

資料 4-1

青森プロジェクトの中間とりまとめ

2013.1.23

1.0.青森プロジェクト全体概要

青森県は被災県であると同時に支援県でもあるため、資料提供元の所在地により、被災関連資料と支援関連資料の両方が収集されることが期待できる。

この両方を収集することにより、東日本大震災の俯瞰的な検証を可能とするアーカイブ構築を目指す。

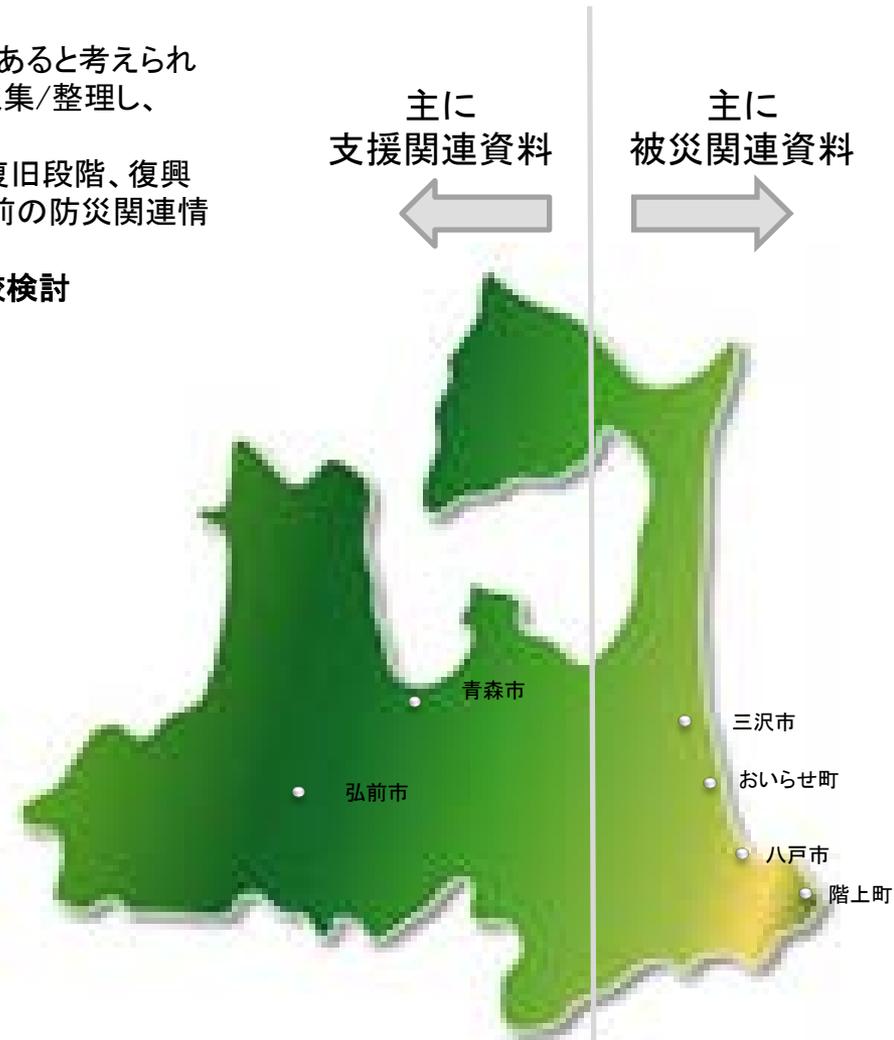
(主な特徴)

青森プロジェクトの意義は、他プロジェクトでは収集することが困難であると考えられる資料、もしくは他プロジェクトにおいて、未整理である資料を中心に収集/整理し、アーカイブ化することである。

他の被災県より一足早く復興が進んでいることから、震災直後から復旧段階、復興段階の資料を網羅的に収集できることが期待される。合わせて、震災前の防災関連情報も収集することにより、

- ・同一箇所、同一テーマに関して、震災前～復興までを時系列で比較検討
- ・防災対策の震災前の想定と、実際の被害を比較検討
- ・復興計画のモデルとしての情報を収集

を実現できることを想定しつつ、整理を進めていく。



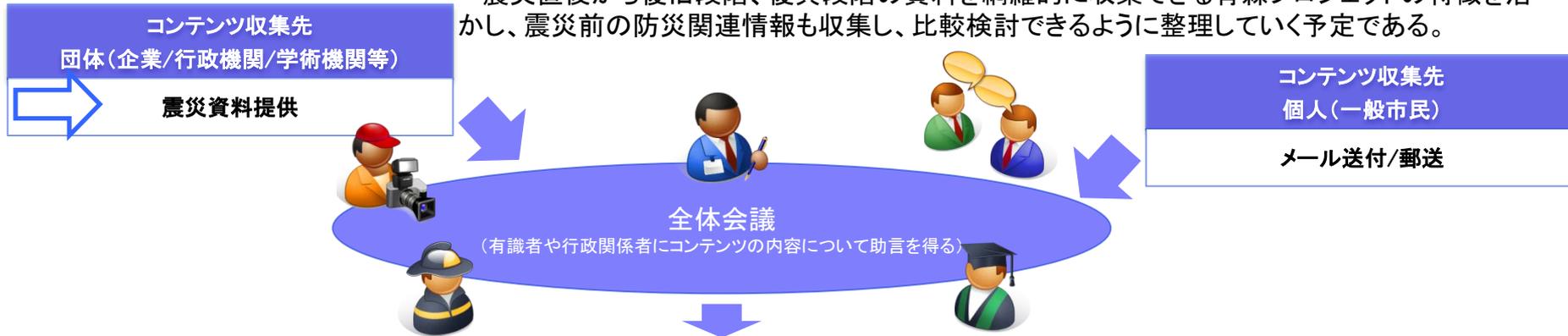
1.0.青森プロジェクト全体概要

(震災関連コンテンツの収集について)

コンテンツ収集においては、残存する震災資料を出来るだけ漏れなく収集するものとする。

青森プロジェクトでは、行政機関をはじめ、企業や学術機関等の団体及び一般住民から震災資料を収集し、有識者や行政関係者のアドバイスを受けながら、コンテンツの整理を進める。

震災直後から復旧段階、復興段階の資料を網羅的に収集できる青森プロジェクトの特徴を活かし、震災前の防災関連情報も収集し、比較検討できるように整理していく予定である。

