

第5章 コンテンツのメタデータ付与について

メタデータとはコンテンツ自身のことを説明するためのデータのことです。デジタル化したコンテンツにメタデータを付与すると、コンテンツの検索・整理、保管の効率化やアーカイブでの提供に役立ちます。第 5 章は、メタデータとは何か、どのように作成、付与すればよいかを説明しています。「4. 震災関連デジタルアーカイブでの事例」では、運用実証調査の事例を紹介します。

1. メタデータとは

メタデータとは、コンテンツ自身のことを説明するためのデータのことで、本について言えば本の「タイトル」「著者」「出版社」「発行日」「値段」などが該当します。

検索用のキーワードをメタデータとしてあらかじめ付与しておくことで、対象となる情報資源を効率的に検索したり、あるいはコンテンツがデジタルデータの場合は、どのようなフォーマットで保存されているのかをメタデータとして記録しておくことで、再生するための方法を識別したりすることができます。

例えば「地名で検索する」「地図で検索する」などの検索機能は、撮影場所（位置情報）をメタデータとして登録しておくことによって実現される機能です。地図アプリは位置情報を緯度経度で表すため、地図で検索させたい場合には、撮影場所のメタデータは地名ではなく緯度経度を使います。また、地区の名称で検索させたい場合は、地区名を使います。

2. メタデータの作成

メタデータの作成は、コンテンツを整理する際に使った目録³⁶をもとに、アーカイブサイトでどのような検索項目を提供するかを考える際に同時に実施すると比較的容易です。

目録には、コンテンツのタイトルや撮影者・作成者、撮影日・作成日、撮影した場所、作成者の住所、資料の種別（動画、写真、音声、文書、書籍等）、許諾条件などが記載されています。これらは全てメタデータとして活用³⁷できます。また、コンテンツをデジタル化した際のフォーマット種別やデジタルデータ化した日付などもメタデータとしてアーカイブ構築の際に有用です。

震災関連のコンテンツを保存、提供する際には、アーカイブを利用するエンドユーザーのことも考慮すると、いつ、どこで、何が起きたのかがわかるメタデータを作成し、キーワードで簡易に検索できるような値を付与しておくことをお勧めします。例えば、エンドユーザーが、主体（地元住民、ボランティア等）、組織（行政、病院、NPO等）、対象（建物、道路、鉄道等）、営み（避難行動、救援活動等）といった視点で検索・活用することを想定した場合、それらに関連するメタデータを作成することが必要です。

また、コンテンツを管理しアーカイブを構築するために、識別子（コンテンツごとに一意に振られたID）やデータのファイル形式、資料種別（文書、写真、音声、動画等の種別）、権利情報・利用条件、コンテンツの権利に関する項目（著作権者、所有者、提供者）を作成しておくとい良いでしょう。

³⁶ 本ガイドライン 「第2章 被災資料の応急措置、修復、保存について」「第3章 震災関連のデジタルデータ化について」を参照。

³⁷ 本ガイドラインでは、「撮影者」「撮影日」などを「メタデータの項目名（プロパティ）」、「東花子」「2013年4月5日」などそのコンテンツ固有の値を「メタデータの値」と言う。

以下に示すメタデータ項目については、アーカイブの提供及びコンテンツの保存の観点等から、入力することをおすすめする項目です。値の記述形式や詳細な説明は、「参考 6 入力することをおすすめするメタデータ項目の説明」をご覧ください。

表 5-1 入力することをおすすめするメタデータ項目

No.	項目名	項目の説明
1	タイトル	コンテンツの内容を端的に示すタイトル
2	識別子	コンテンツを一意に識別するための ID
3	権利情報・利用条件	アーカイブ構築者、運用者やエンドユーザが、コンテンツをどのような条件で利用できるかの説明
4	撮影者、作成者 (著作権者)	写真や動画の撮影者、文書の作成者、著者などコンテンツを作った人や団体の名前
5	公開者、所有者	著作権者に代わり、コンテンツを公開する人、団体（投稿サイトの運営者や出版社など）
6	提供者	著作権者以外からコンテンツの提供を受けた場合は提供者名や連絡先。利用許諾を受けた場合は権利の所有者名や連絡先
7	撮影日、作成日	写真、動画等の撮影日、文書等の作成日
8	掲載日、公開日	新聞、雑誌等の発行日、サイト等での公開日、放送の場合は放映日
9	撮影場所、作成場所 (地名)	コンテンツが作成された場所の地名、住所 ※地名の場合は、なるべく詳しく記載する ³⁸
10	撮影場所、作成場所 (緯度経度)	コンテンツが作成された場所の緯度経度
11	キーワード、内容	コンテンツの内容を端的に表すキーワード
12	資料種別	文書、写真、音声、動画等、コンテンツの種類
13	ファイル形式	コンテンツのファイル形式（画像ファイルなら JPEG、文書ファイルなら PDF 等）やファイルフォーマットのバージョン（PDF なら PDF6.0 等）
14	説明、要約、注記	サイト上のコンテンツを分類するためのキーワード（簡易検索や分類検索の選択肢に当たるもの）やコンテンツの内容を補足するもの

³⁸ 例えば新潟県中越地震、中越沖地震の資料を集めた「中越災害アーカイブ」では、地名検索の際に市町村名ではなく、中越地方で日常的に使われている地域名で検索できるようにしている (<http://map.c-bosai-anzen-kikou.jp/main?m=10030&catt=002000000000052>)。

メタデータ（項目名（プロパティ）と値）は、アーカイブを構築・運営する人がわかりやすい言葉で入力しても構いませんが、他の機関と情報交換したりシステム連携したりすることを考えると、標準的な項目や値、広く普及しているコード体系（緯度経度（世界測地系）、全国地方公共団体コード³⁹等）などがあれば、それを使うことをお勧めします。

NDL 東日本大震災アーカイブでは、国内外の多くの震災関連デジタルアーカイブを参考に作成しましたので、NDL 東日本大震災アーカイブのメタデータ項目⁴⁰を参考にすることも一つの方法です。NDL 東日本大震災アーカイブと連携した際に、検索性が向上します。

また、コンテンツを公開したのち、検索されたキーワードのログ等を保存することにより、それらを分析、活用して、検索の利便性を向上させることも可能です。

なお、デジタルアーカイブを構築する際には、採用するメタデータの項目や、その項目の出現レベル（必須、あれば必須、任意など）、項目の繰り返しの有無、データの入力形式（自由記述の文字列か、コード値か、統制語彙からの選択か、URL のみかなど）などを定める必要があります⁴¹。

3. メタデータの付与

メタデータは、全てのコンテンツに付与し、かつ、全ての項目に値を付与すると検索の精度や性能が向上するという利点がありますが、運用実証調査の事例にみるように、メタデータを付与する作業には時間も費用もかかりますので、どのような検索機能を用意するか、どのくらいの人手や予算をかけられるか等を検討して、付与作業の計画を立てましょう。なるべく多くのコンテンツに効率よくメタデータを付与するためには、必須項目や優先順位を決めておくと良いでしょう。

また、同じコンテンツに対しても、付与する作業員によって異なる値を付与することがあります。いくつかサンプルを用意して、値の基準を決めたり、選択肢を用意してその中から値の一つ選ぶことができるようにしたりしておくなどの工夫をすると、人によるブレを最小限に防ぐことができるでしょう。

³⁹ 財団法人 地方自治情報センターが、地方公共団体の電算主管課の便宜を図るために、平成 9 年（1997 年）12 月 10 日現在で公開し、現在まで継続的にメンテナンスを行っているもの。

⁴⁰ 「参考 5 NDL 東日本大震災アーカイブメタデータスキーマの設計方法」参照。

⁴¹ 採用するメタデータの項目や、その項目の出現レベル、項目の繰り返しの有無、データの入力形式などを定める際には、NDL 東日本大震災アーカイブが定めている「NDL 東日本大震災アーカイブメタデータスキーマ」（以下「NDL 大震災メタデータスキーマ」と言う）を参考にすることができる。<http://kn.ndl.go.jp/static/api> に掲載予定（NDL 東日本大震災アーカイブ HP で最新掲載 URL を確認してください）。なお、運用実証調査にて作成した「連携用メタデータスキーマ」は、「NDL 大震災メタデータスキーマ」に準拠している（「参考 7 連携用メタデータスキーマ」参照。）。

4. 震災関連デジタルアーカイブでの事例

運用実証調査では、震災関連資料・記録の収集、デジタル化後、コンテンツについてメタデータ付与を実施しました。

本項の(1)では、運用実証調査で実施したメタデータ付与の手順について、標準的に実施した内容を説明し、(2)以降では、各地域での実証調査の事例を説明します。

(1) 運用実証調査におけるメタデータ付与

(a) メタデータ付与手順

運用実証調査では、メタデータの内容にばらつきが出ないよう一定の品質を確保したものとするため、「メタデータ付与→検証→メタデータ再付与」の手順を繰り返し実施しました。効率的にメタデータ付与を実施するためには、作業を実施する前に「作業設計」の工程を実施することが大切です。この工程を実施することで、検索性に優れたメタデータを付与することが可能となります。

