7951 個となった。キーワードの中には○（よみがな）などキーワードの後ろに括弧が付いているものも存在したので、括弧を除去し、更に重複除去した結果、7941 個となった。残った単語に対して、LOD へのリンクが可能か調べたところ最終的に DBpedia が 5635 個、NDL が 1846 個となった。両方と合致する単語は 1814 個である（表 20 参照）。

表 20 NHK-DA 全体での総キーワード数とリンク可能な数

総キーワード数	7941
DBpedia にある単語数	5635
NDL にある単語数	1846
両方と合致する単語数	1814

上記 7941 個の単語を NHK-DA 統制語彙としてリソースの作成を行う。リソースの URI は”<http://purl.org/net/tv/vocabulary/>”を基準として、`rdfs:label` プロパティで単語名を付与する。放送コンテンツに主題情報を付与する場合、構造化を行い、`tv:Keyword` もしくは `tv:Category` クラスとして扱う。



## 6.5 メタアーカイブシステムの構築

作成したメタデータを用いて、複数のアーカイブへの統合的な検索を可能にしたメタアーカイブシステムの開発を行った。

本システムは Web ブラウザ上でコンテンツの検索・閲覧を行う Web アプリケーションとして開発を行った。図 35 はシステムの構成を表した図である。

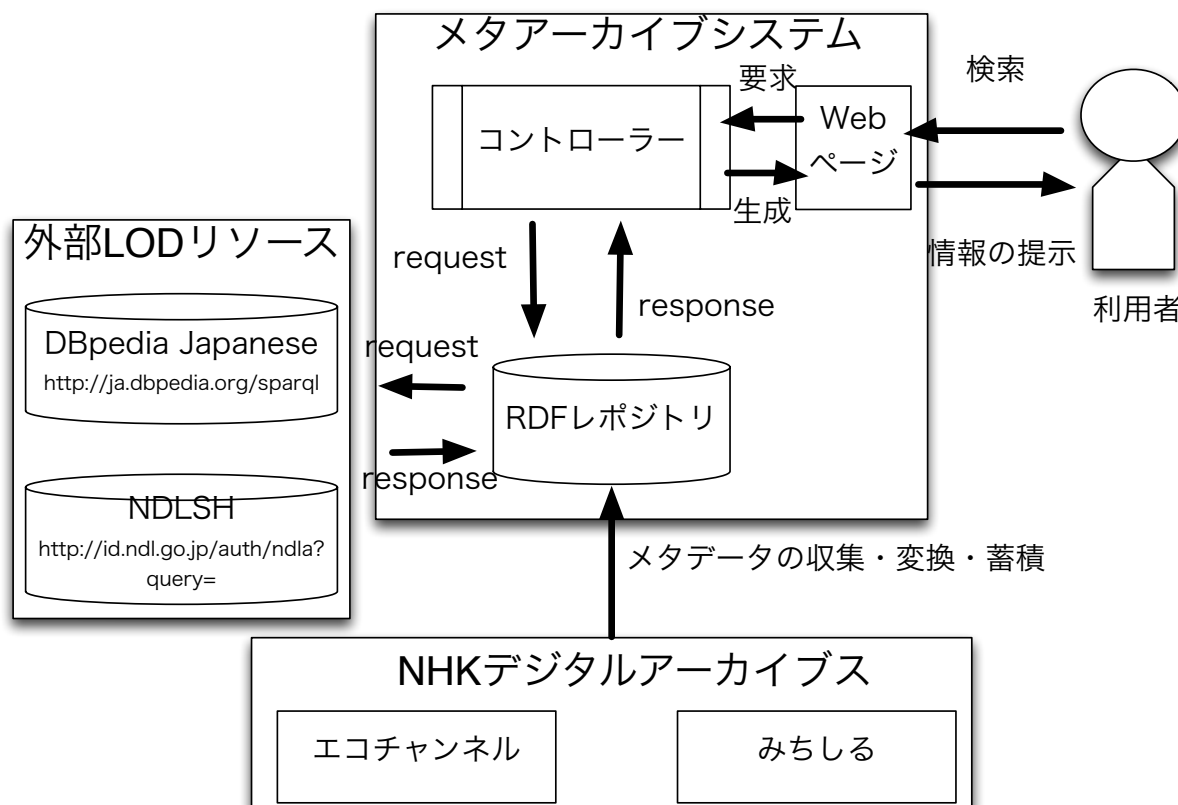


図 35 システム概要図

本システムは、アプリケーション利用者が入力された検索クエリやキーワードから RDF 問い合わせ言語 SPARQL の構文を生成し、メタデータが格納された RDF レポジトリに問い合わせ、求めるコンテンツの情報を持つ Web ページを出力する機能を持っている。NHK デジタルアーカイブスの放送コンテンツのメタデータを本研究で定義したメタデータモデルを用い、RDF で記述することにより、複数のアーカイブを横断した検索を可能としている。RDF レポジトリに対する問い合わせには SPARQL 構文を使用している。