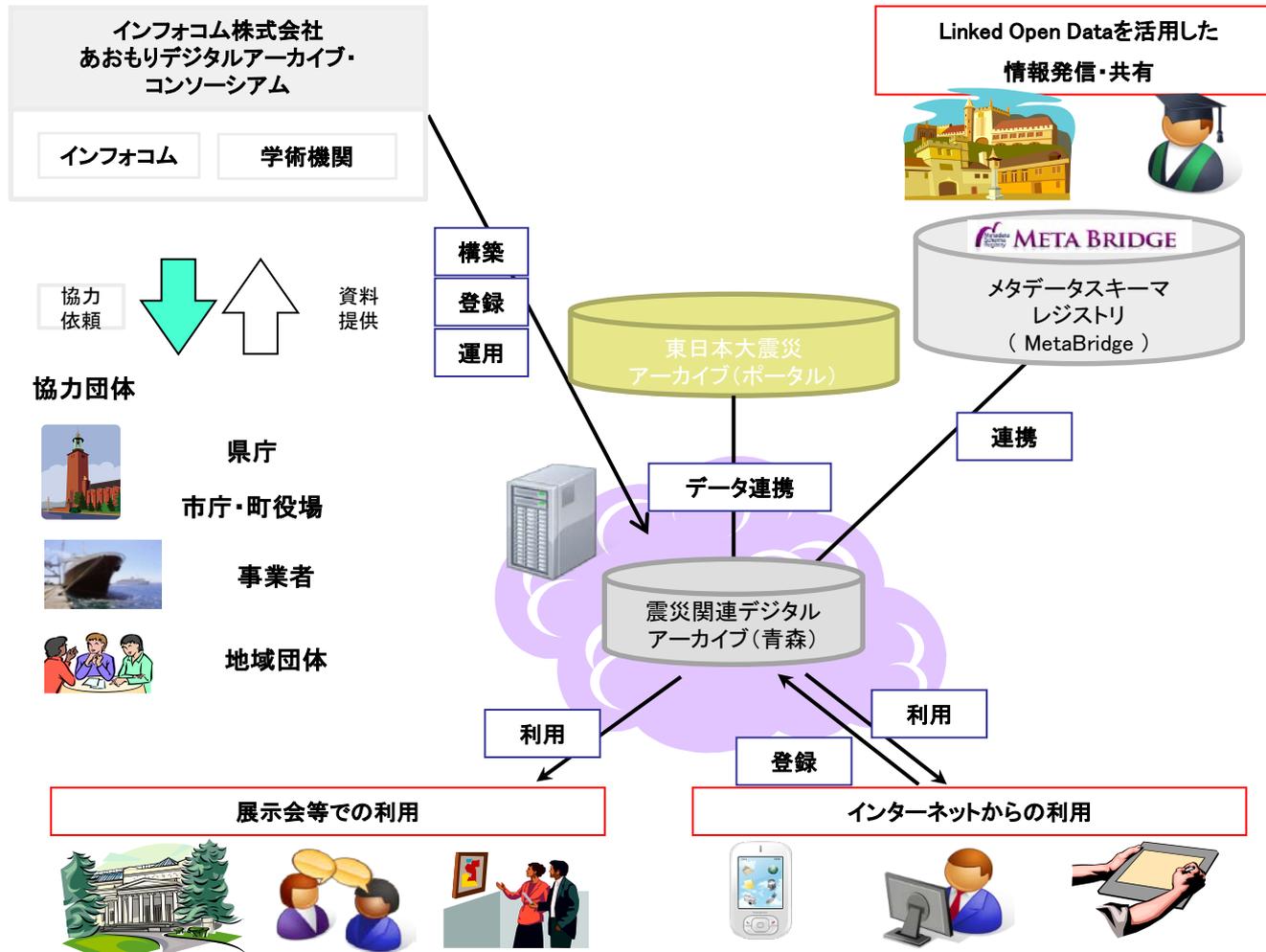


震災関連デジタルアーカイブ				
	被災前	震災直後	復旧段階	復興段階
行政・自治体・独立行政法人	<ul style="list-style-type: none"> 防災マニュアル 	<ul style="list-style-type: none"> 災害対策会議議事録及び配布資料等 	<ul style="list-style-type: none"> 復興計画 	<ul style="list-style-type: none"> 復興対策会議議事録及び配布資料等
研究機関	<ul style="list-style-type: none"> 過去の防災関連研究成果 	<ul style="list-style-type: none"> 調査資料 	<ul style="list-style-type: none"> 調査報告書 研究論文等 	<ul style="list-style-type: none"> 研究論文等
メディア	<ul style="list-style-type: none"> 被災前の写真等 	<ul style="list-style-type: none"> 新聞、雑誌記事 報道番組 	<ul style="list-style-type: none"> 新聞、雑誌記事 報道番組 	<ul style="list-style-type: none"> 新聞、雑誌記事 報道番組
団体	<ul style="list-style-type: none"> 被災前の写真等 	<ul style="list-style-type: none"> 医療機関の負傷者受入状況 	<ul style="list-style-type: none"> 復旧段階の写真等 広報資料 	<ul style="list-style-type: none"> 復旧後の写真等 広報資料
事業者	<ul style="list-style-type: none"> 防災マニュアル 被災前の写真等 	<ul style="list-style-type: none"> 震災直後の写真等 	<ul style="list-style-type: none"> 復旧段階の写真等 広報資料 	<ul style="list-style-type: none"> 復旧後の写真等 広報資料
市民	<ul style="list-style-type: none"> 被災前の写真等 	<ul style="list-style-type: none"> 生活支援情報等 	<ul style="list-style-type: none"> 生活支援情報等 	<ul style="list-style-type: none"> 生活支援情報等 防災関連情報

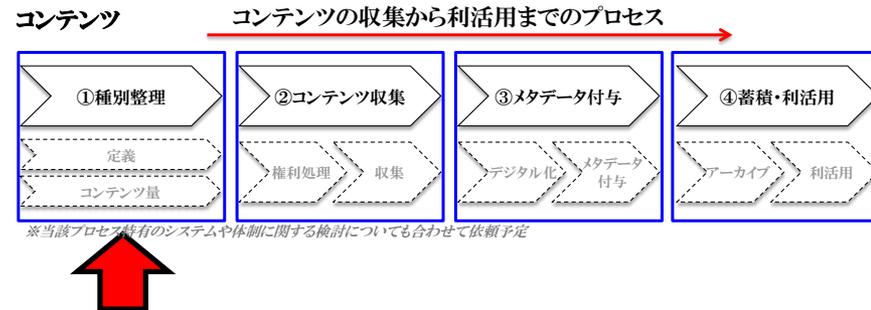
1.0.青森プロジェクト全体概要

(アーカイブシステムの構築について)

蓄積した震災関連コンテンツが長期的に保存され、幅広い利用者による震災関連コンテンツの利活用を推進するための機能を有する震災関連デジタルアーカイブの構築を目標とする。また、地域に根差した長期的な運用とコンテンツ収集が可能なアーカイブとするために、青森県において包括的なデジタルアーカイブの構築を推進している「あおもりデジタルアーカイブ・コンソーシアム」の知見を反映しシステム構築を行うものとする。

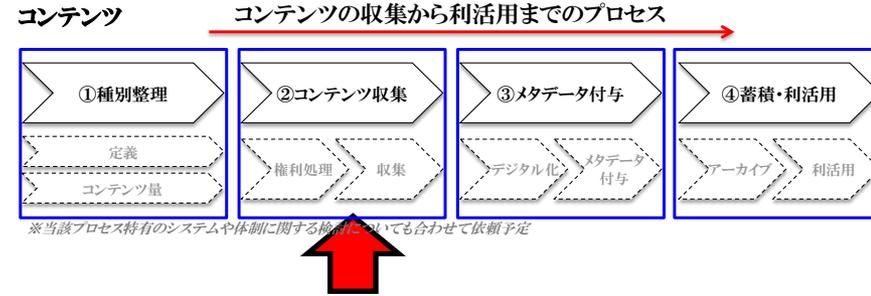


1.1.コンテンツ収集動向



No.	内容	想定種別	想定点数	実績種別	実績点数
1	行政決裁文書類	紙文書	約600点	紙文書	スキャニング作業後にアーカイブへの提供資料を仕分け予定
2	行政機関が収集した写真	JPEG	約24,000点	不明	アーカイブへの提供資料を仕分け中
3	青森県庁が作成した震災関連刊行物	冊子	不明	冊子	現在調整中
4	民間企業が収集した写真	JPEG	約4,000点	JPEG	約2,000点
5	民間企業による震災関連刊行物	冊子	不明	冊子	現在調整中
6	学術機関による被災時のレポートデータ	JPEG他	約1,000点	不明	現在調整中

1.2.コンテンツ収集における方法・課題



権利処理

- コンテンツの収集時に「使用許諾同意依頼書 兼 借用依頼書」を手交し、「使用許諾同意書 兼 借用許可書」を受領している。
- 民間企業からの画像データについては、原則として「公開」することとして収集。提供者から個別に「非公開」とされたものについては「非公開」として取り扱うこととしている。但し、「公開」を前提に収集したコンテンツについても内容を確認し、個人を特定するような情報が掲載されている場合には「非公開」とするか「マスキング」して「公開」する想定。
- 行政文書については、「公開」を前提とした場合、行政サイドとして情報提供に慎重にならざるを得ないため、コンテンツの収集とアーカイブを優先させるため、原則として「非公開」とすることとして収集。公開するコンテンツについては個別に別途、許諾を取得する想定。
- 八戸市庁所蔵の画像データについては、写真に写り込んでいる人すべての許諾を取るまで公開しないものとしている。

- 市民及び民間企業からのコンテンツ収集に関しては、地元企業のチームリーダークラスの従業員5名でグループ企業の取引先等を対象に収集作業を実施。
- 行政文書及び一部の広域企業については、インフォコム社チームリーダークラス以上の職員3名で収集作業を実施。
- コンテンツ収集にあたって教育等は実施していないが、電話及び面談の記録は議事録として共有し、課題についてプロジェクト全体で課題管理表にて共有。
- 八戸市の行政文書(決裁文書)については、その提供範囲について市職員による確認が必要とのことであったため、作業の効率化のため、市庁舎内において追加で八戸市庁OBを契約社員として雇用し1名を派遣。スキャンを行い、デジタル化し、一覧化した上で市職員が最終的にアーカイブへの提供範囲を決定することとした。

権利処理関連において以下のような課題が発生した場合、発生都度問い合わせできるような窓口があるとスムーズな収集作業が行えると考えます。

- 現在コンテンツ収集時に取り交わしている使用許諾及び借用許可に関する文書の記載内容が必要十分なものであるかの判断がつかない。
- コンテンツの提供者から「公開」を前提としてコンテンツの提供を受けた場合、内容によっては「公開」に適さないと思われるものがあるが、コンテンツの受領サイドにおける「公開」、「非公開」の判定基準及び判定プロセスを決めかねている。
- 提供された画像に「表札」、「看板」、「顔」等が映っており、個人が特定される可能性がある場合にマスキングすれば「公開」してもよいのか、その場合のあるべきマスキングの方法について基準を決めかねている。また、マスキングが著作物の改変に当たらないか？等

- 行政文書等は、一般公開にそぐわないものが多いため、アーカイブにおいても非公開とする想定であるが、このようなコンテンツについては、その提供にあたって利用者、保管場所、利用方法等が明示されないと提供が難しいと言われることが多い。具体的な利用イメージ等については収集したコンテンツを踏まえての今後の検討事項であるため、コンテンツの収集がなかなか進まない。結果として収集したいコンテンツと実際収集できるコンテンツに差異が発生している。
- 八戸市の情報については、震災以降行政文書や写真類の文書の整理が間に合っておらず、コンテンツ提供にあたってはこれらの書類の整理からサポートする必要があり、想定外の工数が発生している。
- 民間企業がコンテンツ収集に当たる場合、訪問先からの信頼の獲得に時間がかかる。学術・公的機関等の協力依頼文等の持参が必要と思われる。

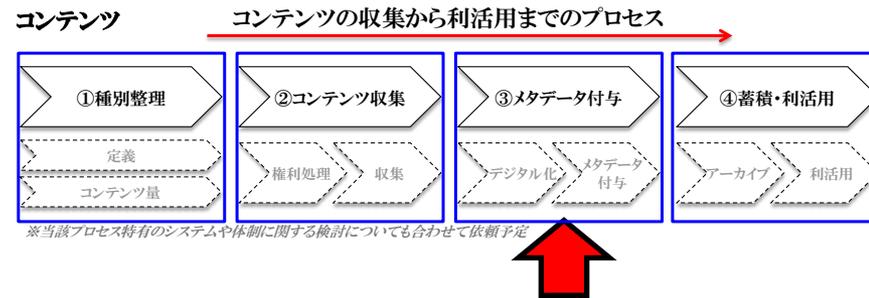
1.2.コンテンツ収集における方法・課題

青森プロジェクトでのコンテンツの収集実績(カッコ内は想定収集点数)

現段階において、当初収集対象として想定していたコンテンツと実際に収集したコンテンツについて収集量の差異(赤色部分)が発生している。行政・自治体関連のコンテンツに関しては現在デジタル化及び仕分け作業中のものが収集できる見込みであるが、これらの収集後にもなお、差異が大きい場合には、コンテンツの収集方法や分類方法の見直し、若しくは、想定する収集コンテンツ数の見直しが必要となる可能性がある。後者の場合においては、コンテンツの収集を来年度以降も継続する必要が発生する可能性が高い。

	被災前	震災直後	復旧段階	復興段階
行政・自治体・ 独立行政法人	0点(10点) •防災マニュアル	0点(500点) •災害対策会議議事録及び配布資料等	1点(50点) •復興計画	3点(50点) •復興対策会議議事録及び配布資料等
研究機関	0点(50点) •過去の防災関連研究成果	0点(750点) •調査資料	0点(50点) •調査報告書 •研究論文等	0点(50点) •研究論文等
メディア	0点(50点) •被災前の写真等	5点(100点) •新聞、雑誌記事 •報道番組	0点(30点) •新聞、雑誌記事 •報道番組	0点(20点) •新聞、雑誌記事 •報道番組
団体	0点 •被災前の写真等	0点 •医療機関の負傷者受入状況	0点 •復旧段階の写真等 •広報資料	0点 •復旧後の写真等 •広報資料
事業者	0点(50点) •防災マニュアル •被災前の写真等	5,160点(3500点) •震災直後の写真等	0点(300点) •復旧段階の写真等 •広報資料	1点(200点) •復旧後の写真等 •広報資料
市民	0点(100点) •被災前の写真等	669点(500点) •生活支援情報等	256点(300点) •生活支援情報等	255点(200点) •生活支援情報等 •防災関連情報