運用実証調査結果を踏まえた作業手順は、以下のとおりとなります。

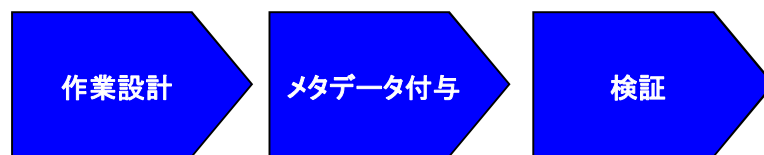


図 5-1 メタデータ付与手順

- 「作業設計」では、作業手順を作成し、一度試験的にメタデータ付与作業を行った上で、メタデータの内容にばらつきが出ないよう、手順書の記載内容、留意事項等を見直します。この工程を実施することで効率的にメタデータ付与を実施することができます。
- 「メタデータ付与」では、「作業設計」で作成した方針、手順書、留意事項等に沿ってメタデータを付与する体制を構築し、管理者と作業者がメタデータ項目等について適宜議論を行いながら実施します。議論を行うことで主観的な判断をなくし、メタデータの品質を高めることができます。
- 「検証」では、コンテンツとメタデータの整合性を確認し、プライバシーや描写の観点から不適切なものが無いかのチェックを行います。検証を実施することでメタデータの品質をさらに高めることができます。

(b) メタデータの付与項目

付与すべきメタデータの項目は、アーカイブ利用者がメタデータを検索する際にどのような項目が必要かを想定し、決定しなければなりません。

運用実証調査で付与したメタデータ項目は、コンテンツ管理の観点とコンテンツ検索の観点から、決定しました。

- ・ コンテンツの管理の観点

「作成者」、「作成日」、「提供者」、「権利関係」、「フォルダ名/ファイル名」などを必須項目とし、「提供者区分」、「サイズ」については、管理の目的に鑑み必要に応じて付与しました。

- ・ コンテンツの内容の観点

「タイトル」、「説明・要約」、「分類」、「場所」、「時期」、「内容（キーワード）」などの項目を設定し、必要に応じて付与しました。

なお、「内容（キーワード）」は、利活用時の検索性に大きく影響する項目であることから、入力にばらつきが出ないように、付与するための方針（ポリシー）を策定し、複数名で検証しながら作業を実施しました。

- ・ NDL 東日本大震災アーカイブとの連携の観点

NDL 東日本大震災アーカイブとの連携を行うため、NDL 大震災メタデータスキーマに準拠したメタデータ項目としました。

以下に各プロジェクトで使用したメタデータ項目について示します。

表 5-2 運用実証調査で使用したメタデータ項目

大分類	中分類	青森	岩手	宮城 (東北大学)	宮城 (河北新報)	福島
管理	作成者	◎	◎	◎	◎	◎
	作成日	○	○	◎	◎	◎
	提供者	◎	◎	◎	◎	◎
	提供者の区別（大）	◎	◎	***	***	◎
	提供者の区別（中）	◎	◎	***	***	○
	提供者の区別（小）	○	○	***	***	○
	権利関係	◎	○	○	○	○
	フォルダ名/ ファイル名	◎	◎	◎	◎	◎
サイズ	***	***	○	○	◎	
内容	タイトル	○	◎	◎	◎	◎

大分類	中分類	青森	岩手	宮城 (東北大学)	宮城 (河北新報)	福島
	説明・要約	○	○	○	○	○
	分類	◎	◎	○	○	◎
	場所	◎	◎	○	○	◎
	時期	○	◎	○	○	◎
	内容 (キーワード)	○	○ 場所、時期は 自動付与。 その他は、任 意	○ 対象 (被写 体) 対象 (文字) 被害等 復旧・復興等 災害原因 災害状況 撮影の意図	○ 東北大学と 同様 ※資料種別 として「新 聞」を追加	◎
その他	必須項目	言語 資料種別 メタデータ ID (システム 自動付与)	言語 資料種別 レコード ID 資料 No 場所よみ	***	***	自治体の属 性情報

(注) 「◎」: 必須項目。「○」: 任意項目。「***」: 未定義。

各プロジェクトで使用したメタデータは、NDL 東日本大震災アーカイブと連携する際に「連携用メタデータスキーマ」⁴²に変換して提供しました。

(c) メタデータ付与の作業体制・作業効率

運用実証調査で実施したメタデータ付与作業について、どのような作業体制で実施し、どれくらいの作業効率で実施したかを以下に示します。

メタデータ付与は、簡単な指導を行うだけで、誰でもできる作業であったことから、作業を実施する上では、作業には特別なスキルは必要ありません。ただし、付与するメタデータの内容によっては、地元の人でなければ分からない内容（地名、伝統行事等）、専門家でなければ判断できない内容等が存在するため、メタデータ付与作業の体制には、こうした判断ができる人員を作業体制に追加することも検討しなければなりません。

⁴² 連携用メタデータスキーマは、NDL 東日本大震災アーカイブが定義する NDL 大震災メタデータスキーマに準拠しており、NDL 東日本大震災アーカイブとの連携が実証されている。連携用メタデータの詳細は、「参考7 連携用メタデータスキーマ」を参照。

- 「場所」や「キーワード」には、その地方での呼び名や利用者にとって検索しやすい内容がメタデータとして付与されていることが重要であることから、作業体制に地元の人を入れることが望ましい。
- 「キーワード」や「説明・要約」については、専門家がコンテンツを利用する場合には専門用語などを、付与しておくことが望ましいが、こうした利用方法は限定的になること、付与できる専門家の確保が困難であることから優先度は低い。

表 5-3 メタデータ付与作業体制・作業効率

地域	役割	人数	人員/スキル	作業効率
青森	・ 管理者 作業指示、監修	2名	アーカイブの専門知識を有する者	40～60 件/時間
	・ 作業員 メタデータ付与	1名	現地で雇用したパート職員 (特別なスキル無し)	
岩手	・ 管理者 作業指示作成、コンテンツ整理、進行管理	2名	・ 高度な IT スキルを有し、オペレータの問い合わせや不具合に対処できる人員 ・ 様々なソフトウェアを駆使し、PC コマンド/システム面のサポートができる人員	6～20 件/時間
	・ 作業員 メタデータ付与、1次検証	18名	・ PC で Excel の入力ができる人員 ・ 管理者の指導に基づいて、支障なく PC 操作及び入力作業ができる人員	・ 文章： 10～20 件/時間 ・ 画像 4～15 件/時間
	・ 最終検証者 1次検証、最終検証、進行管理補佐	3名	・ 運用主体としてコンテンツ/メタデータの品質保証ができる人員	6～200 件/時間
宮城 (東北大学)	・ 作業員 メタデータ付与	7名	《地域性》 ・ 検索ツール (Google) 等を駆使すれば、地域に詳しい人でなくとも付与は可能である 《専門性》 ・ 専門性のあるコンテンツ (研究者) については、専門スキルが必要となるケースもある	44 件/時間

地域	役割	人数	人員/スキル	作業効率
宮城 (河北 新報社)	・ 作業者	5名	事前にツールの簡単な操作を教えることで、作業は可能である	約 18 件/時間
	・ 管理者	1名	コミュニケーション能力があり、土地勘がある者	
福島	・ 管理	1名	多岐にわたる内容の文書を理解できる幅広い知識	6~60 件/時間
	・ オペレータ メタデータ付与	3名	大学院等の学生で、浜通りに一定の土地勘を有する者	

(2) 青森県における実証調査

(a) メタデータ付与の手法・プロセス

青森県では、以下の手順でメタデータ付与を実施しました。

① メタデータ付与手順の詳細

青森県では、メタデータの付与順序について以下の2つのステップで実施しました。

➤ STEP1: 機械的に入力可能な項目の付与

「作成者」、「提供者」といった機械的に入力可能な項目のみ付与しました。

➤ STEP2: 入力にばらつきが出やすい項目の付与

「場所」、「作成時期」のように人によって入力方法にバラつきが出やすい項目であるため、選択式で付与することとしました。

- ・ 「場所」：写っている施設（学校、病院、工場等）、屋外／屋内、具体的な場所（事務所、小屋、自転車置き場、道路等）を選択肢としました。
- ・ 「作成時期」：震災前／震災直後／復旧段階／復興段階を選択肢として付与しましたが、コンテンツの内容から判断すると主観的要素が強くなるため、基本的には作成日を基準とすることとしました。

