

東日本大震災の膨大なデータを収集・活用する アーカイブ・プロジェクト「みちのく震録伝」を推進



国立大学法人東北大学（以下、東北大学）では、東日本大震災やその後の復興状況などの情報を収集・伝承するアーカイブ・プロジェクト「みちのく震録伝」を産学官民と連携しながら推進しています。あらゆる可能性を考慮し、収集の対象を非常に幅広く想定しているため、集められた情報はペタバイト級の膨大な量に上ります。ストレージに蓄積されたビッグデータを有効に活用するために、みちのく震録伝では写真データを検索するためのタグ付け、テキスト情報のテキストマイニングによる分析、映像データの音声認識など、最新技術の活用を進めています。こうした技術活用はプロジェクト活動の今後の成果に大きく影響するものであり、災害アーカイブ活動のスタンダード確立のためにも重要な要素として注目されています。



みちのく震録伝

このロゴは、「東北の力強さと勢い」「山と海を持つ自然の力」「勤勉と実践」を表しています。

災害科学国際研究所を設立し 災害対策の取り組みを強化

東北大学は、1907年の創立以来の伝統である「研究第一」と「門戸開放」を理念に掲げ、世界最高水準の研究・教育を推進しています。東北地方は古くから地震が多発している地域であり、東北大学では以前から地震・震災関連の取り組みを推進してきました。東日本大震災後さらにその取り組みを強化するため、2011年4月に東北大学災害復興新生研究機構を設立しました。東北大学災害復興新生研究機構では3つの理念（表1）に従い、全学からの積極的参画を得て、8つのプロジェクトと復興アクション100+の推進・支援を行っています。

8つのプロジェクトは「災害科学国際研究推進プロジェクト」「地域医療再構築プロジェクト」「環境エネルギープロジェクト」「情報通信再構築プロジェクト」「東北マリンサイエンスプロジェクト」「放射性物質汚染対策プロジェクト

ト「地域産業復興支援プロジェクト」「復興産学連携推進プロジェクト」の8種類で、東北大学が強く推進している活動です。一方、復興アクション100+は「被災者救助」「被災状況把握・調査」「復旧・復興活動」「防災・減災対策」「インフラ等整備」「産業復興・研究開発」などの分野において、さまざまな部門の東北大学教職員が自主的に行っている取り組みです。

8つのプロジェクトの中で東北大学が最も力を入れているものが「災害科学国際研究推進プロジェクト」です。このプロジェクトは2012年4月に設立された災害科学国際研究所を中心に、文系・理系などの分野の垣根を越えた「実践的防災学」の創成を目指して多彩な研究に取り組んでいます。災害科学国際研究所設立の目的について、東北大学災害科学国際研究所 災害リスク研究部門津波工学研究分野 副所長 教授 今村 文彦氏は次のように説明します。

「三陸沖も含めて東北地方は地震・津波の常襲地帯です。特に三陸を中心として津波が頻繁に発生してきましたので、これまでさまざまな側面から対策が行われてきました。東日本大震災以前は過去400年の間に発生した規模の地震・津波を想定した対策が取られていましたが、実際はそれをはるかに上回る規模の地震と津波が発生しました。この問題を解決するアプローチには2つの方法があると思っています。1つは、想定を超えるものをサイエンスでどこまで解明できるかということです。これは古文書の解明だけではなく、地質調査や長期的な変動地形学などの分野でも千年、万年、数千万年といった期間で地震の解明を進めることが重要です。もう1つは、その研究結果を生かして、大規模災害に社会が対応するためのシステムを作り上げることです。そのためには過去の教訓をしっかりと伝承していくことも非常

東北大学災害科学国際研究所
災害リスク研究部門
津波工学研究分野
副所長
教授



今村 文彦 氏
Dr. Fumihiko Imamura

に重要になります」

また東北大学災害科学国際研究所 情報管理・社会連携部門 災害アーカイブ研究分野 准教授 博士（工学）柴山 明寛氏は、災害科学国際研究所の特徴について以下の通りに説明します。