図 36 はメタアーカイブシステム全体のページ遷移図を示したものである。基本的にトップページから検索及びキーワードページへのリンクを辿ることで目的の放送コンテンツページへアクセスすることができる。トップページにはテレビ番組のジャンルページへのリンクがあり、そのジャンルを持つテレビ番組ページにアクセスすることで、関連のある放送コンテンツを閲覧することを可能としている。個別ページにはキーワードページやテレビ番組ページへのリンクが表示されているので、ある放送コンテンツから関連のある放送コンテンツを検索することも容易である。

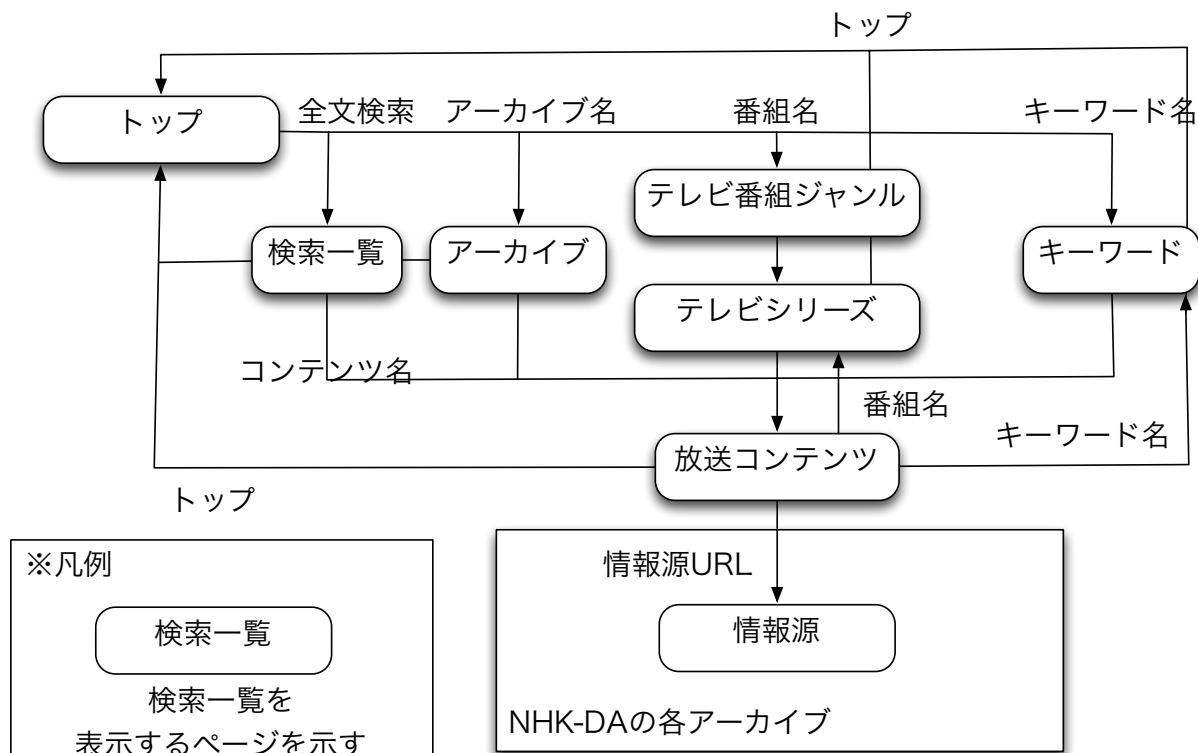


図 36 メタアーカイブシステム遷移図

図 37 はメタアーカイブのトップページ画面である。機能として、所蔵している全てのコンテンツに対する全文検索、キーワードによる検索及びジャンルからテレビ番組の検索による検索が可能となっている。また、情報元の放送コンテンツアーカイブから検索することもできる。