STEP2 の作業完了後、作業管理者による検証作業を実施し、「分類」について、設定が必要と判断し、登録を実施しました。

各 STEP で入力したメタデータの入力例について以下に示します。

《STEP1 で付与したメタデータ項目例》

【作成者】

項目名	作成者		
最大文字数	1000 文字		
繰り返し	最大	10	最小 1
入力必須	1 番目は必須、2 番目以降は任意		
入力制限	なし（任意の文字種別）		
説明	<p>コンテンツの作成者（著者、撮影者等）となる値を設定する。</p> <p>作成者に対応するよみがある場合は、作成者と作成者よみをセットとし、最大で 10 セットまで繰り返して入力する。</p> <p>作成者、撮影者が明確な場合は記載し、特定できない場合は提供機関名とする（※1）。不明の場合は「作成者不明」とする。</p> <p>作成者が個人の場合、個人名は入力せずに「個人」とする（※2）。</p> <p>※1 機関名は正式名称で入力すること。ただし、以下の略号を用いることとする。</p> <p>※2 青森プロジェクトでは、個人情報とは別に管理しており、メタデータには個人情報を掲載しない方針である。個人からのコンテンツ受領の際は、機関を経由しているため、提供機関名で入力する。</p> <p>株式会社 → （株）</p> <p>有限会社 → （有） 合同会社 → （同）</p> <p>合名会社 → （名） 合資会社 → （資）</p> <p>財団法人 → （財）</p> <p>社団法人 → （社） 医療法人 → （医）</p> <p>医療法人社団 → （医社）</p> <p>社会福祉法人 → （社福） 学校法人 → （学）</p> <p>独立行政法人 → （独法）</p> <p>宗教法人 → （宗）</p> <p>特定非営利活動法人 → （NPO）</p>		
入力例	<p>■図書・出版物の場合：著者、翻訳者、共著者、編者等を入力する。</p> <p>例）大日本水産会東日本大震災対策本部、大日本水産会東日本大震災対策本部</p> <p>■行政文書の場合：作成部局名を入力する。</p> <p>例）大蔵省大臣官房政策金融課</p> <p>■写真の場合：撮影者を入力する。（個人名は不可）</p> <p>例）個人</p> <p>■音声・動画の場合：撮影者、録音者を入力する。</p> <p>例）（株）東奥日報社</p>		

【提供者】

項目名	提供者
-----	-----

最大文字数	1000 文字			
繰り返し	最大	1	最小	1
入力必須	必須			
入力制限	なし（任意の文字種別）			
説明	<p>コンテンツの提供者を入力する</p> <p>提供者に対応するよみがある場合は、提供者と提供者よみをセットとする。</p> <p>本項目に入れるのは最終提供者とする。（例えば、個人から提供を受けた機関から提供された場合は、機関名を本項目に設定する。最終提供者までの経路については権利関係の書類（利用許諾書）に詳細を記述する。）</p> <p>あくまでも「最終提供者」を記入し、収集者とはしないこと。</p> <p>※あおりデジタルアーカイブ・コンソーシアム及びシステム開発会社に対してコンテンツの使用を許諾した機関・個人が「最終提供者」となる。</p>			
入力例	例) 国立国会図書館、青森県八戸市広報統計課、個人			

《STEP2 で付与したメタデータ項目例》

【場所】

項目名	場所			
最大文字数	500 文字			
繰り返し	最大	5	最小	1
入力必須	1 番目は必須、2 番目以降は任意			
入力制限	なし（任意の文字種別）			
説明	<p>コンテンツの作成に関わる場所の情報（地名、地域等）を入力する。</p> <p>場所の詳細が不明な場合は、最低限の地域情報として収集場所の都道府県を設定する。</p> <p>場所 1 は以下から選択する。</p> <p>「八戸市」「おいらせ町」「階上町」「三沢市」「青森市」「その他青森県内」「青森県外」</p> <p>場所 2～5 は、詳細な場所情報があれば入力する。</p>			
入力例	<p>場所 1 : 八戸市</p> <p>場所 2 : 八戸市大字河原木字北沼、</p> <p>場所 1 : 青森県外</p> <p>場所 2 : 宮城県仙台市若林区卸町</p>			

【作成時期】

項目名	作成時期			
最大文字数	100 文字			

繰り返し	最大	1	最小	1
入力必須	必須			
入力制限	指定された値から選択			
説明	<p>コンテンツ作成時期の段階を設定する。</p> <p>コンテンツの作成日（撮影日）により「震災前」「震災直後」「復旧段階」「復興段階」から選択する。</p> <p>コンテンツ作成日が不明のものは、内容よりいずれかを選択すること。</p> <p>コンテンツの作成日（撮影日）による選択基準は以下のとおり。</p> <p>震災前 : 震災発生以前（2011年3月11日の地震発生より前）</p> <p>震災直後 : 地震発生より2011年3月31日まで</p> <p>復旧段階 : 2011年4月1日より2011年12月20日まで</p> <p>復興段階 : 2011年12月21日以降</p>			
入力例	震災前、震災直後、復旧段階、復興段階			

《STEP2 完了後、入力したメタデータ項目》

【分類】

項目名	分類		
最大文字数	100 文字		
繰り返し	最大	5	最小 1
入力必須	1 番目の分類は必須、2 番目以降は任意		
入力制限	指定された値から選択		
説明	<p>コンテンツの分類を設定する。</p> <p>選択肢より該当する分類を最大 5 つまで付与可能とする。</p> <ul style="list-style-type: none"> • コンテンツ内容により以下からを選択する。複数選択可能とする。 • 被害 : 被災状況の記録 • 防災 : 防災関連情報 • 復旧 : 復旧・復興作業の様 • ボランティア : ボランティア活動に関する情報 • 支援 : 復旧・復興支援に関する情報 • 行政 : 行政からの公開情報、行政の動き • 避難所 : 避難所関連情報 • 医療 : 医療・介護関連情報 • 報道 : 報道、メディア関連情報 • 生活 : 生活関連情報 • 学校・教育機関 : 小中学校、高校、大学・短大・専門学校 • 産業 : 企業活動に関する情報 		
入力例	<p>■図書・出版物の場合 :</p> <p>例) 報道、被害、行政、避難所、学校・教育委期間</p> <p>■行政文書の場合 :</p> <p>例) 防災、行政</p> <p>■写真の場合 :</p> <p>例) 復旧、ボランティア、生活</p> <p>■音声・動画の場合 :</p> <p>例) 被害</p>		

(b) メタデータ付与に使用した機器等

- メタデータ付与に関しては、Excel を使用しました。
- 写真のメタデータ付与にあたっては、画像ビューワとして、操作性の良いフリーソフト (Irfan View) を使用しました。

(c) 課題と対応策

メタデータ付与に関する課題と対応策については、以下のとおりです。

① 写真の位置情報について

【課題内容と対策案】

写真については、ほとんど位置情報が付加されておらず、提供者、及び、明確に判別可能な被写体の情報、大まかな撮影時期等でしかメタデータの付与ができなかったため、メタデータがほとんど付与できず、付与されているメタデータも多くのコンテンツにおいて重複したものとなってしまいました。

写真は、メタデータでの検索よりも地図を利用したサムネイルによる一覧表示の方が、利用者には分かりやすいことが利用者アンケートからも示されているため、今後、位置情報を何らかの形で付与することを検討していく必要があります。

② 「作成時期」について

【課題内容と対策案】

青森プロジェクトでは、調査テーマとして、コンテンツの作成時期を「被災前」「震災直後」「復旧段階」「復興段階」の4つに分類することとしていましたが、実際には、コンテンツの内容からはどの分類に属するかには判断がつかなかったため、コンテンツの作成時期によってどの分類に属するかを判断し、これをメタデータとして付与しました。

これを検索キーワードの一つとしており、作成時期によるコンテンツ検索が可能となっています。ただし、このようなメタデータは利用目的等に依存するため、今後新たな利用目的・方法等が考えられた場合には新たにメタデータを追加する必要性が生じる可能性があります。

③ メタデータ付与作業者のスキルと付与されたメタデータの検索性の確保

【課題内容と対応策】

- メタデータの付与作業者は、図書館の書誌入力や所蔵目録作成作業の経験者が理想的ですが、専門業者へ委託することとなり、費用も増加してしまいます。
- 運用実証調査では、現地雇用及び継続した運用を考慮する場合、安価で「誰にでもできる」メタデータ付与作業のプロセス確立が求められます。その際の最低限の必要スキルとしては、作業効率化のための基本的なパソコン操作スキルと、管理者とのコミュニケーションを円滑化するためのインターネット、メールを用いたコミュニケーションスキルとなります。
- 青森プロジェクトでは、管理者が作業場に駐在する時には、作業員からの問い合わせに回答でき、解決が図れますが、管理者が不在の時には、メタデータの揺れや入力ミスが発生し、手戻り作業の原因となり、結果として想定外の工数が発生しました。