「災害科学国際研究所は日本最大規模の災害研究所で、『災害リスク研究部門』『人間・社会対応研究部門』『地域・都市再生研究部門』『災害理学研究部門』『災害医学研究部門』『情報管理・社会連携部門』の6部門の研究を分野横断型で推進しています。これまでの災害研究では文系と理系が連携することはなかったのですが、災害科学国際研究所では、例えば工学と法学、医学と文学といったような組み合わせで研究を推進し、『想定外をなくす』ため災害の実態解明を目指しています。あらゆるケースを想定してすべての災害を防ぐことは難しいですが、さまざまな想定にしっかりと対応できることを検討していきます」

震災関連のあらゆる情報を収集する アーカイブ・プロジェクト「みちのく震録伝」

災害科学国際研究所の代表的な活動の1つとして「みちのく震録伝」を挙げることができます。みちのく震録伝は震災の記録を伝承するために、あらゆる情報をアーカイブする活動で、災害科学国際研究所設立以前に発足したプロジェクトです（図1）。

「みちのく震録伝の活動自体は2011年6月ごろから開始され、2011年9月に名称が決まって正式にプロジェクトとしてスタートしました。当初は東北大学の数人のメンバーで自主的に開始したのですが、さまざまな方面から協力をいただき、プロジェクトとしてスタートした時には約20の組

表1. 東北大学災害復興新生研究機構の基本理念

理念1	復興・地域再生への貢献 これまで経験したことのない大震災からの復興・地域再生に被災地の知の拠点として貢献
理念2	災害復興に関する総合研究開発拠点形成 東北・日本のみならず、災害復興を目的とした総合研究開発のための世界的COE（Center Of Excellence：中核的研究拠点）を形成
理念3	分野横断的な研究組織で 課題解決型プロジェクトを形成 災害復興に貢献するため、これまでの部局の枠にとられない横断的な研究組織で課題解決型のプロジェクトを形成し、戦略的・組織的に取り組む



東北大学災害科学国際研究所
情報管理・社会連携部門
災害アーカイブ研究分野
准教授
博士（工学）

柴山 明寛 氏
Dr. Akihiro Shibayama

織が参加されました。みちのく震録伝のアーカイブを開始したきっかけとして、今回の震災では情報メディアの発達によりさまざまな情報が大量に残されているという状況がありました。これらの情報を集めて、研究や伝承を目的として保存し共有することが重要だと考えたのです」（柴山氏）。

みちのく震録伝では、以下の通りに活動の理念を定めています。

- 学術的な観点からあらゆる可能性を否定せずに幅広く情報を収集し、アーカイブすること。
- 東日本大震災をはじめとする、東北地方の過去・未来の災害についてもアーカイブすること。
- 地震・津波災害の防災・減災の研究の進展に貢献し、防災・減災対策に活用する情報を発信すること。
- アーカイブと利用者間の対話の中から、継続的に成長

するシステムであること。

- 被災地の復旧・復興過程の現況を東北地方の沿岸部から内陸部まで空間的に記録し、発信すること。
- 他機関のさまざまな震災アーカイブや社会サービスと横断的に連携すること。
- 災害アーカイブのグローバル・スタンダードを目指すこと。
- 大学、研究所、行政、企業などの防災・減災システムと連携し、対策・対応を支援すること。
- さまざまな環境に応じた防災教育に関する情報を提供できること。
- アーカイブ・システムの構築に伴い、東北地方を中心に雇用を創出すること。

「この理念の中で特に重要なポイントは、1番目に掲げている『あらゆる可能性を否定せずに幅広く情報を収集』することです。研究者は自分の研究目的や専門分野のみに注目がちですが、将来研究対象となる可能性を考慮して、現時点では必要とは思えない情報であっても漏らさずに収集しています。例えばフリーペーパーなども、掲載されているスナップショットが今後の研究に役立つかもしれません。学術的な観点であらゆる可能性を考慮して収集しアーカイブしているのです」