メタアーカイブシステム -NHKデジタルアーカイブス-

全文検索

キーワードから探す

[たばこ屋](#)
[オアフ島](#)
[ダイヤモンドヘッド](#)
[ワイキキビーチ](#)
[ホノルルビーチ](#)
[スパイ・スパイス](#)
[スリル](#)
[揺れ動く](#)
[男](#)
[やっつける](#)
[スイーツ](#)
[フレジェ](#)
[ケーキ](#)
[再](#)
[スキャット](#)
[小満](#)
[タブローム](#)
[遺跡](#)
[スイカ](#)
[イングランド](#)
[オンドリ](#)
[食数](#)
[美観地区](#)
[倉敷川](#)
[倉敷川](#)
[評伝](#)
[的](#)
[建造物群](#)
[保存地区](#)
[湯船](#)
[風呂](#)
[水上](#)
[湯量](#)
[すべり台](#)
[お化け](#)
[おぼけ](#)
[ハイキング](#)
[隅田川](#)
[清盛](#)
[グレンデ](#)
[スズメ](#)
[黒い](#)
[繁殖期](#)
[探れたて](#)
[NGC3132](#)
[8の字](#)
[星雲](#)
[着地](#)
[株毛](#)
[駄菓子](#)
[小銃](#)
[菓子](#)
[ガム](#)
[せんべい](#)
[ピンポン](#)
[あさり](#)
[ハナムグリ](#)

テレビ番組から探す

[特集](#)
[報道](#)
[ドキュメンタリー](#)
[自然](#)
[科学](#)
[紀行](#)
[音楽](#)
[教養](#)
[情報](#)
[連続テレビ小説](#)
[ドラマ](#)
[大河ドラマ](#)
[こども](#)
[教育](#)
[人形劇](#)
[アニメ](#)
[クイズ](#)
[バラエティー](#)
[伝統芸能](#)
[趣味](#)
[実用](#)
[特集ドラマ](#)
[くらし](#)
[趣味](#)

収録されているデジタルアーカイブから探す

[クリエイティブ・ライブラリー](#)
[東日本大震災アーカイブス](#)
[戦争証言アーカイブス](#)
[NHK映像マップ](#)
[みちしる](#)
[エコチャンネル](#)

図 37 メタアーカイブ トップページ画面

全文検索機能は前節で述べたメタデータの内、文字列の値を持つメタデータに対して、入力された語句を含むかどうかを判断して結果を出力する。例えば、「ラーメン」という語句に対して、定義したプロパティである `dc:title` や `dc:description` などの文字列の値を持つ目的語に「ラーメン」が含まれている場合、主語となる放送コンテンツの情報を検索結果としてページに表示する（図 38 参照）。

## 「ラーメン」の検索結果(45)



図 38 全文検索「ラーメン」による検索結果の表示

キーワードによる検索はクラス `tv:Keyword` であるリソースを RDF レポジトリから検索し、キーワードページへのリンクをトップページに表示する。図 39 はキーワード「兵庫県」に関するページ画面である。このページでは `tv:hasKeyword` や `tv:hasCategory` プロパティを介して、`tv:Keyword` クラス「兵庫県」を目的語としている放送コンテンツへのリンクを表示する機能を持つ。図 40 はキーワードページ「兵庫県」が表示している放送コンテンツを取得するための SPARQL 構文である。この構文では `rdf:label` の値が「兵庫県」であるキーワードリソースを検索し、そのキーワードリソースを `tv:hasKeyword` の目的語として持つ放送コンテンツリソースとそのタイトル名を取得している。また、`tv:Keyword` クラスには `rdfs:seeAlso` プロパティが定義されており、その目的語として DBpedia Japanese のリソースへのリンクが存在する。本システムでは、このリンクを用いて DBpedia Japanese の RDF レポジトリからリソースの概要説明文を取得、ページに表示する機能を持っており、利用者は関連情報を閲覧することができる。また、全文検索による検索結果一覧を表示するページやキーワードページには、放送コンテンツが収蔵されているアーカイブを段落ごとに分けてその数を表示する機能や各放送コンテンツのサムネイル画像を表示する機能を実装することで機能性の向上を行った。