(3) 岩手県における実証調査

(a) メタデータ付与の手法・プロセス

岩手県のメタデータ付与作業の手法・プロセスについて、以下に示します。

① メタデータ付与の詳細

a. 作業設計

作業設計では、メタデータ付与作業にばらつきが出ないように以下のような作業指示書を作成しました。

提供者	遠野まごころネット	資料No.	100005
コンテンツ概要	震災後の大槌町の画像		
メタデータ付与方針	1画像1メタ		
コンテンツ選択基準	意図の伝わらない画像は除外する、連続した似たような画像は除外する		
コンテンツ加工の有無	有 (無)		
コンテンツ整理方法 (フォルダ/ファイル構造)			
メタデータ詳細	タイトル	場所+状況説明 <NG>単語のみの表記にはしない	
	タイトルよみ	「タイトル」をカタカナで記入	
	作成者1	遠野まごころネット	
	作成者よみ1	トオノマゴコロネット	
	提供者	遠野まごころネット	
	提供者よみ	トオノマゴコロネット	
	説明・要約	画像を確認しタイトルをより詳細に説明できる文章を作成する	
	場所1	写真から判定	
	コンテンツ作成日(自)	Null	
	コンテンツ作成日(至)	画像のプロパティより、撮影日を記入	
	時期	コンテンツ作成日(至)から判断	
	個人特定情報	プライバシー判定基準に準ずる	
	機関特定情報	プライバシー判定基準に準ずる	
	公開NG描写		
	権利関係	「個人特定情報」「機関特定情報」「公開NG描写」に準じて自動記入	
	著作権者名	遠野まごころネット	
	二次利用可否	否	
	コンテンツ無償アクセス(実証調査期間中)	Null	
	コンテンツ無償アクセス(実証調査期間後)	Null	
	メタデータアクセス制限(実証調査期間中)	特定の組織、コミュニティへの公開	
	メタデータアクセス制限(実証調査後)	特定の組織、コミュニティへの公開	
	コンテンツアクセス制限(実証調査期間中)	特定の組織、コミュニティへの公開	
	コンテンツアクセス制限(実証調査後)	特定の組織、コミュニティへの公開	
コンテンツアクセス期間制限	Null		
キーワード	共通: jpg 画像 個別: 画像を見て判断(地区名や店舗名)		
注記	Null		
形態	その他		
数量	1		
分類1	写真から判定し記入		
言語	jpn		
資料種別	写真		
	必須項目	共通メタデータ	
検証手順	<p>①入力シートを開きCtrl+Qでメニュー立ち上げ、「入力データ検証」を実施</p> <p>②メタデータ「フォルダ/ファイル名」のセル上で、Ctrl+Rにより該当コンテンツを呼び出し画像とメタデータを照らし合わせ以下を確認</p> <ul style="list-style-type: none"> ・タイトル ・説明・要約 ・キーワード ・個人特定情報 ・機関特定情報 ・公開NG描写 <p>※上記において、修正箇所が発生した場合はセルの色を変更 修正済みセル →  未修正セル → </p>		

図 5-2 メタデータ付与作業指示書サンプル

また、「分類」「提供者の区分」「場所」等の共通入力項目は、作業者によってばらつきが出ないように専用シートを使用し、選択肢から選択させるよう設計しました。

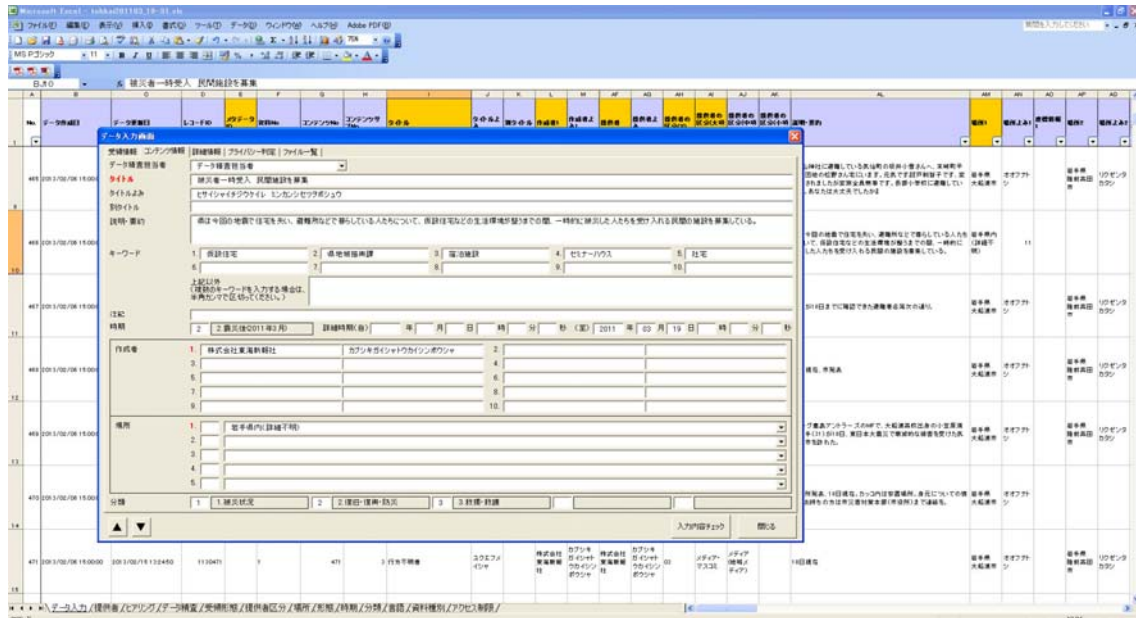


図 5-3 メタデータ入力専用シート

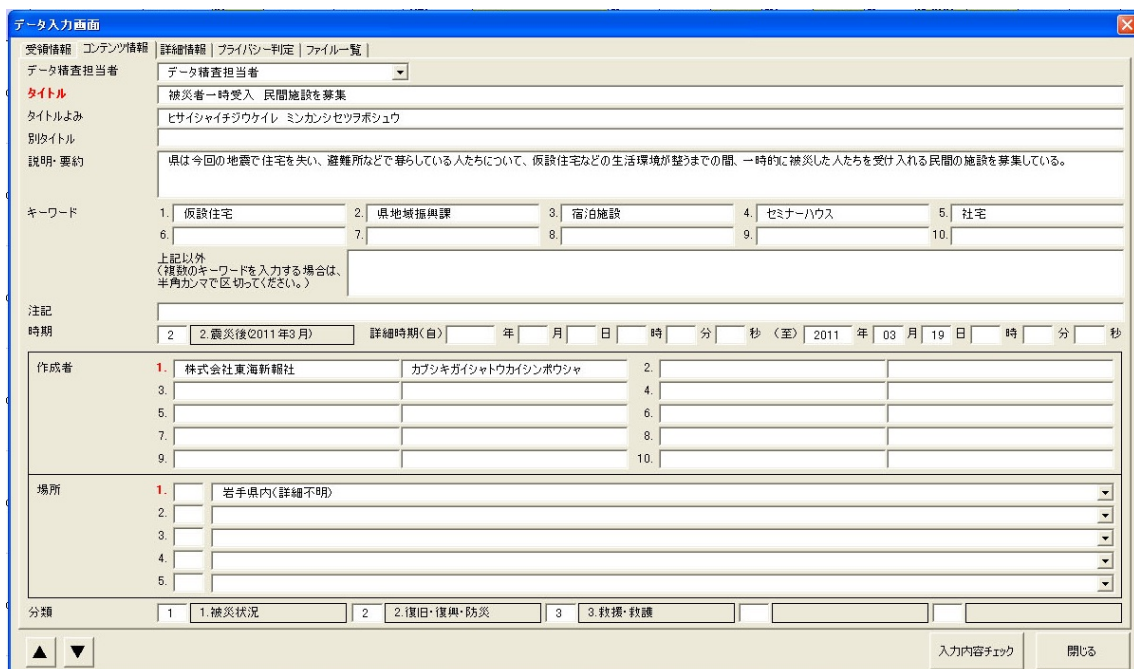


図 5-4 メタデータ入力ツール

分類	中分類CD	中分類
1	1. 被災状況	01 地理・地形・測量データ
		02 災害・被災箇所
2	2. 復旧・復興・防災	01 復興計画
		02 仮設住宅
		03 仮設商店街
		04 復興住宅
		05 写真関連活動
		06 プロジェクト
		07 防災
3	3. 救援・救護	01 ボランティア活動
		02 医療支援
		03 企業支援
		04 団体支援
		05 海外支援
		06 消防
		07 自衛隊
		08 避難所
4	4. 暮らし・まち・風景	01 なつかしいまちなみ
		02 あれからのまちなみ
		03 風景
		04 まちのこえ
		05 祭り・イベント
		06 学校
		07 コミュニティ
5	5. 観光・産業	01 農業
		02 漁業
		03 地場産業
		04 その他の産業
		05 新たな産業(復興後)
		06 観光・交流
		07 特産、てしごと
6	6. 自然・歴史・文化	01 地図
		02 芸術・文化遺産
		03 民族・芸能
		04 風俗・慣習
		05 衣食住文化
		06 農漁村文化
		07 産業
		08 自然
		09 昔話、津波教訓

提供者の区分 (大項目)	提供者の区分CD	提供者の区分 (中項目)
1 個人	01	市民
2 メディア・マスコミ	02	メディア(全国メディア)
	03	メディア(地域メディア)
	04	メディア(海外メディア)
	05	インフラ企業
3 民間企業	06	地域企業
	07	企業
	08	金融機関
	09	経済団体
	10	新規起業
	11	医療・救護
	12	農協、漁協
	13	その他 民間企業
4 公共機関・団体	14	日本政府/中央省庁
	15	地方自治体
	16	公共施設
	17	諸外国
	18	その他 公共機関・団体
5 NPO・NGO等	19	社会福祉協議会(ボランティアセンター)
	20	NPO、NGO、一般社団法人 他
6 学術研究	21	全国 大学/公的研究機関/企業研究機関
	22	外国 大学/公的研究機関/企業研究機関
	23	その他 学術研究機関