産官学民が連携して推進されているみちのく震録伝には、日本アイ・ビー・エム株式会社（以下、日本IBM）も早い段階から参加しています。

「みちのく震録伝は大量の情報をアーカイブする活動で

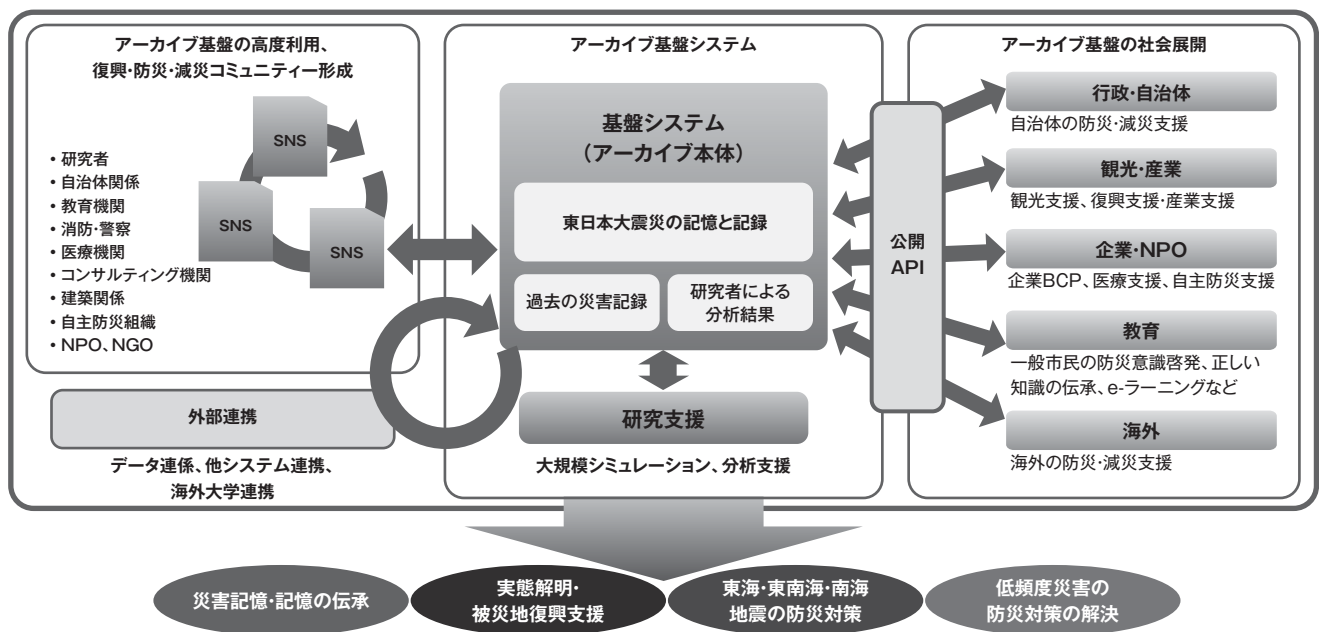


図1. みちのく震録伝の概要

すので、IT 企業の協力も必要になります。日本 IBM では Smarter Cities やビッグデータの取り組みを推進しており、みちのく震録伝のアーカイブ活動との親和性が高いということから、共に活動を展開することになりました」（柴山氏）。

画像データを分類・検索するための タグ付け作業を自動化

みちのく震録伝では、さまざまな方法で情報を収集しており、その中の1つとして「みちのく・いまをつたえ隊」の活動があります。

『「みちのく・いまをつたえ隊」は2012年1月から始めた被災地密着型の情報収集活動で、復旧・復興の現状を中心に、GPS内蔵カメラによる写真撮影、住民へのインタビュー、チラシ類を含めた現地の方々が保有している情報を収集するといった活動を沿岸部の15市町村で展開しています」（柴山氏）。

「みちのく・いまをつたえ隊」以外では、災害科学国際研究所のスタッフによる各自治体からの情報収集、賛同協力機関からの情報収集、新聞社などのメディア関連企業からの情報収集など、多岐にわたる収集活動を展開しています。

こうして収集した膨大な量の情報を有効に活用するためには、必要な情報を素早く検索できることが求められます。しかし、情報の中にはテキスト・データが付されていない写真や動画のデータも大量に含まれているため、これらのデータを分類し、検索のためのタグ付けを行わなければならない。

「現時点で非公開のものも含め、20～30万点ほどの写真データが集まっていますが、これらの写真を検索す

るためにはそれぞれの写真データにキーワードをタグとして付けておく必要があります。しかし、これだけの量の写真を手作業で処理することは膨大な労力とコストを要するので、タグ付けをある程度自動化する手段はないかと日本 IBM に相談したところ、IBM Multimedia Analysis and Retrieval System（以下、IMARS）の活用を提案していただきました」（柴山氏）