図 39 メタアーカイブ キーワードページ画面

```

PREFIX tv:<http://purl.org/net/tv/ns#>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX dc:<http://purl.org/dc/elements/1.1/>
select distinct ?uri ?title where {
  ?x rdfs:label "兵庫県".
  ?uri tv:hasKeyword ?x.
  ?uri dc:title ?title.
}

```

図 40 キーワード「兵庫県」を持つリソースの URI とタイトルを取得する SPARQL 構文

次にテレビ番組による検索による機能の説明を行う。トップページからテレビ番組のジャンルページへとアクセスすることができる。テレビ番組のジャンルは前節で述べた `tv:Series` クラスがプロパティ `dc:subject` の値として持つ 16 項目からなる。ジャンルページではそのジャンルの値を持つ `tv:Series` リソースへのリンクを表示する。図 41 はジャンルページ「報道・ドキュメンタリー」にアクセスしたときの画面である。このページでは主にニュースに関するテレビ番組が並んでおり、タイトルを示したリンクから個別のテレビ番組ページにアクセス可能である。そのうち、「NHK ニュースおはよう日本」へのリンクにアクセスしたときの画面が図 42 である。プロパティ `dc:title` や `dc:description`、`dc:data` などの値からタイトル、説明文、放送年をページに出力しており、キーワードページと同様に関連する放送コンテンツへのリンクを一覧で表示する。図 42 で示す SPARQL 構文は関連する放送コンテンツへのリンクを作成するために取得するためのものである。この構文では「NHK ニュースおはよう日本」を `dc:title` として持つリソースをまず取得し、そのリソースを `tv:hasSeries` の目的語として持つ放送コンテンツリソースとそのタイトル名を取得している。

## 報道・ドキュメンタリー

[NHK特集 のぞみ5歳 ～手探りの子育て日記～](#)  
[NHK特集 ドキュメント昭和 ～世界への登場～](#)  
[にっぽん列島ただいま6時](#)  
[NHK特集 調査報告 チェルノブイリ原発事故](#)  
[NHK特集 救援～ヒロシマ・残留放射能の42年～](#)  
[NHK特集 奥羽山系 マタギの世界](#)  
[NHK特集 命もえつきる時～作家 檀一雄の最期～](#)  
[NHK特集 鶴になった男 ～釧路湿原 タンチョウふれあい日記～](#)  
[NHK特集 世界の中の日本 土地はだれのものか](#)  
[NHK特集 ミツコ 二つの世紀末](#)  
[NHK特集 蒲田・町工場物語](#)  
[NHK特集 二・二六事件 消された真実 ～陸軍軍法会議秘録～](#)  
[NHKニュース・トゥデー](#)  
[イブニングネットワーク](#)  
[NHKスペシャルドラマ ドラマ 失われし時を求めて—ヒロシマの夢—](#)  
[NHKスペシャル 世界一愛されたウサギ ?イギリス・ピーターラビットの田園から?](#)

図 41 テレビ番組ジャンルページ画面



図 42 テレビ番組の個別ページ画面

```

PREFIX tv:<http://purl.org/net/tv/ns#>
PREFIX dc:<http://purl.org/dc/elements/1.1/>
select distinct ?uri ?title where {
  ?x dc:title "NHK ニュースおはよう日本"@ja.
  ?uri tv:hasSeries ?x.
  ?uri dc:title ?title.
}

```

図 43 「NHK ニュースおはよう日本」をシリーズに持つリソースの URI とタイトルを取得する SPARQL 構文

図 44 は放送コンテンツの個別ページにアクセスしたときの画面である。個別ページで表示する情報量と内容はその放送コンテンツが持つメタデータによって、動的に変更する。このページではエコチャンネルの放送コンテンツについての情報を表示しており、タイトル、概要、放送日、カテゴリー、キーワード、収録元テレビ番組、収録先アーカイブ、情報元の URL を閲覧することができる。このページを生成するための SPARQL 構文を示したものが図 45 である。ID から上記のメタデータを取得するが、カテゴリーやキーワードを複数持つものが存在するため、`group_concat` 関数を用いることで一つの文字列として扱い、処理している。ページの生成の際、カテゴリー、キーワード、テレビ番組、アーカイブにはハイパーリンクを作成する。これは関連するキーワードページやテレビ番組ページ、アーカイブページへのアクセスを可能とし、他の放送コンテンツの閲覧を容易にするためである。