1.3.メタデータ付与における方法・課題



デジタル化

- 総務省監修の「デジタルアーカイブの構築・連携のためのガイドライン」にて参照している「国立国会図書館 資料デジタル化の手引き2011年版」の①保存用画像の項において「原資料に対して300～400dpi」との記載がある。これに従い、画像については300～400dpi、文書については実際にスキャン作業を行い、文字の認識が十分可能との検証結果が得られたため、作業効率を優先し、200dpiでスキャンを行うという方針とした。
- 民間企業から収集している画像についてはデジタルデータを受領しているため、デジタル化の作業は一部を除き発生していない。
- 行政文書については原則として紙文書となるため、デジタル化の作業が必要となる。デジタル化については、インフォコム社のセキュリティ規約に従い、鍵のかかる部屋に鍵で固定下パソコンと高速スキャナを持ちこみ、スキャンしたデータを暗号化機能付のディスクに保管している。

- 八戸市庁における行政文書のスキャンに関しては、ADF付のスキャナを持ち込んでいるが、付箋や写真の貼付のある資料や、折り込んである資料は結果的に1枚1枚をスキャンする必要があり、非常に時間がかかる。また、ホチキス留めの文書が多く、ホチキスを外してスキャンした後の原状復帰作業にも時間がかかっている。スキャン作業は1枚当たり1分から1分半程度の時間を要するため、1時間で40枚程度が限界である。

メタデータ付与

- メタデータ付与作業は当事業用において新たに現地で雇用したパート職員1名が入力作業を行っている。
- メタデータについては、主観的な判断を行わずに付与できる情報のみを付与するものとしている。設定するメタデータ項目の属性値については、はじめは大まかな分類で設定し、データのばらつき具合(カーディナリティ)を判断しつつ、属性値の分類の粒度を見直していくというプロセスで付与を行っている。

- 画像データについては、ほとんどが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いため、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属性値毎のカーディナリティが極めて低い結果となっている。したがって、メタデータでの検索が難しいため、検索方法としてサムネイルによる一覧表示による検索等の対応が必要と思われる。

1.3.デジタル化の例

現在、デジタル化作業については八戸市庁内で実施しており、まだ、外部持ち出しが許可されていないため、市庁舎外部におけるデジタル化の例はありません。

デジタル化の体制

- 八戸市庁内において市庁OBを1名契約社員として雇用し、派遣し、スキャナーを持ち込んで書類のデジタル化を実施中。
- 鍵付の部屋にて、ワイヤーロックで固定したパソコンにスキャンデータを保存。

デジタル化の手法

- 文書については200dpiでスキャン。画像については300～400dpiにてスキャン。
- スキャン作業に当たっては、スキャン対象がA4だけでなく、A3資料も存在すること、及びスキャンスピード(1分当り70枚)を考慮し、エプソン社のA3フラットベッドの最上位機種DS-7000を使用。しかしながら、資料はホチキス留めや付箋の貼られているものも多く、ADF(自動紙送り装置)を使用できず、1枚1枚スキャンするため、1分当り07～1枚程度の読み取り速度となっている。
- 総務省監修の「デジタルアーカイブの構築・連携のためのガイドライン」にて参照している「国立国会図書館 資料デジタル化の手引き 2011年版」の①保存用画像の項において「原資料に対して300～400dpi」との記載がある。これに従い、画像については300～400dpi、文書については実際にスキャン作業を行い、文字の認識が十分可能との検証結果が得られたため、作業効率を優先し、200dpiでスキャンを行うという方針とした。

1.3.メタデータ付与の例

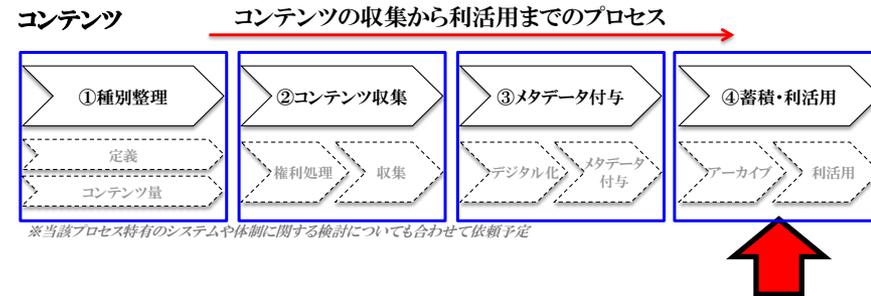
メタデータ付与の体制

- メタデータについては、アーカイブの専門知識を有るインフォコム社社員2名による監修のもと、パート職員1名がメタデータの付与作業を実施。
- メタデータの付与方針については、インフォコム社社員が決定し、その指示に従って機械的な作業によってパート職員がメタデータの入力を実施。

メタデータ付与の手法

- 標準語彙として、Dublin Core, FOAF, PREMISを採用。また国内のメタデータ語彙としてDCNDLを採用。
- NDLとのメタデータ連携のために連携メタデータへの変換機能を実装。
- メタデータ付与に関してはエクセルを使用。エクセル上でメタデータの入力を行い、最終的にアーカイブに一括登録する想定。
- メタデータについては、主観的な判断を行わずに付与できる情報のみを付与するものとしている。設定するメタデータ項目の属性値については、はじめは大まかな分類で設定し、データのばらつき具合(カーディナリティ)を判断しつつ、属性値の分類の粒度を見直していくというプロセスで付与を行っている。

1.4.蓄積・利活用



現状地域で想定される利活用

- 利活用の要素としては、利用者及び利用目的・利用方法となるが、地域で想定される利用者としては、行政、住民、学術機関、教育機関等が考えられる。それぞれの利用者における利用目的・利用方法は以下のものが想定される。

行政：災害対策立案時の参照資料等

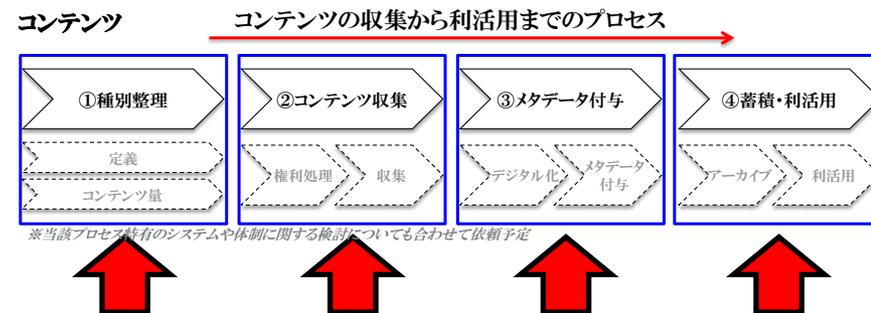
住民：郷土学習

学術機関：災害対策、建築等に関する学術研究資料

教育機関：災害啓発、災害の伝承、郷土学習

- これはあくまでも現段階の想定の利用であり、今後のコンテンツの収集状況及びアーカイブ閲覧システムの検証を通じて新たな利活用方法を検討する。

1.5.その他



その他特記事項

- 青森プロジェクトでは、同一箇所、同一テーマに関して、震災前～復興までを時系列で比較検討できるようなアーカイブとする想定であるが、前述の通り、現段階においては、想定収集コンテンツと実際に収集したコンテンツの差異が発生している。実際に自治体や民間企業が所蔵しているコンテンツの総体をどのような粒度で時間軸、空間軸で分類していくか、収集方法について、現在の収集状況をにらみつつ再度検討しなおす必要が生じる可能性がある。若しくは、想定収集規模を拡大し、来年度以降も継続してコンテンツ収集を継続していくことも必要と考える。

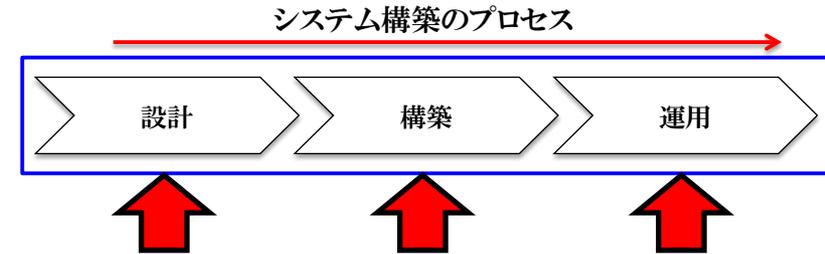
2.システム

運用モデルの特性

方針

- 青森プロジェクトでは、同一箇所、同一テーマに関して、震災前～復興までを時系列で比較検討できるようなアーカイブとする想定であり、現在震災資料の収集を行っている。これら震災資料に対して利用者が、場所や時間軸に沿って震災関連コンテンツを閲覧できるよう、タイムライン上に震災関連コンテンツを表示できるインターフェイスを実現する。
- 地域に根差したアーカイブを実現するにあたり、メタデータの相互運用性が重要である。青森プロジェクトでは設計するメタデータスキーマをメタデータスキーマレジストリMetaBridgeに登録することで、今後、他の機関が震災関連アーカイブを構築する際のメタデータスキーマ設計の一助になることを目指す。
- 幅広い利用者に活用してもらうため、一般的なWebブラウザのみではなく、スマートフォン・タブレットに最適化したユーザインターフェイスを提供する。