IMARSはIBMが開発を進めている画像の自動分類・検索システムで、高度な機械学習に基づいて画像の視覚的特徴を理解し、自動的にラベル付けを行います（本誌69ページ以下：IBM技術解説参照）。みちのく震録伝では、膨大な画像のタグ付けにIMARSを活用できるかどうか、その有効性を検証するための技術検証を行いました。

「検証の結果『Beach』『Building』というような大まかな分類については一定の成果が得られました（図2）。しかし、崩壊した建物やがれきなどに関しては機械学習による自動分類が非常に難しいので、IMARSで大まかに分類した上で、さらに詳しいタグを付けを人手で行っています。結果として、それらの負荷を減らすことに役立っています。今後は、さらに高度な自動分類を目指し、建物の抽出および被害レベルをどこまで判別できるのか検証を行うことを予定しています」（柴山氏）。

テキストマイニング技術により 膨大なテキスト情報を自動分析

新聞記事やインタビュー記事など、テキスト・ベースの記事はキーワード検索が可能ですが、記事内容が表す現象を把握するためにはキーワード検索だけでは分かりに

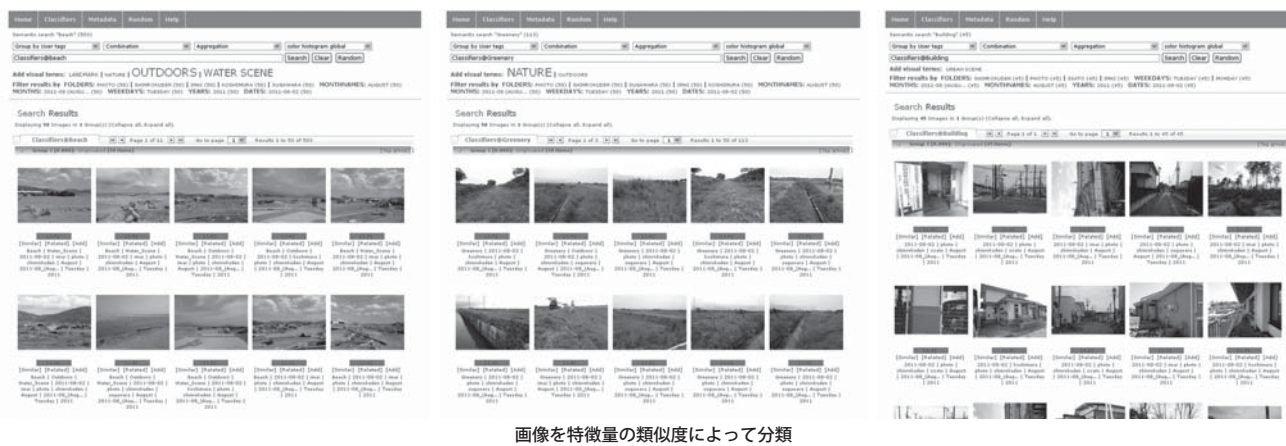


図2. IMARSを用いた画像自動分類例

くいため、テキストマイニング技術を応用した仕組み作りが進められています。

「例えば、『復旧途上』『完全復旧』『復旧を断念』など、同じ『復旧』というキーワードに対しても意味の異なる内容が検索されます。このような文章の違いを認識できなければ記事内で表記されている現象を把握することはできません。そこでテキストマイニング技術を活用して文章内容を自動分析し分類しておくことにより、例えば時系列で現象がどのように変化しているのかを把握できるようになり、より有効なデータの活用が実現します(図3)」(柴山氏)。

テキストマイニングにはIBM Content Analytics(以下、ICA)を活用しています。ICAは大量データであっても一括分析を実現し、高度な構文解析にも対応しています。また「ファセット分析」「時系列分析」「偏差分析」「トレンド分析」「ファセット・ペア分析」などさまざまな分析が可能なので、多角的に現象を把握することが可能です。

「今後は分析結果を使って次の支援活動を決定するなど、意思決定に活用していくことを計画しています。大量のデータを活用しやすい形で提示して、その情報を必要とする人の役に立てていくためにICAを有効活用していきたいと思っています」(柴山氏)。