# アスベスト汚染

概要

アスベストは天然の鉱物繊維で、比較的値段が安い上、耐熱性や耐久性に優れていることから、長年、住宅や建物の建材に多く使われてきました。しかし、空気中に漂う繊維を長い間吸い続けると、肺がんや中皮腫（ちゅうひしゅ）を引き起こすことが知られています。1972年にはアスベストの発がん性が世界的に知られるようになっていましたが、日本での対策は大きく遅れました。2006年に使用が全面禁止になりましたが、健康被害が出るまでに30年から40年の潜伏期間があるとされ、被害の拡大が懸念されています。放送番組：特集 環境アーカイブス「伝えよう かけがえのない地球 第1回 豊かさの代償 四大公害病」

放送日 2006-08-27

カテゴリ 環境汚染・公害

キーワード 兵庫県 アスベスト 健康被害

テレビ番組 NHKアーカイブス

収録先アーカイブ エコチャンネル

情報源 [http://cgi4.nhk.or.jp/eco-channel/jp/movie/play.cgi?did=D0013771787\\_00000](http://cgi4.nhk.or.jp/eco-channel/jp/movie/play.cgi?did=D0013771787_00000)

[TOP](#)

図 44 メタアーカイブ 放送コンテンツ画面

```

PREFIX tv: <http://purl.org/net/tv/ns#>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX dc: <http://purl.org/dc/elements/1.1/>

select distinct * where
{
  {
    select distinct ?content (group_concat(?k ; separator =",")AS ?keywords)
    {
      ?content tv:hasKeyword ?keyword.
    }
    ?keyword rdfs:label ?k.
  }group by ?content
}

{
  select distinct ?content (group_concat(?c ; separator =",")AS ?categories)
  {
    ?content tv:hasCategory ?category.
  }
  ?category rdfs:label ?c.
}group by ?content
}

?content tv:identifier "D0013771069_000000";
dc:title ?title;
dc:description ?description;
dc:identifier ?source;
dc:date ?date;
tv:hasSeries ?series;
tv:isArchivedFrom ?archive.
?archive dc:title ?atitle.
?series dc:title ?stitle.
}

```

図 45 アーカイブ「エコチャンネル」の放送コンテンツのメタデータを ID から取得する SPARQL 構文

## 7. 考察と課題

本研究では、放送コンテンツアーカイブのメタアーカイブの構築とメタデータモデルの提案を行った。本節では、メタアーカイブの構築手法とメタデータモデルについての考察と課題について述べる。

表 21 は提案したメタデータ記述規則を NHK デジタルアーカイブス (NHK-DA) のアーカイブに適応した場合にその記述項目が存在するか示したものである。

「戦争証言アーカイブス」の場合、概要 (コンテンツ紹介文) の公開方法が異なるコンテンツがあった。「戦争証言アーカイブス」は証言、番組、戦時録音資料、日本ニュースという 4 つの種類のコンテンツが公開されており、そのコンテンツ自信の概要にあたるメタデータが記述されているものは番組コンテンツだけであった。その他のコンテンツは再生テキストという形式で公開されており、セグメント毎にその内容が変わるため、構造化したセグメントの概要としてメタデータの記述を行った。この事からは同じアーカイブの中でも放送コンテンツの内容によって記述項目を変更して公開している事例もあると考えられる。

「戦争証言アーカイブス」と「東日本大震災アーカイブス」では放送日ではなく、公開日と収録年月日、取材時期という項目名で公開されているコンテンツがあった。同じ時間情報でもその指し示す内容が微妙に異なるため、他のアーカイブと同様のプロパティで表現するのは適切ではなかった可能性がある。また、この 2 つのアーカイブでは、カテゴリ情報が付与されていなかった。これは両方共公開されているコンテンツのカテゴリに差がなかったためだと考えられる。戦争証言アーカイブスの場合、基本的にコンテンツの内容は戦争に関するものなので、大まかな主題情報であるカテゴリを設定する必要がなかったためである。東日本大震災アーカイブスの場合、その内容は東日本大震災に関するものが中心のため、カテゴリを設定する必要がなく、キーワードも元々付与されていなかった。

テレビ番組情報が付与されていないコンテンツを持つものは「戦争証言アーカイブス」と「みちしる」の 2 つであった。「戦争証言アーカイブス」はコンテンツの内容によって付与されている場合もあるがみちしるはどれもテレビ番組情報が付与されておらず、メタデータを作成することができなかった。