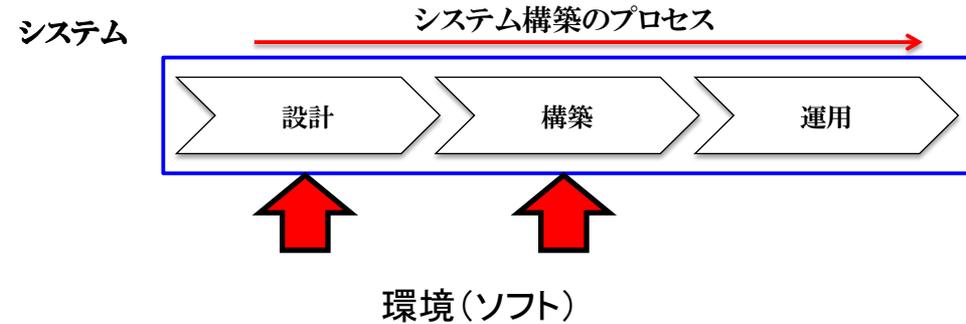
システム



考慮点

- 年表 (TimeLine) や地図 (Google Map) 上から震災関連コンテンツを探し出せる利用者インターフェイスを提供する。
- NDLとのメタデータ連携のために青森プロジェクトのメタデータ形式から連携メタデータへの変換機能を実装する。
- 画像、動画、音声、文書等はそれぞれ配信形式を決定し、国立公文書館「電子公文書等の管理・移管・保存・利用システムに関する調査報告書」の長期保存フォーマットを参考として一般的なアーカイブに適用可能なフォーマットを採用する。また、長期保存のためのメタデータをシステムで極力自動付与できることとした。
- 利用者の検索性向上のために、検索キーワードのサジェスト機能を提供する。
- スマートフォン・タブレット端末向けのインターフェイスを開発し、検証する。

2.システム



- サーバは青森プロジェクトの共同実施主体となる学術機関に設置。
- システムの想定利用頻度(最大1日100人程度)、システムダウン時に要求される復旧時間等から大容量で安価なSATAのディスクをRAID 5構成で採用しサーバのコストを削減。
- 青森プロジェクトでは実証期間内に収集する震災関連コンテンツが1万点の行政文書等のドキュメントデータとなるため、検索とコンテンツ配信は小規模なサーバ構成でも対応可能と判断(1CPU、MEM6GB、HDD500GB)。初期導入コストを削減。
- 収集した震災資料をデジタル化するために「デジタルアーカイブの構築・連携のためのガイドライン」の5章、2項「デジタルアーカイブをつくらう」を参考にスキャナを準備。
- 青森震災関連アーカイブでは幅広いユーザインターフェイスへの対応を検討するために検証用タブレット端末(iPhone、iPad、Android端末を複数台)を用意。
- コンテンツの保存に関しては、大容量の遠隔地バックアップという観点からリムーバブルストレージシステムRDX(1TB)を選定。
- 必要最低限の停電対策として1500VAのUPSを設置し、停電時にシャットダウンを行うよう設。

- 選定したサーバ機器は後年運用におけるコンテンツデータの増加を考慮にいれ、拡張可能(CPU、メモリ、HDD)である。

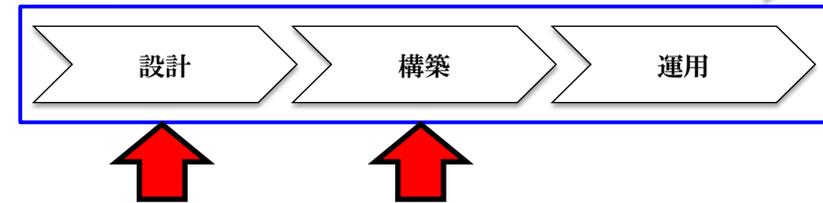
- 情報発信のための国際標準規格としてZ39.50、OAI-PMH、またそれに準じた規格として、一般的に利用されているSRU/W、Open Searchを採用することとした。これら規格を実装したシステムを実現するため、アーカイブ領域で実績を多く持つ商用ソフトウェアを採用。
- 今後のメタデータの項目追加等の拡張性を考慮し、XML形式でデータを保持。データはRDBMSで管理。
- XML対応の検索エンジンであるOpenText7を採用する。
- バックアップは重要であるため、多くの実績を有する商用ソフトウェアを採用。
- 一般的な公開サービスと同等なセキュリティ対策ソフトウェアを導入。

- スマートフォン・タブレット端末向けのインターフェイスを開発し、検証する。

2.システム

システム

システム構築のプロセス



機能

- 震災関連アーカイブの基本的要点を満たすため、仕様書記載の機能を全て提供する。
- 将来的なメタデータの項目追加等の拡張性を考慮し、XML形式でデータを保持し、設定による項目の追加・変更を可能としている。
- 震災関連コンテンツはWeb配信に適したファイル形式に変換する。元ファイルはオリジナルファイルとして保存する。
- 年表(TimeLine)や地図(Google Map)上から震災関連コンテンツを探し出せる利用者インターフェイスを提供する。
- 利用者の検索性向上のために、検索キーワードのサジェスト機能を提供する。
- 動画や音声の配信に対応する。
- 画像、動画、音声、文書等はそれぞれ配信形式を決定し、国立公文書館「電子公文書等の管理・移管・保存・利用システムに関する調査報告書」の長期保存フォーマットを参考とし一般的なアーカイブに適用可能なフォーマットを採用する。また、長期保存のためのメタデータをシステムで極力自動付与できることとした。
- 幅広いデジタルコンテンツのフォーマットに対応するために、ファイル変換のためのフリーソフト(ImageMagic、wvText、docx2txt、ffmpeg、iconv等)を組み合わせて利用する。

- 現時点で特になし。

データモデル(メタデータ)

- 震災関連アーカイブに採用するメタデータスキーマは、震災標準メタデータ、およびNDL震災標準メタデータを参考とし、また幅広い情報発信と相互運用性確保のため、国際的な標準語彙をできるだけ利用する方針とした。(Dublin Core, FOAF, PREMISを採用。また国内のメタデータ語彙としてDCNDLを採用。)
- NDLとのメタデータ連携のために青森プロジェクトのメタデータ形式から連携メタデータへの変換機能を実装する。
- 東日本大震災アーカイブとの連携プロトコルとしてはOAI-PMHを採用する。
- 連携メタデータはRDF/XML形式で表現可能なようにして管理する。
- 青森プロジェクトでは外部サービスであるメタデータスキーマレジストリ(MetaBridge)にメタデータスキーマを登録する。

- 青森プロジェクトのメタデータスキーマに使用する分類体系として、他の運用モデルプロジェクトでの成果も参考とする予定。

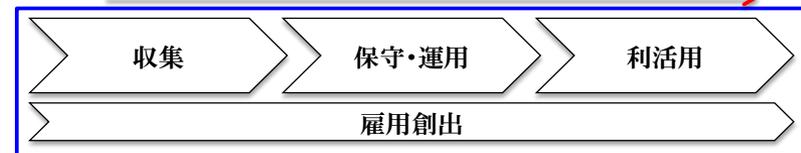
現状の決定状況

今後の想定・課題

3.体制・保守運用

体制

運用フローの中で体制を検討すべきポイント



収集

保守・運用

利活用

現状の決定状況

今後の想定・課題

- 青森プロジェクトではコンテンツ収集において、残存する震災資料をできるだけ漏れなく収集するものとする。
- プロジェクト構成企業・団体の協力のもとに、中でも青森県八戸市を中心とした震災資料を保有する企業、団体等に広く依頼していく。
- 震災直後から復旧段階、復興段階の資料を網羅的に収集するだけでなく、震災前の防災関連情報も収集し、比較検討できるように収集を行う。

- 現時点で想定しているコンテンツと収集済コンテンツの差異が発生しているが、これを解消するためにはコンテンツの分類方法の見直し、もしくは、収集コンテンツ数の拡大が必要となると考えられる。この場合、来年度以降も継続してコンテンツの収集が必要となる。
- 但し、コンテンツの収集は、収集作業及びデジタル化、メタデータ付与に関し相当程度の工数を要し、1点当たり数千円の費用が発生すると考えられるため、継続的なコンテンツ収集に当たっては別途予算の確保が必要となる。

- 運用管理を行う学術機関にネットワーク構築を行いサーバーを設置済
- 運用管理を行う学術機関にてサーバー及びネットワークの運用を行う。

- 次年度以降の運用にあたり、運用を行う学術機関でのアーカイブシステム設置に伴う電力料、インターネットプロバイダ料金及び、インフォコム社での遠隔での保守・運用に関する費用が年間数百万円単位で発生すると考えるが、これらの資金的手当ての目途が立っていない。

- 行政機関をはじめ、企業や学術機関等の団体及び一般住民から震災資料を収集し、有識者や行政関係者のアドバイスを受けながら、コンテンツの整理を進める。
- 震災直後から復旧段階、復興段階の資料を網羅的に収集できる青森プロジェクトの特徴を活かし、震災前の防災関連情報も収集し、比較検討できるように整理していく。
- インターネット上で利活用するだけでなく、展示会などでも収集したコンテンツを公開するような利活用方法を検討する。

- 以下の利活用方法が想定される。
- 学術機関での防災に関する研究資料としての利活用
 - 青森県民による郷土学習用コンテンツとしての利活用
 - 地方自治体における災害対策計画立案時の参照資料としての利活用
- 実際に想定される利活用方法が実現可能かどうかについて、コンテンツがある程度集まった段階で各主体に検証の協力を依頼する必要があると考える。

資料 4-2

岩手プロジェクトの中間とりまとめ

2013.1.23

1.0.岩手プロジェクト全体概要

本プロジェクトの特徴

【目的】

被災地域の中でも地震・津波によって市街地全域および行政機関が甚大な被害をこうむった陸前高田市、大槌町を中心に、震災関連コンテンツを収集する。震災によって資料の流出が想定される一方で、震災後の資料、記事や写真等、膨大な量の震災関連コンテンツを所有していると思われるが、資料の整理がなされておらず、震災関連資料の所在とその種類などの調査が必要である。

また、壊滅的な被害にあった歴史ある町並みや、人々の離散によって失われつつある固有の伝統や文化、知恵等、文化的価値の復興にも焦点をあてて、記憶や記録の残し方を、ヒアリング及びワークショップなどの手法を通じて実証する。特に、記憶の可視化や対話によるアーカイブについては、住民の意見や意向を反映することを特徴とする。