また映像データの音声について分析するためには、音声情報をテキスト化して整備することが求められます。そこでみちのく震録伝では、IBMの技術を活用した音声の自動認識にも取り組んでいます。

「インタビューなどの映像情報を分析するためには、音声認識が必要になるので、その検証も行いました。結果としてアナウンサーなどの専門家が話しているものについては高い精度で認識されました。しかし、一般の方へのインタビューなどにおいては、方言やマイクの性能、周りの雑音などに影響されて認識率が周りの雑音などに影

響かれて認識率が下がってしまいました。現在音声認識の取り組みは中断しているのですが、今後はどこまで方言に対応できるか研究を進めたいと思います。さらに、細かい地名の辞書を組み込むことも計画していますので、これらの取り組みにより音声データの認識精度がさらに上がってくれば、テキストマイニングへの応用も可能になり、さらなる有効活用が期待されます」(柴山氏)。

360度カメラでの撮影を行い 復興状況の進展を把握

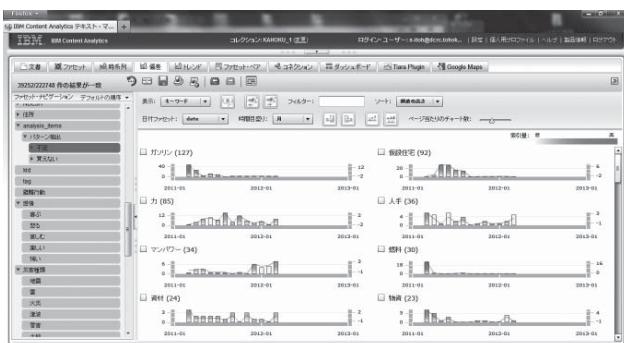
みちのく震録伝では、被災地の復旧・復興状況の進展を把握するために、車載カメラによる道路からの360度の撮影を行っています。

「東日本大震災は被災地の範囲が非常に広く、人手ですべての地域の現地調査を行うことが難しいこと、衛星写真では詳細な被害状況が判断できないことから、車載した360度カメラによる撮影が有効だと判断しました。南相馬から宮古までの約400kmの範囲をこれまでに4回撮影しましたが、協力各社が撮影したのものも含めると現在50テラバイト程度のデータが蓄積されています。撮影データはパノラマ写真や連続写真として閲覧することが可能で、詳しい活用方法は今後検討していく予定になっています」(柴山氏)。

収集された写真や映像のデータは撮影場所が重要となるため、みちのく震録伝では地図データと連携させる仕組みを開発しています。

「みちのく震録伝以外の災害アーカイブの取り組みにおいても、さまざまなデータを地図上で表現するという手法が主流になっています。どこでどのようなことが起こったのかを記録することで地理的特徴による分析も可能となります」(柴山氏)。

みちのく震録伝では、IBMのPhoto-based Virtual World(以下、PVW)を活用して、写真を位置情報と連携させるだけでなく、同じ場所で異なる時期に撮影された写真を比較しながら閲覧できる仕組みを実現しています。さらに、広範囲を撮影した写真と詳細部分を撮影した写真をシームレスに連携した機能も備えています。広範囲の写真を閲覧している際に特定の建物を詳しく調べたくなった場合、その個所をマウスで指定するだけで、あたかも建物内に入っていかのように詳細な写真が表示されます(図4)。



震災時に不足していたものを抽出して分析

図3. ICAのパターン抽出機能を用いた新聞記事の分析例

国民一人一人の命を救うことを目的として より活発な取り組みを推進

これまでさまざまな形で情報収集や情報分析を進めてきたみちのく震録伝では、Web サイトでの情報公開など、情報共有の取り組みも進められています。

「アーカイブした情報の公開・共有としては、国内外の展示会などへの貸し出しや Web サイトや Facebook などの SNS での公開という形で推進しています。今後は、ハーバード大学などとの連携をさらに強化して、海外に対する情報発信も強化していきたいと考えています。みちのく震録伝の活動は各方面から評価をいただき、各種メディアでも紹介されていますので、こうしたアーカイブ活動の重要性は広く認知されてきていると思います。2013 年 3 月までには約 10 万点のデータを公開して、誰でも自由に検索できる新しい検索サイトを公開する予定です」(柴山氏)。