図 5-5 「分類」「提供者の区分」のサンプル

あ	岩手県 一関市	イチノセキシ
	岩手県 一戸町(二戸郡)	イチノヘマチ(ニノヘグン)
	岩手県 岩泉町(下閉伊郡)	イワイズミチヨウ(シモヘイグン)
	岩手県 岩手町(岩手郡)	イワテマチ(イワテグン)
	岩手県 奥州市	オウシユウシ
	岩手県 大槌町(上閉伊郡)	オオツチチヨウ(カミヘイグン)
	岩手県 大槌町(上閉伊郡)赤浜	オオツチチヨウ(カミヘイグン)アカハマ
	岩手県 大槌町(上閉伊郡)安渡	オオツチチヨウ(カミヘイグン)アンド
	岩手県 大槌町(上閉伊郡)大槌	オオツチチヨウ(カミヘイグン)オオツチ
	岩手県 大槌町(上閉伊郡)大町	オオツチチヨウ(カミヘイグン)オオマチ
	岩手県 大槌町(上閉伊郡)大ケ口	オオツチチヨウ(カミヘイグン)オガクチ
	岩手県 大槌町(上閉伊郡)金沢	オオツチチヨウ(カミヘイグン)カネザワ
	岩手県 大槌町(上閉伊郡)上町	オオツチチヨウ(カミヘイグン)カミチヨウ
	岩手県 大槌町(上閉伊郡)吉里吉里	オオツチチヨウ(カミヘイグン)キリキリ
	岩手県 大槌町(上閉伊郡)小鎧	オオツチチヨウ(カミヘイグン)コヅチ
	岩手県 大槌町(上閉伊郡)栄町	オオツチチヨウ(カミヘイグン)サカエチヨウ
	岩手県 大槌町(上閉伊郡)桜木町	オオツチチヨウ(カミヘイグン)サクラギチヨウ
	岩手県 大槌町(上閉伊郡)新町	オオツチチヨウ(カミヘイグン)シンチヨウ
	岩手県 大槌町(上閉伊郡)新港町	オオツチチヨウ(カミヘイグン)シンミナトマチ
	岩手県 大槌町(上閉伊郡)末広町	オオツチチヨウ(カミヘイグン)スエヒロチヨウ
	岩手県 大槌町(上閉伊郡)須賀町	オオツチチヨウ(カミヘイグン)スカチヨウ
	岩手県 大槌町(上閉伊郡)本町	オオツチチヨウ(カミヘイグン)ホンチヨウ
	岩手県 大槌町(上閉伊郡)港町	オオツチチヨウ(カミヘイグン)ミナトマチ

図 5-6 「場所」のサンプル

b. メタデータ付与

メタデータ入力作業について、文章を中心としたコンテンツ（新聞記事・書類等）については、メタデータ入力画面とコンテンツを表示させる画面を同時に表示させることで、整合性を確保しました。（図 5-7 参照）

また、オペレータが上述の作業指示に基づいて入力を実施しました。



図 5-7 メタデータ入力風景

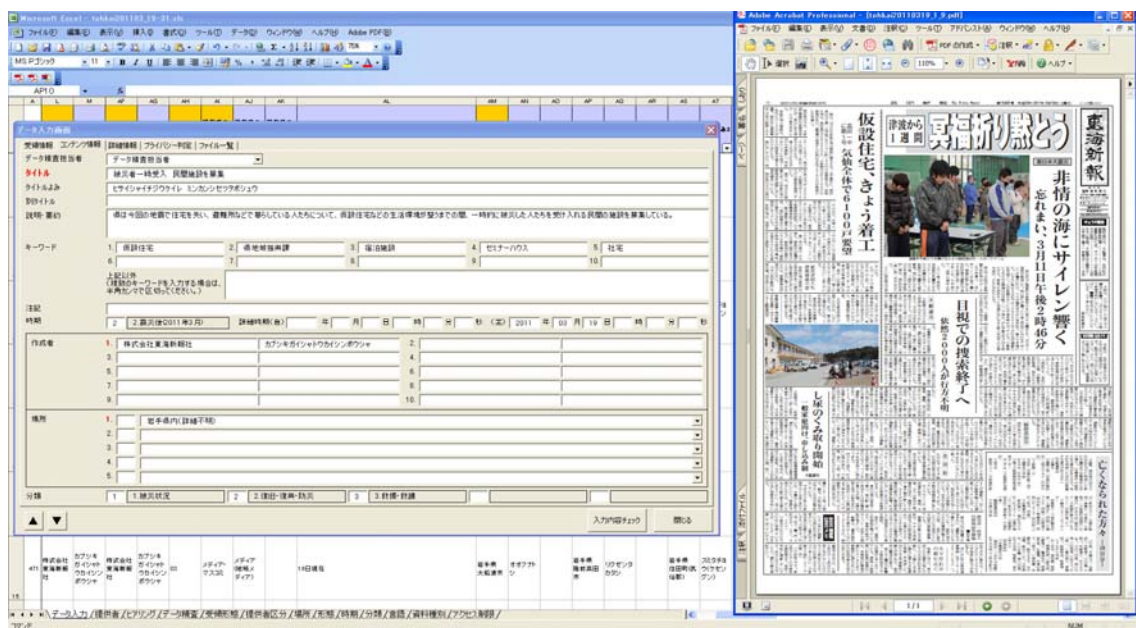


図 5-8 メタデータ入力画面

c. 検証

検証シートを用い、入力必須項目のうち共通項目で無いもの（タイトル・タイトルよみ・場所・コンテンツ作成日・キーワード）及び、プライバシーチェック（個人特定情報、機関特定情報、公開 NG 描写等）について、メタデータ付与内容が適切か否かを作業員以外の第三者が確認し、チェックマーク入力を実施しました。

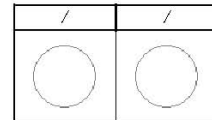
さらに、「個人情報特定」「機関特定」「公開 NG」といった情報については、描写の有無を岩手プロジェクトにおいて策定した品質管理項目と照らし合わせ最終確認を行いました。

(b) メタデータ付与に使用した機器等

メタデータの付与作業には Excel を使用しました。さらに、作業効率を上げるため、メタデータの設計に併せて、全てのコンテンツのメタデータを共通で入力できるツールを作成しました。ツールの使用により、項目によってはプルダウン形式で設計し、入力の手間を省くと同時に、作業員が多数になっても統一のフォーマット入力・データ管理が可能となりました。

入力画面のイメージは、以下のとおりです。

レコードID		資料No		検証	
コンテンツNo		コンテンツサブNo			
タイトル					
タイトルよみ					
説明・要約					
キーワード					
場所1					
場所2					
分類1		分類2			
分類3		分類4			
分類5		時期			
機関特定情報					
個人特定情報					
公開NG描写					



img1.jpg



1/1

図 5-9 メタデータ入力画面イメージ

(c) 課題と対応策

メタデータ付与に関する課題と対応策については、以下のとおりです。

① 地元住民オペレータのメタデータ付与作業実施について

【課題内容と対応策】

メタデータ付与については以下の3つのパターンを想定していました。

- i. コンテンツ提供先でメタデータをコンテンツ毎に付与
- ii. コンテンツ収集時にメタデータ付与のための基礎情報をヒアリングし、収集後にメタデータを付与
- iii. 提供者からの基礎情報なしでメタデータを付与

提供者が付与、又はヒアリングして収集者が付与するパターンは、提供者側の負担を伴うので、状況に合わせて対処が必要です。

提供者からの情報無しでメタデータ付与を行う場合は、オペレータは、地元雇用の人材が適していました。特に、写真や映像のみから内容を判断し、メタデータを付与する場合は、「タイトル」「場所」「説明・要約」「キーワード」について、精度が高いメタデータ付与が可能と考えられます。

反面、オペレータの中には震災被害に遭った地元の住民が多いため、被災時の写真を見ることで精神的苦痛を伴う方もおり、メンタル面での配慮が必要と考えられます。

② メタデータのばらつき

【課題内容と対応策】

メタデータ付与作業においてオペレータの入力内容にばらつきが見受けられました。

a. 使用している語彙にばらつきがある場合

「がれき、ガレキ、瓦礫」「車、自動車、車両」といった例があげられます。統一することで、情報が整理されますが、検索ヒット率が下がる可能性もあります。似通った内容のコンテンツに対しては、オペレータがメタデータ付与作業前に、共通キーワードとして多様な語彙を設定するといった対応策が考えられます。

b. コンテンツ内容の理解度によるばらつきがある場合

コンテンツの内容の理解度に差があると、同一のコンテンツに対して各オペレータが具体的な名称と一般的な名称（「国道45号線」と「道路」等）を付与する例が見受けられました。

上記①の課題にも関連しますが、具体的な名称を入れられない場合は、現地に詳しいオペレータ等に聞いて入力するフローにすることである程度ばらつきが抑えられると考えられます。

(4) 宮城東北大学における実証調査

(a) メタデータ付与の手法・プロセス

宮城東北大学におけるメタデータ付与作業の手法・プロセスについて、以下に示します。

① メタデータ付与の詳細

メタデータの付与において、データの種類によって、機械的に自動でメタデータが付与される「自動タギング」、作業員が手動で付与を行う「手動タギング」、利用者を含めて広く付与を許可する「ソーシャルタギング」の3つに分かれます。メタデータ付与の全体の方針は以下の表のとおりです。