【方針】

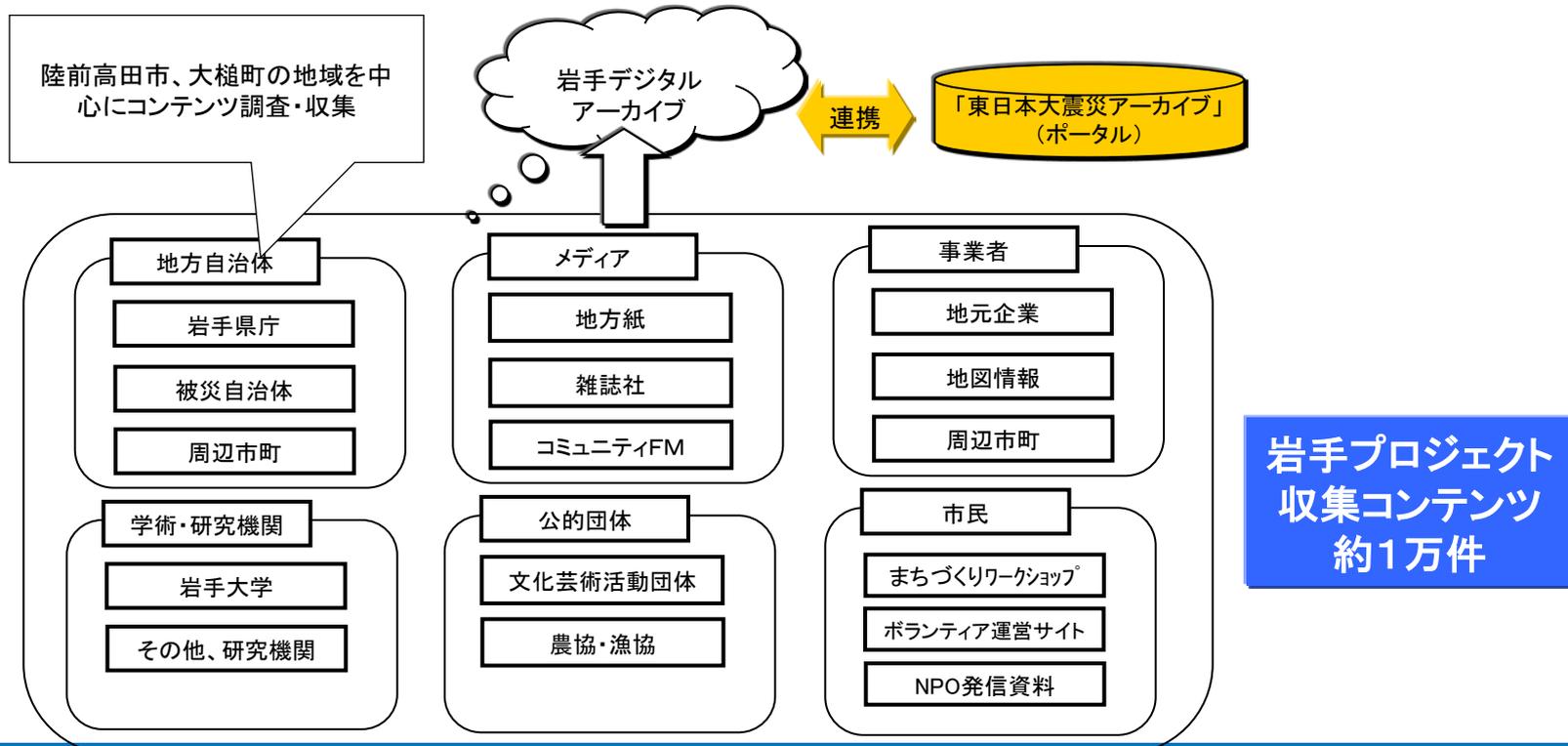
- ・収集時、コンテンツ種類のカテゴリごとに提供先のコンテンツ量や整理状況などから優先順位をつけて調査・協力打診を行う。
- ・調査の回答回収100団体程度、コンテンツの収集40団体程度を目安とする。
- ・震災関連コンテンツは陸前高田エリアに重点を置いて収集する

■コンテンツ収集の方針

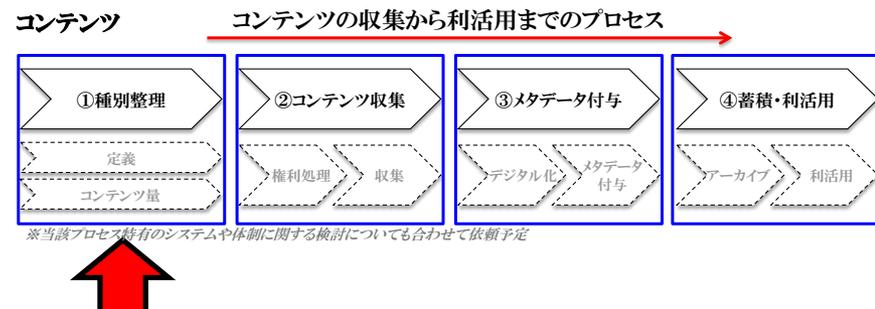
本プロジェクトでは、東日本大震災に関して、記録・記憶が必要だと考えられる震災資料を領域や分野に捕われずに、震災資料の収集先の体系を調査し、広く収集先を特定していく。収集にあたっては、岩手県、岩手大学、および被災自治体の陸前高田市、大槌町の協力により、資料提供及び、企業、マスメディア、NPOからの資料提供を基に収集活動を実施する。

期間の限られた中での実証にあたり、以下の点に留意して収集活動を行う。

- 甚大な被害を被った地域(陸前高田市・大槌町)に関する情報は重点的に収集先の調査およびアーカイブモデル構築を行う
- 破壊された歴史ある町並みや、消失していく可能性が高い固有の文化伝統知恵の記憶、記録収集を優先的に実施する
- 岩手県や被災自治体の情報拠点整備等への利活用を考慮したモデル作りを目指す



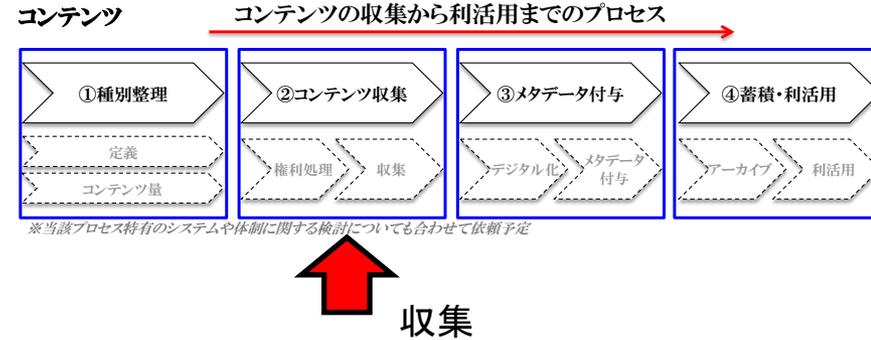
1.1.コンテンツ収集動向



No.	内容	想定種別	想定点数	実績種別	実績点数
1	震災以前のまちなみ、文化伝統等 震災後の被災状況	新聞縮刷	2,000	PDF	20
2	震災後の支援活動	写真	2,000	JPEG	20
3	震災前の風景、行事、観光地	写真	2,000	JPEG	20
4	震災後の自治体情報	文書	1,000	紙媒体	20
5	災害FM放送記録	音声	100	MP3	20
6	新聞記事	文書	200	html	20
7	震災後の被災状況	写真	500	JPEG	20
8	震災前、震災後の写真	写真	500	JPEG	20

- 150件の収集先調査のうち、100点以上コンテンツを収集する大規模収集先の選定を行っている。
- 12月前半には、そのうち上記8団体から20サンプルを収集し、メタデータ付与、デジタル化についてのフローの確立、検証を行う。
- 12月後半～1月より、その他12件の大規模収集先、20件程の小規模収集先に対して、本格的にコンテンツ収集を行う。

1.2.コンテンツ収集における方法・課題



権利処理

【同意許諾書】

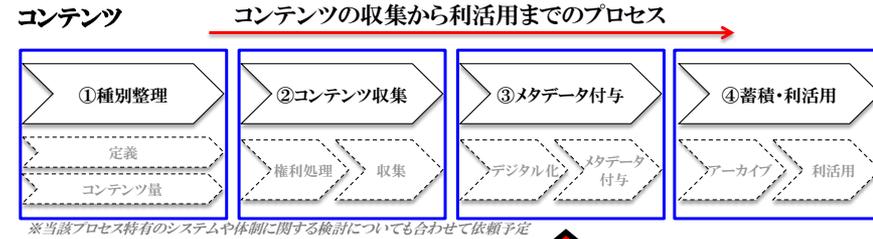
- 3つの公開レベル「ID付与者のみ(研究目的等)公開」「一般公開」「非公開」を想定し収集を行う。
- コンテンツ使用許諾については、収集時、またはそれ以降に口頭説明し、同意許諾書をもらうこととしている。
- NPO等では、コンテンツの提供者と著作権者が異なることが多いので、提供者の協力をえて、権利処理を検討する。

- 甚大な被害を被った地域であるがゆえに、震災前の固有の文化伝統知恵の記憶、記録をアーカイブの対象としており、人の写った祭りの写真等、肖像権等の処理がしえないものが多数存在する。ぼかしをいれるか、非公開とするか、地域のみでの限定とするか検討が必要。
- 図書館では自治体の震災前の書類が存在するが、図書館内のみ利用を前提にしているため、自治体の許可をえる等の手続きが必要である。
- WEBサイトで利用している人物写真について、二次利用許可の取り直しの検討が必要。
- 権利処理の状況により、公開／非公開のファイルと判断する基準。写っている内容によっても上記の判断の指針の策定。

- 収集先リストを洗い出し、コンテンツの所在調査・収集フローを設定し、収集計画を設計。
- コンテンツのカテゴリ、保有形態、時系列等を設定し、大規模(100点以上)の収集先と、小規模収集先に分けて、コンテンツ種類を網羅的に集めて実証実験を行う。
- 収集体制としては、今後も継続的に活動するNPO、企業、団体と連携を図っている。
- 「コンテンツ保有状況の調査票」を作成し、初回訪問時にディレクターによるOJTを行って、ヒヤリング内容の均質化を図った。加えて、日報により活動状況の管理を行う。

- 被災自治体の場合、被災直後の重要な情報整理ができていない。
- 全国の自治体、広域連合等が救援・救護をした報告書が存在するが、集められていないのが現状。記憶とともに薄れている被災直後の資料、記憶については、重要度の高いものから優先順位を決めて、来年度以降にデジタル化する必要がある。