表 21 NHK デジタルアーカイブス 記述項目対応表

	戦争証言アーカイブス	東日本大震災アーカイブス	エコチャンネル	みちしる	クリエイティブ・ライブラリー
コンテンツ	○	○	○	○	○
タイトル	○	○	○	○	○
概要	△	○	○	○	○
放送日	△	△	○	×	○
ID	○	○	○	○	○
情報源	○	○	○	○	○
カテゴリ	×	×	○	○	○
キーワード	○	×	○	○	○
テレビ番組	△	○	○	×	○



今回提案したメタデータモデルは NHK-DA に適応する項目が多かったが、アーカイブによっては記述されているメタデータの種類が多く、メタデータモデルにない項目も多数存在した。例えば、「戦争証言アーカイブス」と「東日本大震災アーカイブス」の場合、メタデータとしてセグメント情報が記載されていた。これは映像の内容を分割し、その再生内容がテキスト化したものであり、提案するメタデータ項目にはないものであった。そのため、新たにプロパティとクラスの作成を行い、記述規則の定義を行った。

放送コンテンツアーカイブの中には地理情報について詳細に記述されているものも存在した。「戦争証言アーカイブス」と「東日本大震災アーカイブス」、「NHK 映像マップ みちしる」が今回それにあたる。特にみちしるは撮影地、都道府県といった項目名が記載されており、地理情報の語彙を新たに策定する必要があった。また、地理情報として記載されていないもののキーワードとして都道府県名が付与されているコンテンツもあった。今回作成したメタデータでは、都道府県名であってもキーワードとして扱うことにしたが、地理情報となるメタデータはプロパティを変更した方がより関連するコンテンツ同士を結び付けられたと考察できる。

提案したメタデータモデルは放送コンテンツの一般的なメタデータのみを対象に記述項目の選定を行ったため、足りない項目も多数存在する結果となった。今回のモデルでは放送情報と主題情報を中心にリンクを行ったが、より網羅的にコンテンツ同士をリンクするために地理情報の整理を行う必要があると推測できる。キーワードの項目に存在する地理情報を都道府県の項目にマッピングすることでデータを作成することが求められる。

次に放送情報のメタデータ作成に関する分析を行う。放送情報のメタデータは NHK-DA の「みのがしなつかし」からの収集を行い、RDF ファイルを作成した。しかし、収集したデータだけでは他のアーカイブのコンテンツとの連携が難しいパターンが存在した。NHK には「NHK スペシャル」というドキュメンタリー番組がある。この番組の映像は多数のアーカイブのコンテンツとして収録されているが、特集番組であるためか放送毎に内容のジャンルが全く異なっている。例えば、「戦争」に関する放送内容もあれば、「動物」に関する放送内容もある。そのため、アーカイブによって、「NHK アーカイブ」という大枠のシリーズ名ではなく、「NHK スペシャル サブタイトル名」のようにそのとき、放送されたサブタイトルがタイトルに付与されている場合も存在した。この問題を解決するために「NHK スペシャル」と「NHK スペシャル サブタイトル名」を別のリソースとして扱うことにした。そして、お互いの関連を示すため、プロパティ「dc:relate」を用いてシリーズ同士の関連を作成した。NHK のテレビ番組では「NHK スペシャル」のみならず「クローズアップ現代」のような特集番組をシリーズに持つコンテンツはサブタイトルがタイトル名に従属されているパターンがあるとわかった。

番組情報のメタデータ記述項目は「みのがしなつかし」を参考に作成を行ったため、番組情報のリソースが足りない場合があった。特に「エコチャンネル」のコンテンツはそれが顕著であり、「みのがしなつかし」から作成した番組情報のリソースで「エコチャンネル」のコンテンツと共通するリソースは「NHK ニュースおはよう日本」と「NHK ニュース 7」のみであった。足りないリソースはエコチャンネル内の番組情報から RDF ファイルを作成して、解決を行った。このことから「みのがしなつかし」だけでなく、より多くの番組情報が記載された情報源が必要だと考える。

番組情報リソースを作成する際、テレビ番組の変遷の関係を記述する必要があると考察する。例えば、「NHK スペシャル」には前身のタイトル「NHK 特集」が存在している。アーカイブのコンテンツの中には前身のタイトルで収録されている場合もあり、「NHK スペシャル」のコンテンツと「NHK 特集」のコンテンツはお互いに関連していると推察できる。しかし、文字列からその関連を発見することは非常に困難である。そのため、あらかじめ番組同士の関係を示したオントロジーや辞書資源が必要であると考察する。