表 5-4 メタデータ付与方針

情報	自動	手動タグ	ソーシャルタグ
時間情報	写真画像の EXIF から撮影日時刻を取得	文章などに時間記載があるもの	—
空間情報	写真画像の EXIF から緯度経度情報を取得し、住所情報への変換	文章などに住所情報があるもの	写真画像などの場所が特定できないもの
テーマ（内容、対象）	—	写真画像の場合、文字として読み取れるもの、文章などは、文章から抽出	曖昧な情報、不確かな情報など

宮城東北大学プロジェクトでは、自動タグと手動タグについてメタデータ付与を行いました。ソーシャルタグについては、一部のコンテンツについて検証したところ、不確かな情報が混在することが明らかになりました。

さらに宮城東北大学プロジェクトでは、多くの写真を取り扱うことと、撮影者とは異なる第三者の作業員がキーワードを付与する作業となるため、写真に対しての手動でのキーワード付与の基本方針を定め、作業者に周知を行いました。

② キーワード付与の事例

手動での写真のキーワード付与の基本方針の決定に当たり、事前に複数の人が同じ画像に対してキーワード付与する検証を行いました。キーワード付与の検証の結果は、撮影者とは異なる作業員がキーワードの付与を行った場合、憶測や推測によるキーワードが付与されていることがわかりました。また、作業員によっては、見落としなどでキーワードが不足していることも確認されました。これらの結果から、憶測や推測を排除し、写真から見て取れる内容をキーワードにする方針とし、見落としが無いように複数のチ

チェック体制を設けることとしました。

(例 1) 思い込みによるキーワードの違い (1)



【検証前のキーワード】

海岸林、防災林、防潮林、青空、更地、跡地、田畑、平野、地面、津波被害、塩害

【検証後のキーワード】

地面、木、林

(例 2) 思い込みによるキーワードの違い (2)



【検証前のキーワード】

キャベツ、看板、田畑、農地、メッセージ、絆、塩害

【検証後のキーワード】

キャベツ、ブルーシート、土手、畑、絆

例 1、例 2 の事前検証では、津波被害をイメージし、被害後の状況を連想する語彙をキーワードとして付けています。この写真は、位置情報が正確にわからないため、沿岸部の写真かどうか判断がつかず、思い込みでキーワードを付与するという結果が得られました。

(例 3) ボキャブラリに起因するキーワードの違い (1)



【検証前のキーワード】

掲示物、寄せ書き、応援メッセージ、消防署、神栖市立太田小学校、がんばろう

【検証後のキーワード】

がんばろう日本、平成23年6月、神栖市立太田小学校、児童・教諭一同、応援メッセージ、掲示物、消防署の方へ

(例 4) ボキャブラリに起因するキーワードの違い (2)



【検証前のキーワード】

寺社、神社、地藏菩薩、鳥居、浪分神社

【検証後のキーワード】

お地藏様、フェンス、仙台市、好日庵、供花、建物、弊殿、注連縄、浪分神社、看板、賽銭箱、鳥居

例 3、例 4 から見て取れるように、写真の内容から判断できるキーワードがあるにも関わらず拾い切れていないケースが確認できました。これは、作業者がキーワードとして判断する基準が異なることと、個人のボキャブラリの問題が起因していることが挙げられます。本プロジェクトにおいては、見た写真の内容から判断できる、文字、建造物や物の名称は、必ずキーワードとしてあげることとし、複数人によるチェックを設けることとしました。

このような結果から、宮城東北大学プロジェクトでは手動による写真のキーワード付与においては、以下のような方針を定めました。

表 5-5 キーワード付与基準

方針	説明
写真から読み取れる事実をキーワードとする	<ul style="list-style-type: none"> 写真からわかる対象物名 写真から読み取れる文字

「曖昧」「不確実」な情報をキーワードとして用いない	<ul style="list-style-type: none"> ・位置情報が無いものに関しては、被災地と被災地外の切り分けが不可能 ・時間情報が無いものに関しては、震災前と震災後の切り分けが不可能（例えば、建物被害が見られない建物など） ・震災被害かどうか、被害原因等が判断できないものに関しては、被害を示すキーワードを用いない ・構造物の全壊や半壊、一部損壊など、被害程度を示す指標については、専門家しか判断できないため、被害程度を示すキーワードは用いない
ほぼ確実ではあるが、確定できないものは、ソーシャルタグとして付与する	<ul style="list-style-type: none"> ・撮影の意図 ・地図や検索で知り得た情報 ・その他、明らかに事実と認められる情報
キーワードを単語毎に分割せずに用いる	<ul style="list-style-type: none"> ・例えば、キーワードとして「浪分神社」とした場合、「浪分」「神社」のように分割することはしない

③ ソーシャルタグの事例

本プロジェクトにおいて特徴的なソーシャルタグについて、検証を行いました。

検証については、写真画像を用いて、撮影者本人と本プロジェクトとは関係していない第三者にお願いし、付与したキーワードに、新たにソーシャルタグの付与を依頼しました。その結果、撮影者本人については、撮影した状況や背景を理解しているため、撮影の背景がわかるキーワードが追加されました。一方、関係しない第三者がソーシャルタグを付与した場合、不確かなキーワードが追加されました。また、今の状況を表すキーワードを振るケースも確認することができました。

(例 5) プロジェクトで付与したキーワードとソーシャルタグとの違い (1)



【プロジェクトで付与したキーワード】
PARADISO、がんばろう石巻宮城に生まれ宮城とともに、復興 絆、配布物

【撮影者が追加したソーシャルタグ】
書道家

【第三者が追加したソーシャルタグ】
石巻、パチンコ屋、ポスター、

(例 6) プロジェクトで付与したキーワードとソーシャルタグとの違い (2)



【プロジェクトで付与したキーワード】
ドラム缶、建物、損傷物、被害、被災家屋、車、道路、雨

【撮影者が追加したソーシャルタグ】
復興写真展で入賞写真場所、残すかどうかで議論があった

【第三者が追加したソーシャルタグ】
撤去された、シンボル

例 5 のケースでは、撮影者本人が付与したキーワードは、有名な書道家が書かれたものであることを知っており、「書道家」という言葉が付与されました。第三者に関しては、「PARADISO」というキーワードが、パチンコ屋の会社名と知っており、「パチンコ屋」と付与されました。写真の中には、パチンコ屋という名称が記載されておらず、「PARADISO」から推測されたキーワードが付与されたこととなります。

例 6 のケースでは、撮影者本人が付与したキーワードは、復興写真展で入賞写真をみた風景ということと、建造物を残すかどうかの議論があったが撤去されたこと背景を知っていたことであったため、その 2 点の「復興写真展で入賞写真場所」、「残すかどうかで議論があった」という言葉を付与しました。第三者については、ニュース等で見た

ことがある風景であり、建造物が撤去されたことを知っているため、「撤去された」、「シンボル」というキーワードが付与されました。

ソーシャルタグのキーワード付与に関しては、キーワードとしては不確かではあるが、検索キーワードが増えることで検索の広がりがあることが分かりました。また、撮影した時点の情報のみだけでは無く、復旧・復興過程で得られた情報についても付与可能であることが検証で確認できました。一方、このソーシャルタグは不確実なことが多いため、検索方法については、一般的な検索方法と同様に行うかは今後の課題となります

尚、本プロジェクトでデジタル化したコンテンツにおいては、処理に応じてショック映像に分類されるコンテンツは無かったため、それを分類するメタデータは不要でした。

(b) メタデータ付与に利用した機器等

写真へのキーワード付与については、ツールとして **GooglePicasa** を使用しました。**GooglePicasa** は、写真にタグを振ることができるため、キーワードを振る際のツールとして活用しました。

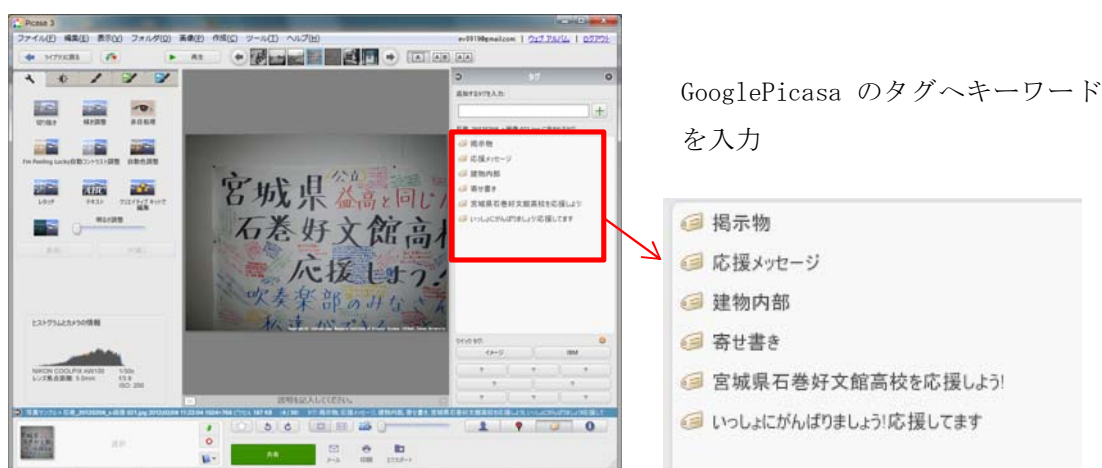


図 5- 10 GooglePicasa 画面イメージ

GooglePicasa では、写真にキーワードを埋め込め、写真とキーワードを別々に管理しなくても良い他、写真を見ながら作業ができ、写真の縦横変換も合わせて作業することが可能でした。さらに位置情報が入っている場合は、それも合わせて確認できるメリットがあります。登録したキーワードは、**Google** の **API** を使用したツールで、ファイル名とキーワードを抜き出し、これをシステム登録時のデータとして使用しました。