1.3.メタデータ付与における方法・課題



■メタデータ付与における方法・課題

デジタル化

- 図書館等から収集した文書関連(ちらし、パンフレット、報告書、論文)について、デジタル化を実施する。
- 文書関連のデジタル化に関して、国立国会図書館「資料デジタル化の手引き2011年版」の①保存用画像の項において300dpi～400dpiと規定されており、文書については作業効率を優先し、また容量の制約などを考慮しスキャナでの読み取り解像度は300dpiとする。
- 保存形式は、PDFやJPEG形式を想定している。
- その他画像、音声、動画の保存形式はMP3、MPEG4形式などを想定している。

メタデータ付与

- 付与についてはコンテンツ提供先でメタデータをコンテンツ毎に付与するパターンと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するパターンを想定している。
- メタデータの必須項目としては、「タイトル」「作成者」「提供者」「提供者区分」「説明・要約」「場所」「時期」「権利関係」「権利関係」「分類」等を設定。
- 検索しやすいメタデータ(キーワード、タイトル等)を付与する方針で対応する。他アーカイブの事例から「場所」に関しては市町以下の地区名も入力することとし、詳細検索区分を実現する。
- チェック機能の一元化とメタデータ付与マニュアルを作成して、作業の(効率化と)均質化を図る。
- 入力ミス減少させるため入力フォーマット(一部プルダウン)を活用する。
- タイトルは検索性能向上のためメタデータの必須項目として付与する。

- サンプルコンテンツのデジタル化のテストを行い下記を検証した。
 - ①デジタル化処理工数(時間、費用等)算定
→実証実験に必要な処理量算定
 - ②音声データの品質確認(アップする際の基準)
 - ③写真データの品質確認(アップする際の基準)
 - ④動画データの場合全部確認に必要な工数(時間)と実際確認作業との調整
- 利活用の用途により、上記デジタル化の基準づくり(利活用によって基準も異なる)を検討する。

- コンテンツとメタデータのつけあわせ管理手法の確立を検討する。
- メタデータを付与、整理するための人件費等、コストを算定する。
- 協力者について費用支払いの場合の対処について、団体の状況にあわせて検討する。
- 顔が載っている、団体名等が特定できる画像などについてどこまでの範囲なら、肖像権などをクリアーできるのか基準の検討を行う。
- 位置情報、時間、内容などが特定できないコンテンツのメタデータ付与をどのように行っていくか検討する。
- 情報セキュリティ管理を徹底する。

■ デジタル化の例

デジタル化の体制

- 専門スキルを持ったスタッフ2名で実施する。
- 文書処理300dpiのスキャナーを整備している。
- 当該自治体に所在する専門デジタル会社と協働し、ベースキャンプとしてコンテンツ収集、管理、デジタル化、サーバーアップ等の処理を行う。
- 品質管理については凸版から専門員を派遣し、現地にて管理体制を構築する。
- マニュアルにて、各種データのデジタル化の作業フロー等を構築する。

デジタル化の手法

- 下記にてサーバーアップを行う。
 - ①文書関連(ちらし、パンフレット、報告書、論文)
→処理スピードとデータ容量から、アナログからデジタル化処理について300dpiスキャニングを採用し、JPEG、PDFにデータ変換する。
 - ②音声データ(災害関連FM放送)→mp3
 - ③写真データ(メインとなるコンテンツ)→JPEG、PDF
 - ④動画データ →Windowsmedia、Quicktime及び変換ソフトを利用しMPEG4

■メタデータ付与の例

サンプルテスト

著作権未処理のため削除

著作権未処理のため削除

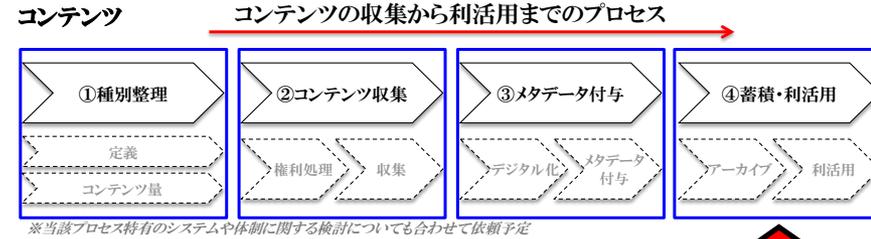
メタデータ付与の体制

- 現地雇用を重視して、現地企業から雇用をする体制を構築している。
- 現地に品質管理の専門員を派遣し管理体制を構築する。

メタデータ付与の手法

- メタデータ付与はコンテンツ提供者付与と収集後付与の2パターンを想定している。
- メタデータ入力は専用入力フォーマットで入力ミス防止のためプルダウンを採用し、必須項目については空白チェック、その他属性指定のある項目については属性チェック機能の設定を検討する。
- 活用しやすいメタデータの付与を目指し、ケーススタディ方式のキーワード、タイトル付与方法を検討する。
- コンテンツ所有者に直接ヒアリングを行うことは、コンテンツの情報整理に有効である。

1.4.蓄積・利活用



現状地域で想定される利活用

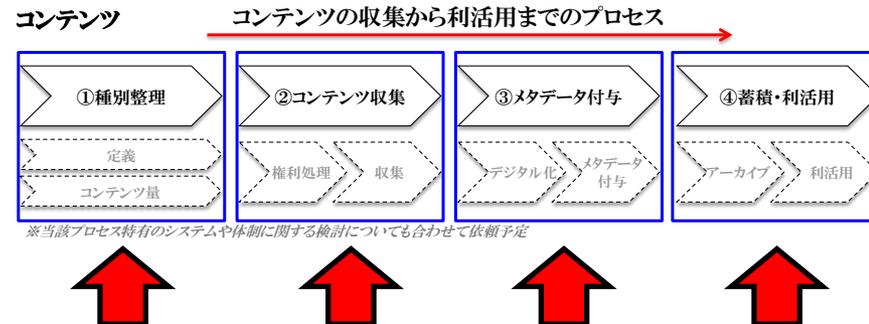
【自治体、地域団体の意見】

- 地域のコンテンツは、「地域資源」という考え方をしており、どこでも見れるのではなく、地域に来てもらってはじめて、見てもらえるようにしたいという意見が多い。
- 震災メモリアルパーク構想している自治体もあり、施設のコンテンツ活用を検討している。
- NPOは被災地ツアーに写真を活用しており、観光振興の素材としての活用を検討している。
- 防災検証事業を推進している自治体もあり、被災状況の記録集作成に、アーカイブコンテンツの活用の要望がある。(2次使用許可について検討課題あり)
- 商工会では、融資の相談で、過去の店舗の様子などの写真を提出する予定があり、被災前の資料は活用できる。また、震災前の記録から、復旧のあゆみを行政とともに記録し、書籍として発行したい。
- 沿岸エリアの共有の財産として、広域での防災検証をはじめ、防災計画や危機管理として役立つようにしたい。また、一地域ではなく、沿岸エリアの共有の財産として、広域での防災検証としても活用し、沿岸防災計画に役立てる。
- 小中学一環で町の伝統芸能や地域産業、防災教育などを学ぶ科目で活用したらよいのではないかと意見もある。

【調査計画】

- コミュニティリーダーに、アーカイブ運用、利活用にかかわる意向を調査し、利用者と利活用ニーズを整理する。
- 特に重要と思われる利活用(観光、教育を想定)については、観光ツアーを行う語り部や教育関係者、自治体を対象にして、アーカイブしたコンテンツを公開し、利用者評価の調査を行う。
- 後年運用の観点から、検索サイトを地域の方々に使っていただき、ユーザビリティ等の実証検証を行う。
- 地域での利活用を活性化させるために、「地域での活用」の限定利用、限定公開の方法を検討する。
- 岩手県の自治体に対しては、震災関連資料の収集とデジタル化の実態調査アンケートを行い、岩手県における東日本大震災アーカイブのあり方を検討する。(12月～1月アンケート協力依頼、実施予定)

1.5.その他



その他特記事項

- 広域コンテンツホルダーへの依頼について
複数の県にまたがるコンテンツを保有すると思われるホルダーへの対応に苦慮。地域サーバ以外で一括集約する等の検討とするか、地域サーバで格納するかの検討を行いたい。
- 後年運用に関して
システム運用拡張、メタデータメンテナンス追加、コンテンツ作成収集、ワークショップ等の後年運用の主体(バラバラに又は一括で)の調整。後年運用の為の費用拠出、自治体対応の推進についての検討を行う。

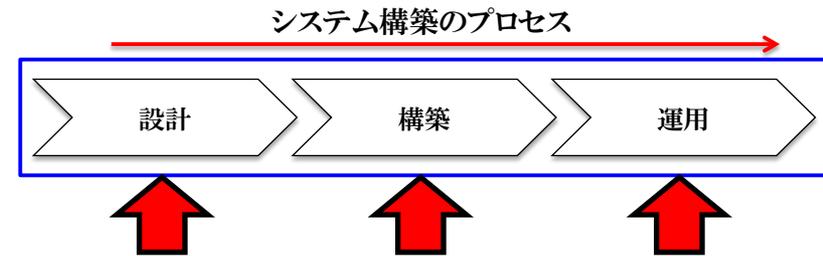
2.システム

運用モデルの特性

方針

- 岩手プロジェクトでは、主にヒアリングやワークショップの手法を通じてコンテンツ収集することで、被災の記憶、記録を残していく。
- 特に被害の大きい陸前高田市、大槌町より多くのデータを収集し公開する。
- サーバは岩手大学に設置し、ネットワークポリシー等は、岩手大学に準拠する。

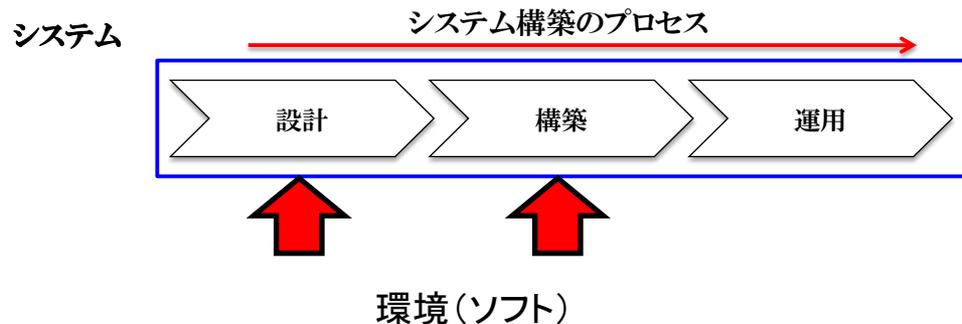
システム



考慮点

- 動画、音声、画像ファイルについて、1万件を登録できるストレージを準備する。また実証調査後も登録できるよう余裕をもった容量とする。
- コンテンツ収集先は地域団体、マスコミ、企業など多数となるため、様々なファイルフォーマットにも変換、保存が可能なシステムとする。(元ファイルもオリジナルファイルとして保存)
- またコンテンツの種類も多数なので、検索の利便性向上のため、メタデータの言葉の統一等を配慮した設計とする。
- 自治体がメタデータを作成することを考慮して、入力補助機能を実装したシステムとする。