今回、メタアーカイブシステムは **NHK-DA** を対象に構築を行った。今後の課題として、他の放送コンテンツアーカイブを対象にした構築を行うことが出来るか評価を行う必要がある。その上で、よりメタアーカイブの構築手法を一般化したものを考察する。提案したメタデータモデルが他のアーカイブにも適用できるか、利用性の向上のために番組情報や主題情報に何が足りないか考察することも求められる。

## 8. おわりに

映像や音声などのテレビ番組で放送されたコンテンツが公開されているアーカイブの利用を向上させるために、本研究では複数のアーカイブに対して統合的な検索を行うことができるメタアーカイブの構築とメタデータモデルの提案を行った。

WWW 上で公開されているデジタルアーカイブが増えている一方、その利用性の向上が求められている。放送コンテンツアーカイブは映像や音声など有益なコンテンツを公開しているが、アーカイブ同士の連携がなされていない場合が多い。それはアーカイブ毎に作成者や運営方針の違いからメタデータの記述規則が異なっているからである。そこで本研究ではアーカイブを連携し、統合検索を可能にするメタアーカイブシステムを構築する。そのために放送コンテンツアーカイブに共通するメタデータ項目と構造を見出し、LOD を用いたメタデータモデルの提案を行った。放送コンテンツアーカイブの共通情報である放送情報、特にテレビ番組の情報を整理して LOD 化することで、関連するコンテンツ同士のリンクを行う。また、メタアーカイブで統一された主題情報を用いるために、DBpedia や NDL SH といった他の LOD とリンクした統合語彙の作成手法について説明した。

提案した手法をもとに NHK デジタルアーカイブスを対象にしたメタアーカイブシステムの構築を行った。複数のアーカイブを対象にした統合検索を可能にしたが、メタデータを作成する上で問題も発生した。テレビ番組のシリーズの構造に関する例外があり、提案した放送情報のメタデータモデルでは適応できない場合が存在した。また、アーカイブによってメタデータの記述項目が足りない場合が存在し、地理情報の項目を充実することでよりコンテンツ同士の連携を行えることも推察できた。NHK デジタルアーカイブスだけでなく、他の放送コンテンツアーカイブにも適用し、本手法の評価を行うことも課題に挙げられる。

## 謝辞

本研究を進めるにあたり、指導教員の杉本重雄先生と副指導教員の永森先生には研究の議論、アドバイス、論文としてのまとめ方や添削など様々な面でご指導を頂きました。心より感謝を申し上げます。併せて、研究目的の設定やメタデータモデルの問題のご指摘など、研究を進める上で重要なご意見を下さった阪口哲夫先生と森嶋厚行先生、そしてメタデータを作成する上で必要なデータの提供をして下さった日本放送協会オンデマンド室室長、辻泰明様と日本放送協会の方々に深く感謝を致します。

また、研究を進める議論の中で様々なご意見をくださった研究室の三原鉄也先輩とメタデータ作成及びシステムの作成に関して指導頂いた研究室の本間維先輩に心より感謝を申し上げます。本研究のメタデータの作成をお手伝いいただいた後輩の池元貴哉君にも深く感謝致します。また、共に研究を進め、数多くの助言を与えていただいた杉本・永森研究室の同期と後輩の皆様に感謝致します。