その他のツールとしては、Excel を用いてファイル名とキーワード管理を行いました。

(参考) ツール実行結果

塩竈_20120309_宮城県塩竈市楓町 2 丁目_s-画像 472. jpg	学校 指定避難場所 看板
塩竈_20120309_宮城県塩竈市楓町 3 丁目_s-画像 473. jpg	中学校 校舎 看板
塩竈_20120309_宮城県塩竈市楓町 3 丁目_s-画像 474. jpg	塩釜第二中学校校舎 看板

ファイル名

キーワード

図 5- 11 Excel を用いたファイル名とキーワードの管理

(c) 課題と対応策

メタデータ付与に関する課題と対応策については、以下のとおりです。

① メタデータのばらつき

【課題内容と対応策】

入力者によりメタデータの付与にばらつきがみられました。具体的には、「がれき、ガレキ、瓦礫」という表記や、「オートバイ、バイク、自動二輪、車両」という例が存在しました。

メタデータは、コンテンツの説明であるという側面と、検索するためのキーワードとしての側面を持っています。前者は、コンテンツに付随する様々な情報といった内容ですが、利活用の観点から考えると、後者が重要な要素となります。検索をしやすくするためには、キーワードを充実させる必要があります。この語彙をどのように増やしていくのかが、大きなテーマとなります。例えば、「ガレキ」「がれき」「瓦礫」という言葉は、一般的に使われている語彙ですが、キーワードを付与した担当により、違いが出てきてしまっています。これを統制することは、検索の観点から望ましくないため、同義語の定義を何等かの形で、システム的に実装するのが良いと考えます。

② 震災関連デジタルアーカイブとしてのメタデータの独自性

【課題内容】

コンテンツを検索する際、「時間」、「空間（位置情報）」、「テーマ（検索語彙）」の 3 軸による検索としました。一般的な美術館、図書館、文書館等で運営しているアーカイブと震災関連デジタルアーカイブでは、下記のように検索内容に違いが存在することがわかりました。

・ 時間

後年運用の利活用を考えた時に、年月日時分までの情報が必要なコンテンツが存在しました。

・ 空間（位置情報）

被災地域毎に、事象や事情に違いがあるため、どの地域のコンテンツかが、重要であるとともに、詳細な住所が必要なケースもありました。

・テーマ（検索語彙）

建物、重機、Twitter、Facebook、イベント、風景等、身近に見える全ての物が検索対象となり、様々な語彙が使われます。

これらは全てメタデータとして付与された内容ですが、何処まで精度を上げてくかが課題になります。今回の実証調査では、第三者がメタデータを振っているため、キーワードが不足していると考えられます。

③ 専門用語の付与

【課題内容】

研究者の視点で考えた場合、専門用語がキーワードとして付与されていないことが課題として上げられます。研究者による利活用を想定した場合、研究者による検索の精度を高めるため、専門用語が付与されていることは、重要な要件です。しかし専門用語は、研究者しか使わない用語であり、実際にキーワードを振る際も、その領域に精通している人が振る必要があります。したがって専門用語によるキーワードをどの様に充実させていくのかは、後年運用における課題となります。

(5) 宮城河北新報社における実証調査

(a) メタデータ付与の手法・プロセス

メタデータ付与作業の手法・プロセスについて、以下に記載します。

① メタデータ付与の詳細

メタデータの付与は、作業開始前に東北大学とディスカッションし、東北大学の考え方を参考にしました。

以下にメタデータ入力例を示します。



時間：20110315

住所：仙台市青葉区一番町

コメント：一番町4丁目商店街

キーワード：張り紙、安否情報、伝言板、マルシェジャポン、シャッター、とよはし、きぶや、あずま寿司

図 5-12 メタデータ入力例 1



時間：20110317,15:17
 住所：宮城県石巻市中央2丁目
 コメント：松栄石巻パーキング付近
 キーワード：雪、鉄塔、道路石巻線、
 自転車、破損者、流木、津波被害、
 松栄不動産、木片、打ち上げ線、内
 海橋

図 5-13 メタデータ入力例 2

メタデータ付与については、キーワードの内容や場所の特定に苦勞をしました。宮城河北新報社プロジェクトでは、出来るだけ正確な内容を振ることを目標としたため、時間をかけて検討を行いました。

一方、正確にわからないものも多く、抽象的な表現を使うこととしました。

また、メタデータを付与する担当者の思い込みにより、キーワードを付与するケースもありました。

(例 1) 場所の特定



例 1 の写真は、同じ施設ですが、施設名称が時期によって変化しており、検索の際にはどの名称を入力しても検索できるよう全ての語彙を付与する必要があるケースです。

これは仙台市社会福祉協議会が震災直後、宮城野区新田東の宮城野体育館内に設置した災害ボランティアセンターの写真ですが、左の写真が示すように設立直後は「宮城野区災害ボランティアセンター」でした。(2011年4月26日まで)。4月27日以降は「仙台市北部津波災害ボランティアセンター」と変更、さらに6月1日からは、若林区にあったセンターを統合して「仙台市津波災害ボランティアセンター」に変更されました。この間、若林区のセンターも、同様に「若林区」「南部」と変更になっています。

(例 2) 背景の特定



【検証前】

キーワード：被害、川、倉庫、工場、タンクローリー、

【検証後】

キーワード：被害、川、建物、タンクローリー、

(例 3) 被写体の呼び名



【検証前】

キーワード：がれき、防潮林、戸建、損壊、残存、津波被害 . . .

【検証後】

キーワード：がれき、並木、戸建、損壊、残存、津波被害 . . .

(例 4) 津波被害後の状態を表す語



【検証前】

キーワード：がれき、基礎、外壁、戸建、損壊、流木 . . .

【検証後】

キーワード：がれき、基礎、外壁、戸建、損壊、残存、流木

例 2 のケースでは、背景に写っているのが、「倉庫」や「工場」なのかが、正確に特定できないため、「建物」という言葉に置き換えました。例 3 のケースは、海岸部の林を指す名称として、「防風林」が正しいか「防潮林」が正しいか判断できなかったため、「並木」としたケースです。しかし、この場合は、「防風林」「防潮林」あるいは「防砂林」などと複数入力しておけば済んだケースであり、抽象的な単語 1 つに絞る必要はなかったといえます。このように一度決めた判断でも後からみると変えた方がよいということもありました。

例 4 は、津波被害で残った状態の表現について、議論をおこないました。この場合、「残存」という言葉を使うこととしました。

(例 5) 事実が確認できない場合

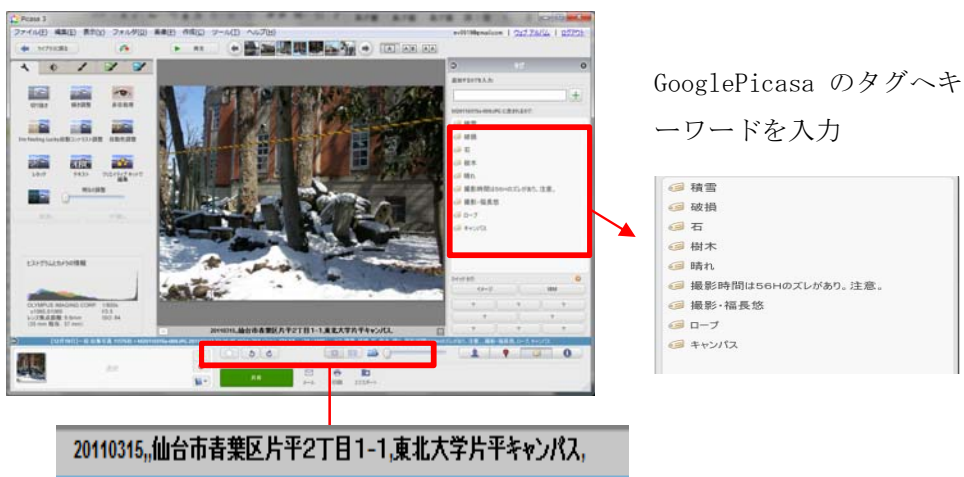


【検証前】 キーワード：おとな、毛布、男性、避難所、避難者、・・・
【検証後】 キーワード：おとな、毛布、男性、避難所、避難者、 <u>被災者</u> 、・・・

例 5 のケースは、大震災発生の数日後、仙台市青葉区川内の川内コミュニティーセンターで撮影された写真です。場所や時期から判断して、津波の直接の被害者がいない避難所です。避難所で撮影された写真に避難者が映っている場合は検索しやすくするために「避難者」、「被災者」の語句を付与するようにしました。

(b) メタデータ付与に使用した機器等

写真へキーワードを付与する際は、東北大プロジェクトと同様に GooglePicasa を使用しました。GooglePicasa は、写真にタグを振ることができるため、キーワードを付与する際のツールとして活用しました。