2.システム



環境(ハード)

- ・ サーバは、共同実証をおこなう岩手大学に設置。
- ・ 対障害性として、ディスクはRAID5で構成、電源についても冗長化構成を採用。業務系のシステムではないので、機器自体の冗長化はおこなわない。
- ・ 実証調査で収集するコンテンツ1万点を考慮し約2TBの容量を確保。
- ・ 必要最低限の停電対策としてUPSを設置、停電時に正常にサーバをシャットダウンさせるよう設定。
- ・ 長期保存と大容量を考慮し、データのバックアップはテープオートローダー(Ultrium5 データ・カートリッジ)を採用。

環境(ソフト)

- ・ 仕様書にある基本的性能とアーカイブ機能(公開機能/登録機能/管理機能)、そして各種プロトコルによる連携機能において、多くの実績を有する商用パッケージソフトを選定。
- ・ 今後のメタデータの項目追加等の拡張性を考慮し、XMLベースでのデータ保持を可能とする。
- ・ バックアップはテープオートローダーに対応し、実績のあるNetVaultを採用。
- ・ ウィルス対策ソフトを導入。
- ・ 遠隔地からの保守対応を考慮し、サーバ、アプリケーション、ネットワークの障害およびパフォーマンスの監視が可能なソフトを導入。

- ・ Disk容量が足りなくなったときの拡張方法。

- ・ 脆弱性などのセキュリティ上の脅威への対応。

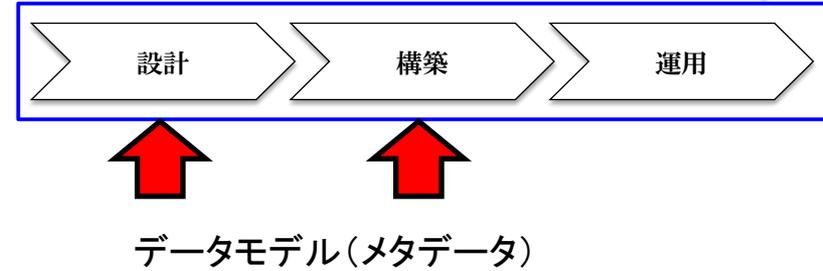
現状の決定状況

今後の想定・課題

2.システム

システム

システム構築のプロセス



機能

データモデル(メタデータ)

- ・ コンテンツはインターネットでの配信に適したファイル形式に変換し、元ファイルはオリジナルファイルとして保存。
- ・ 年表や地図(Google Map)上からコンテンツを検索できる利用者インターフェイスを提供。
- ・ 利用者の検索性向上のために、検索キーワードのサジェスト機能を提供。
- ・ 動画や音声の配信に対応。
- ・ 登録作業を簡略化するため、メタデータやコンテンツを一括で登録・更新できる機能を提供。
- ・ 画像、動画、音声、文書等はそれぞれ配信形式を決定、国立公文書館「電子公文書等の管理・移管・保存・利用システムに関する調査報告書」の長期保存フォーマットを参考とし一般的なアーカイブに適用可能なフォーマットを採用。また、メタデータをシステムで極力自動付与できるようにした。
- ・ 幅広いデジタルコンテンツのフォーマットに対応。

【メタデータスキーマ】

- ・ 震災標準メタデータ、およびNDL震災標準メタデータを参考とし、国際的な標準語彙をできるだけ利用する方針とした。(Dublin Core, FOAF, PREMISを採用。また国内のメタデータ語彙としてDCNDLを採用。)
- ・ 区分の大・中など階層を意識して設定が必要な項目には、大・中で一意なコードを定義し、コードからシステムで逆引き設定する方式を採用。
- ・ 場所(住所)での検索を正確におこなうために、よみをセットとし、メタとして設定する内容をリスト化。

【メタデータ連携】

- ・ NDLとのメタデータ連携のために連携メタデータへの変換機能を実装。
- ・ 連携プロトコルとしてOAI-PMHを採用。

- ・ 実証調査で発生する追加機能要件等の対応方法。

- ・ 検索を正確におこなうため、入力する内容をリスト化しているが、入力時に正しくリストから選択する方法。
- ・ 入力メタデータのエラーチェック方法。

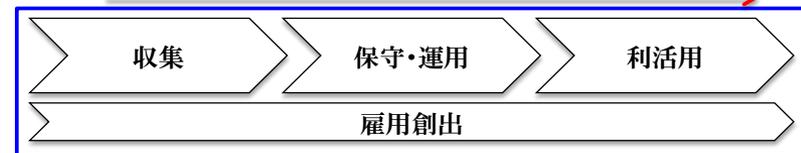
現状の決定状況

今後の想定・課題

3.体制・保守運用

体制

運用フローの中で体制を検討すべきポイント



収集

保守・運用

利活用

現状の決定状況

今後の想定・課題

- 地域NPO、コミュニティの協力による収集体制の整備に向けて検討中である。

- 今年度は、自治体では、震災関連コンテンツの所在調査を主にしているが、特に震災直後の被災状況、救援・救護、復旧については、貴重な紙文書(職員の記憶等含む)情報が、当該自治体、救援支援した自治体等に保管されている。これを来年度以降、デジタル化する必要がある、継続的なコンテンツ収集体制の整備が必要。

- 実証期間中サーバーの保守、運用については岩手大学にて実施協力により行う。
- 実証実験用として、岩手プロジェクトでのメタデータを決定し、デジタル化までの一連のテスト行い、トレーニング、マニュアルの整備が必要。

- 後年運用システム実施主体の検討
サーバー管理、システム変更、メタデータ変更メンテナンス等

- 岩手大学と継続的な研究、体制について検討を行う。

- 地域のためのコンテンツ活用として、商業者、地域NPOによるビジネス利用について、権利処理、地域限定活用の方法について検討が必要。

資料 4-3

宮城(東北大・河北新報)プロジェクトの中間とりまとめ

2013.1.23

1.0.宮城東北大学プロジェクト全体概要

東北大学は、被災地の大学として、震災直後から多くの研究者が被災地に入り、震災関連コンテンツを収集し、研究に活用すると共に、アーカイブに関するシンポジウムを開催するなど様々な活動を行っている。

また、既に震災関連デジタルアーカイブ「**みちのく震録伝**」を運営し、収集した一部のコンテンツを公開している。

宮城東北大学プロジェクトでは、この「みちのく震録伝」をベース基盤として、東日本大震災アーカイブ(NDL)へのデータ連携の実装や、実証調査の仕様に即した機能の実装を図り、より広いコンテンツの利活用を目指すものとする。

・研究者視点での震災関連コンテンツの収集・公開

震災後まもなく、被災地に入り、収集した写真・動画・オーラルヒストリなど、研究者視点で収集した多くのコンテンツを整備し、公開することにより、防災・減災研究の推進に寄与するものである。

・災害科学国際研究によるコンテンツの利活用の推進

災害科学国際研究所に所属する様々な分野の研究者により活用されるため、将来に向けて研究者の要望を取り入れやすい環境にあり、震災関連デジタルアーカイブを柔軟に拡張することができる。

・充実したオーラルヒストリ記録促進

人々の記憶が薄れる前に、被災地域の雇用により、オーラルヒストリや記録をまとめる作業を組織化し、収集コンテンツをアーカイブする。

・東北大学附属図書館の収集コンテンツの公開

東北地方を中心とした図書館グループの震災記録収集活動(右図2参照)と連携し、電子化された情報を「みちのく震録伝」を介して、東日本大震災アーカイブ(NDL)に連携する。

・震災関連コンテンツの分析基盤の提供

河北新報社から提供される1991年以降の全新聞記事を東北大学の分析基盤に登録する。東北大学では、これらの新聞記事に対してテキストマイニング技術を活用し、阪神・淡路大震災との比較や、被災状況、発生事象など、様々な研究に役立つ基盤を提供する。なお、河北新報社と東北大学においては、協定により、新聞記事本文の参照が可能な環境となっている。

・オープンソースと商用システム双方を活用したアーカイブシステムの構築

コンテンツを公開、登録、管理する機能についてはオープンソースを活用し、東日本大震災アーカイブ(NDL)との連携機能については、実績のある商用システムの導入することで、システムの安定性と、将来の機能面での拡張を最大限に考慮したシステム構成を取る。

図1 「みちのく震録伝」ホームページ



「みちのく震録伝」では、左記のホームページを立ち上げ、震災に関する様々な情報の発信を行っている。

<http://shinrokuden.irides.tohoku.ac.jp/>

図2 図書館グループの活動



【図書館グループ キャンペーン抜粋】

震災では甚大な被害が生じ、復旧への道のりは未だ道半ばですが、被災地での生活、救援活動、復興へ向けた取り組みなどの中から、数々の震災の記録が生まれてきています。

このキャンペーンに参加する図書館では、このような震災の記録を広く収集し、今後永く保存することにより、震災で得た教訓を後世にしっかりと引き継いでいきます。

収集した記録は、防災・減災といった視点だけでなく、今後の私たちのあり方を考える上でも、必ず役に立つものと考えています。

つきましては、皆様がお持ちの震災記録を、ぜひ図書館にご寄贈くださいますようお願いいたします。

1.0.宮城河北新報プロジェクト全体概要

河北新報社は被災地域の報道機関として、東日本大震災の発生直後から、被災地域での取材活動を継続的に続けており、新聞記事や報道写真をはじめ、膨大な量の震災関連コンテンツを所有している。

また、河北新報社が運営する地域SNS「ふらっと」に投稿写真サイトを設けたり、NPOや学生ボランティアと共に被災地に入り、その様子をブログにまとめたりするなど、報道とは別の角度からも、一般市民の証言や記録の収集に努めている。