## 参考文献

- [1] 知のデジタルアーカイブに関する研究会開催要綱.  
[http://www.soumu.go.jp/main\\_content/000101009.pdf](http://www.soumu.go.jp/main_content/000101009.pdf)(アクセス:2014.10.17)
- [2] BBC Archive. <http://www.bbc.co.uk/archive/>, (アクセス : 2015.1.5)
- [3] Ina.fr : vidéo, radio, audio et publicité - Actualités, archives de la radio et de la télévision en ligne - Archives vidéo et radio Ina.fr. <http://www.ina.fr/>, (アクセス : 2015.1.5)
- [4] NHK デジタルアーカイブス. <http://www.nhk.or.jp/archives/digital/>, (アクセス : 2015.1.5)
- [5] エコチャンネル. <http://www.nhk.or.jp/eco-channel/>, (アクセス : 2015.1.10)
- [6] NHK クリエイティブ・ライブラリー. <http://www1.nhk.or.jp/creative/>, (アクセス:2015.1.10)
- [7] デジタルアーカイブの連携・構築のためのガイドライン.  
[http://www.soumu.go.jp/main\\_content/000153595.pdf](http://www.soumu.go.jp/main_content/000153595.pdf), (アクセス : 2014.10.17)
- [8] 武田英明. 「Linked Data の動向」『カレントアウェアネス』 2011 (308) CA1746 pp. 8-11.  
<http://current.ndl.go.jp/files/ca/ca1746.pdf>, (アクセス : 2015.1.10)
- [9] Linked Data -Connect Distributed Data across the Web .<http://linkeddata.org> ,(アクセス : 2015.1.13)
- [10] DBpedia. <http://dbpedia.org/>, (アクセス : 2014.12.10)
- [11] 国立国会図書館件名標目表. <http://id.ndl.go.jp/auth/ndla>, (アクセス : 2014.10.19)
- [12] RDF -Semantic Web Standards. <http://www.w3.org/RDF/> , (アクセス : 2014.10.19)
- [13] NHK スペシャル | NHK エコチャンネル 環境情報の動画サイト【ECO CHANNEL】 .  
<http://cgi4.nhk.or.jp/eco-channel/jp/program/detail.cgi?prgid=special>, (アクセス : 2015.1.5)
- [14] DBpedia Japanese. <http://ja.dbpedia.org>, (アクセス : 2015.1.5)
- [15] Europeana. <http://www.europeana.eu/>, (アクセス : 2014.10.18)
- [16] Europeana Data Model Primer.  
<http://pro.europeana.eu/documents/900548/770bdb58-c60e-4beb-a687-874639312ba5> (アクセス : 2014.11.05)
- [17] 国立国会図書館東日本大震災アーカイブ. <http://kn.ndl.go.jp/>, (アクセス : 2014.10.17)
- [18] 国立国会図書館ダブリンコアメタデータ記述,  
[http://dl.ndl.go.jp/view/download/digidepo\\_8295098\\_po\\_dcndl201112.pdf?contentNo=1](http://dl.ndl.go.jp/view/download/digidepo_8295098_po_dcndl201112.pdf?contentNo=1), (アクセス : 2014.10.18)
- [19] DCMi Metadata Terms.  
<http://dublincore.org/documents/2012/06/14/dcmi-terms/?v=elements#>, (アクセス:2015.1.10)
- [20] Nokogiri. <http://www.nokogiri.org/>, (アクセス : 2015.1.10)
- [21] Selenium WebDriver. <http://docs.seleniumhq.org/projects/webdriver/>, (アクセス : 2015.1.10)
- [22] RDF 1.1 Turtle. <http://www.w3.org/TR/turtle/>, (アクセス : 2015.1.10)
- [23] SPARQL 1.1 Query Language . <http://www.w3.org/TR/sparql11-query/>, (アクセス : 2015.1.10)
- [24] 戦争証言アーカイブス. <http://www.nhk.or.jp/shogenarchives/>, (アクセス : 2015.1.10)
- [25] 東日本大震災アーカイブス. <http://www9.nhk.or.jp/311shogen/>, (アクセス : 2015.1.10)
- [26] NHK 映像マップ みちしる. <http://www.nhk.or.jp/michi/>, (アクセス : 2015.1.10)

- [27] vCard Ontology - for describing People and Organizations .  
<http://www.w3.org/TR/vcard-rdf/>, (アクセス : 2015.1.10)
- [28] FOAF Vocabulary Specification 0.99. <http://xmlns.com/foaf/spec/>, (アクセス : 2015.1.10)
- [29] W3C Semantic Web Interest Group: Basic Geo (WGS84 lat/long) Vocabulary.  
<http://www.w3.org/2003/01/geo/>, (アクセス : 2015.1.10)
- [30] 神崎正英. セマンティック・ウェブのための RDF/OWL 入門. 森北出版. 2005.
- [31] NPO 知的資源イニシアティブ. アーカイブのつくりかた 構築と活用入門. 勉誠出版. 2012.
- [32] 長井暁 (2008) .世界の映像アーカイブの現状と課題. 放送研究と調査 (月報) 2008 年 3 月号,p46-49,NHK 放送文化研究所.