さらに柴山氏は今後のみちのく震録伝の展望について説明します。

「情報の収集もさらに力を入れていくことを考えています。例えば震災の記録を数多く所有している建設業や物流業の企業にも協力いただきたいと思います。さらに国内の取り組みとしては、地元の復旧・復興に添いながら、社会生活の中にアーカイブを内包させていくことを考えています。震災の被害や復興の道のりをすべてためてから一度に公開するのではなく、社会生活の中で常にアーカイブしながらその内容を通常時にも使えるようにすることで

循環が生まれ、アーカイブが成長するとともにコミュニティの再形成支援や日ごろの防災・減災対応の支援、新しい街づくりの PR につなげることができるでしょう。アーカイブにはコストも発生しますので、ビジネスとして展開することも大事だと考えています。今後、震災記録のデータ容量はペタバイト級になりますので、ストレージの大容量化も検討しなければなりません。クラウドの利用や遠隔地でのアーカイブも検討しながら、大容量化と処理スピードの向上を実現していきたいと思います。

IMARS については 2013 年 3 月までに、機械学習によってどこまで建物被害を判読できるのかを検証する予定です。検証結果を踏まえた上でさらなる判読の精度向上に取り組んでいきたいと思っています。精度の高い分類が自動的にできるようになれば、例えば車載した 360 度カメラで撮影を行うと同時に建物の被害マップを作成することができ、災害対策のためのリアルタイムの意思決定支援が可能になるでしょう。これは、台風や洪水なども含めた災害発生直後の復旧対策や減災対策にも活用できると考えています。ほかのセンシング情報と組み合わせられるようになればその可能性はさらに広がります。このように、蓄積したビッグデータを知識インフラとして活用するために、現象同士をいかに有機的に結び付けるかが今後は非常に重要になってくるでしょう」

今村氏は現時点でのみちのく震録伝の活動の到達点を踏まえ今後の活動に対する期待について以下のように語ります。

「以前の研究では専門家の視点で限られたデータのみを集めていましたが、みちのく震録伝の活動はアーカイブ活動として新しい可能性を見いだしたといえるでしょう。現時点では断片的な情報が中心になっていますので、今後は人々の心理状況なども含めた多彩な情報収集を実現していきたいと思っています。みちのく震録伝の活動の出発点は『国民一人一人の命を救う防災学または実践学』にあります。以前の研究では、分かっていることを解析・評価して対策につなげるというものでしたが、それが命を守るために十分かどうかという検証はできていませんでした。東日本大震災の経験と教訓を受けて、一人一人の命を救うために何が必要かという従来と違う出発点から研究を推進することにより、従来とは異なる新しい成果が出てくるのではないかと期待しています」

東北大学およびみちのく震録伝の取り組みの発展は、復興計画の推進、新しい防災・減災対策の策定などに大きく貢献していくでしょう。

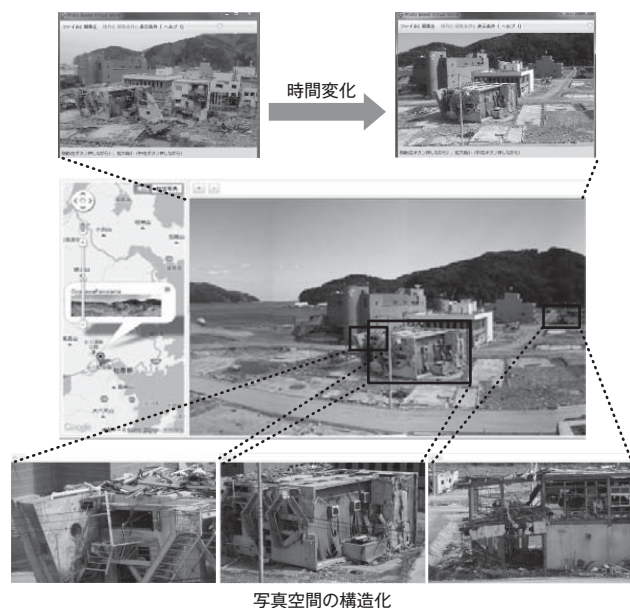


図4. PVWのみちのく震録伝への適用例