GooglePicasa の説明欄に年月日、住所、コメント（建物）を入力

図 5- 14 GooglePicasa 画面イメージ

(c) 課題と対応策

メタデータ付与に関する課題と対応策については、以下のとおりです。

① キーワードのバラつき

【課題内容と対応策】

キーワードについては、語彙として、「ガレキ」、「がれき」、「瓦礫」など3種類の表現がある等、作業者によりバラつきがありました。また、住所については、県名がないものや市区町村名がないもの等、様々なケースがありました。

キーワードのバラつきを統制することは、検索の観点から望ましくないため、同義語の定義を何等かの形で、システム的に実装するのが良いと考えます。

住所については、手作業での補正を実施しました。

② 施設名の違い

【課題内容と対応策】

仮設住宅や避難所は、通称があり、時期によっても名称が違っていることが確認されました。

施設名は、位置を調べ、インターネットで検索し、自治体のサイト等を確認しながら、名称を入力しました。

【参考としたサイト】

Google マップ : <https://maps.google.co.jp/>

Actiz : <http://www.actiz.jp/>

Wiki : <http://ja.wikipedia.org/wiki/>

その他各自治体のサイト

(6) 福島県における実証調査

(a) メタデータ付与の手法・プロセス

メタデータ付与作業の手法・プロセスについて、以下に記載します。

① メタデータ付与の詳細

- 福島プロジェクトでの収集情報は文書が多かったため、メタデータの付与にあたっては、内容属性について文書の内容を判断しながら付与する必要がありました。付与に際しては、主観的に判断しないように文書内で使用されるキーワードを抽出し、これを属性値として使用しました。

- メタデータ付与作業にあたっては、Excel シートにコンテンツのファイル一覧を作成し、メタデータを付与しました。付与結果を確認後、アーカイブへ一括で取り込みました。
- メタデータ付与については、数人で実験的に数千単位のコンテンツのメタデータ付与を実施した上で、工数及び必要なスキルを精査し、メタデータ付与のための体制を整備しました。
- 以下のとおり「ファイル情報」「コンテンツ情報」については、サンプルを作成し、入力のばらつきを防止しました。

付与するタグ情報		説明
ファイル情報	ファイル名	ファイルの物理名
	作成日	ファイルの作成日付(自動セット)
	サイズ	ファイルサイズ(自動セット)
コンテンツ情報	タイトル	コンテンツの概要がわかるようなタイトルを付与
	作成日	コンテンツの作成日、または公開日
	作成者	情報の作成者
	発信者	情報の発信元
	提供者	情報の提供者(1次提供者)
	時点情報	コンテンツがいつ時点のものかを表す情報(該当する日付をFrom Toで指定若しくは全文検索)
	地点情報	コンテンツがどの地点のものかを表す情報(場所を特定するキーワードを抽出)
	コンテンツ分類	コンテンツがどのような情報かを識別する分類(別紙参照)
	内容分類	コンテンツの内容を識別する分類(別紙参照)
	内容	コンテンツがどのような内容かについての詳細情報。コンテンツ内の単語を抽出してセット(若しくは全文検索)

コンテンツ分類			
写真	被害状況	被害状況が判別できる写真	
	避難所	避難所における写真	
	風景	まちの風景を撮影した写真	
	イベント	各種イベント風景を撮影した写真	
	その他	上記以外の写真	
	動画	災害記録	災害・被災状況を撮影した動画
		放送番組	放送局等で放送された動画
		広報ビデオ	各種PR用に作成された動画
		イベント風景	イベント風景を撮影した動画
	文書	被災・被害状況	被災・被害状況に関する情報
避難所・避難先情報		避難所の状況に関する情報	
放射性物質検査情報		環境放射能、飲料水、農作物等に関する放射性物質検査結果に関する情報	
支援情報		支援物資、義捐金等に関する情報	
お知らせ		上記以外のお知らせ情報	
その他		上記以外の情報	
その他			

図 5-15 「ファイル情報」「コンテンツ分類」のサンプル

この他、メタデータとして、自治体に関する属性情報を付加しておくことで、共通の

属性を含む自治体を横断的に検索することが可能となるため、自治体に関する属性情報をメタデータとして付与することとしました。自治体に関する属性情報の例は、以下のとおりとなります。

- 避難指示区域、警戒区域の指定の有無とその内容
- 避難指示解除準備区域、居住制限区域、帰還困難区域、警戒区域、計画的避難区域、緊急時避難準備区域
- 自治体避難先の避難先情報
該当自治体：双葉町、大熊町、川内村、楡葉町、広野町、浪江町、富岡町

(b) メタデータ付与に使用した機器等

- メタデータ付与に関しては、特別なハードウェアは使用しておらず、Excel において、マクロやワークシート関数を使用しました。

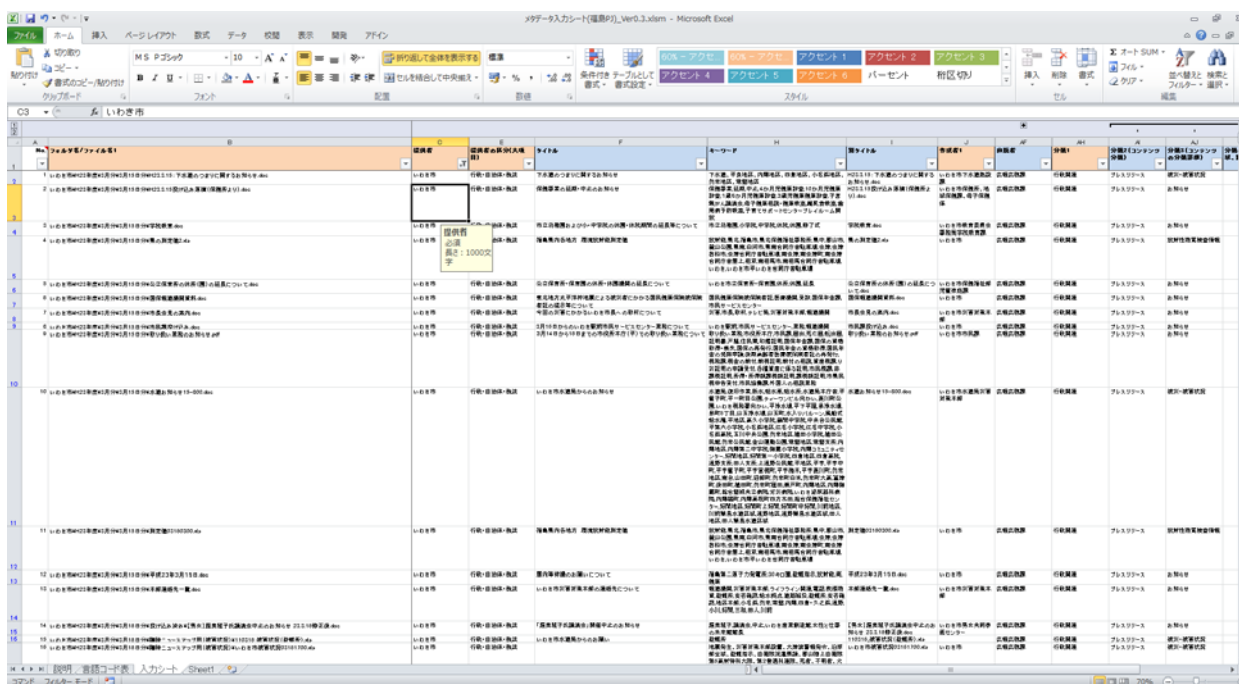


図 5-16 メタデータ登録画面イメージ

(c) 課題と対応策

メタデータ付与に関する課題と対応策については、以下のとおりです。

① 位置情報の付加

【課題内容と対応策】

コンテンツを地図データへマッピングしながら閲覧する利用イメージを想定して、全てのコンテンツに緯度・経度情報を付与する方針でしたが、自治体の広報誌のような市内の様々なスポットの情報を掲載している場合は、位置情報を一つに絞ることができませんでした。

また、特定の位置情報が含まれないケースも発生しました。

どのスポットを緯度経度情報として登録するか、広報誌に掲載されたスポット情報を収集先自治体と協議し、決定しました。

② メタデータの付与方法について

【課題内容と対応策】

メタデータは、主にコンテンツの検索時に使用されますが、コンテンツの検索方法としては、Google の検索サイトのような、キーワードによる検索と、主に Yahoo 等のポータルサイトのようにあらかじめ用意された階層型のカテゴリから、1 階層ずつカテゴリを選択しつつ、絞り込みを行う検索の大きく 2 種類あります。

前者のキーワード検索の場合、検索キーワードとして使用すべき対象は文書であれば、全文検索が望ましいと考えられます。メタデータ付与の作業効率を考えるとデジタルデータの全文をキーワードとして登録する時間は確保できなかったこと、PDF から機械的なテキスト抽出機能においては、改行等により単語が分断された場合には、うまくキーワード検索できなくなるため、全文検索は行わないこととしました。対応策として目次、見出し等から抽出したキーワードのみを検索対象として登録しました。

カテゴリ検索の場合には、どのようなカテゴリでコンテンツを分類するかが重要なポイントとなります。福島プロジェクトにおいては、図 5-15 のように実証テーマに即した分類を設定しました。