実証実験に当たっては、これらの震災関連コンテンツの永続的な蓄積とともに、より多くの方々に適切な情報が共有できるよう東日本大震災アーカイブ(NDL)とのデータ連携を実現し、世界の防災や減災研究の発展に寄与するものとする。

なお、新聞社として取材し、記録した記事・写真は有償で提供していることもあり、本実証プロジェクトを通じて、公的機関が運営するシステムとの適切な境界線を確立し、有償コンテンツとの切り分けについても整理するものとする。

・被災地域の新聞社が有する膨大な量の震災関連コンテンツ流通の仕組み作り

震災関連の新聞記事の提供(本文は有償)により、震災当時、被災地で何が起っていたかを新聞記事を通して、把握することができる。また、その新聞記事を補足する写真や動画との関連付けを表現することにより、将来の防災・減災研究の促進に寄与する。

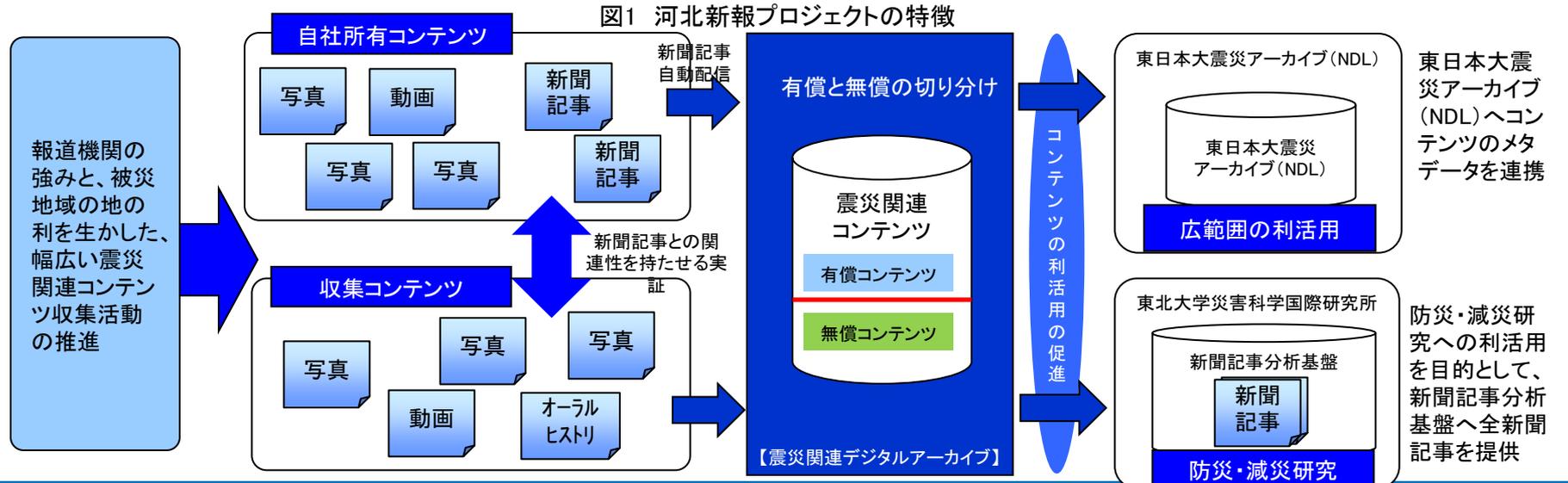
・既存システム(商用)と震災関連デジタルアーカイブ(無償)との有機的な結合と境界線の確立

マスメディアが所有する既存の商用データベースと、本実証事業で構築する震災関連デジタルアーカイブとの連携を、新聞記事(見出し、タグ情報など)の自動配信の仕組みにより有機的に結合する。また、有償コンテンツと無償コンテンツの境界線を明確にする。

・報道機関の強みと、被災地域の地の利を生かした、幅広い震災関連コンテンツ収集活動の推進

報道機関としての機動力と、地域コミュニティとのリレーション、被災地域の地の利を最大限に活用し、地域で活躍するボランティア団体などからのコンテンツ収集や、一般市民からの投稿促進のための施策により、幅広い震災関連コンテンツの収集が可能。

図1 河北新報プロジェクトの特徴



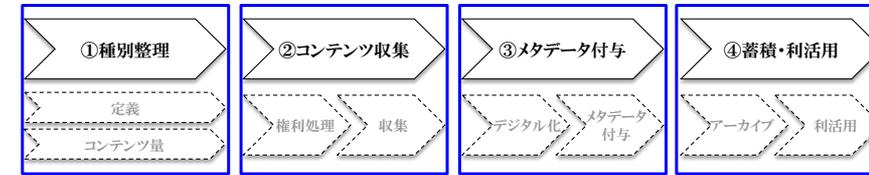
1.1.コンテンツ収集動向

(※1)動画形式

.MOV .AVI .3GP .wmv .MP4 .mp4 .m4v .mif .MTS .MPG .m2ts

コンテンツ

コンテンツの収集から利活用までのプロセス



※当該プロセス特有のシステムや体制に関する検討についても合わせて依頼予定

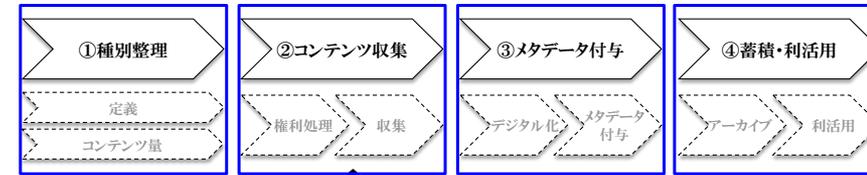
No.	プロジェクト	定義	内容	想定種別	想定点数	実績種別	実績点数
1	河北新報社	震災関連新聞記事見出し	震災を伝える新聞記事の見出し	テキスト	約15,000	テキスト	20,000点以上
2	河北新報社	新聞未掲載・記者撮影写真	取材現場で撮影した写真	写真	約100	JPEG	8255点 (内100点選定予定)
3	河北新報社	河北社内システム掲載写真	河北新報社が運営するSNSサイト掲載写真	写真	約300	JPEG	約800点
4	河北新報社	各種ボランティア団体提供写真	ボランティア団体より提供された震災関連の写真・動画	写真・動画	約10,000	JPEG 動画 (※1)	11,000点以上
5	東北大学	研究者の写真	研究目的で収集した震災関連写真	写真	約24,000	JPEG	約23,000点
6	東北大学	みちのく・いまをつたえ隊	宮城県沿岸部の16市区町村で撮影した写真と被災者へのインタビュー	写真 音声	約76,000	JPEG	約27,000点
7	東北大学	東北大学付属図書館収集	収集した震災関連のチラシ・パンフレット	アナログ	約1,400	TIFF又はpdf	約1,440点

1.2.コンテンツ収集における方法・課題

東北大学

コンテンツ

コンテンツの収集から利活用までのプロセス



※当該プロセス特有のシステムや体制に関する検討についても合わせて依頼予定



収集

権利処理

収集コンテンツの特性により、利用許諾の取得の仕方に違いがあるものの、権利上の取り扱いが難しいものは存在しない。

【被災者へのインタビュー】

•被災者のインタビューを取得時に、利用目的を伝えた上で、収集時に許諾の取得している。

【チラシ・パンフレットの収集】

•チラシ掲載の連絡先へ、事前に電話にて許諾の意思を確認し、後日、郵送にて許諾書を送付し、返送頂く。

- 被災地域在住の方へ収集をお願いし、被災地域の写真を撮影すると共に、被災者へインタビューを実施し、音声として記録する。また、活動を理解してもらう為に、チラシを作成し、お店や役所に置かせてもらっている。
- 図書館保有の震災関連のチラシやパンフレットは、図書館職員が日常生活や出張時などに、街中に置いてあるものを収集したり、国内学協会に寄贈して頂いたもの、「みちのくいまをつたえ隊」の活動者が収集して頂いたものである。
- 研究者が震災記録として撮影した写真を個人や学内サーバーに保持しているものを収集。
- 収集したコンテンツは、肖像権、個人情報の観点でチェックを行うと共に、ファイル名の重複の排除を行い、メタデータ付与担当者の作業性を考え、フォルダ分けを行う。

【チラシ・パンフレットの収集】

•電話で事前に許諾依頼をお願いする際に、2回電話をするケースがあり、チラシ・パンフレット記載の連絡先が同じ団体でも違う名前で記載されているケースがあり、名寄せしきれなかった。現在は、電話番号を調べ、電話番号で寄せてまとめて取得する。

•連絡先に電話をすると、別団体や別部署に回されるケースがある。既に、電話をしている団体等については、一度電話をした担当者より電話をし、「度々すいません」などの対応をしている。

- インタビューとして取得した音声データは、そのままでは公開できない。人の語り口や方言があり、要点がまとまった内容ではないことと、内容を確認するのに、膨大な時間が係ることが理由として上げられる。
- 「震災復興」の文字が入っているチラシで、震災との関係性で悩むケースがある。コンサートは対象としたが、商品のチラシは対象外としている。一貫した判断が難しいのが現状である。
- 都心部で収集活動を行うケースでは、住民の理解を得られないケースがあり、コミュニティ主催のイベント等に参加することで、理解を得られ、ようやくインタビューに答えてもらえる。
- 研究者のコンテンツで、GPS情報があり、居住できるレベルの被災状況の建物を撮影しているケースがある。GPS情報より個人を特定出来てしまう可能性がある為、アーカイブ対象外として対応を行う。

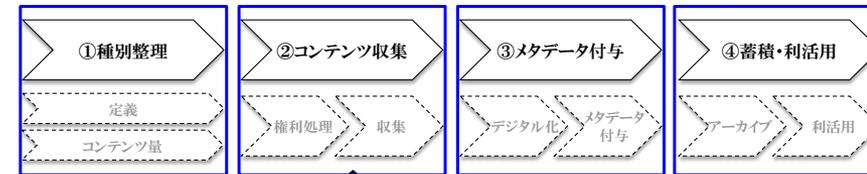
手法・プロセス

現状の課題

河北新報社

コンテンツ

コンテンツの収集から利活用までのプロセス



※当該プロセス特有のシステムや体制に関する検討についても合わせて依頼予定



収集

権利処理

- 震災関連新聞記事や社内保有の写真等については、河北新報社の社内承認を得た上で対応を行うが、各種団体又は個人より受け入れるコンテンツは、権利処理を行う必要がある。コンテンツとしては、デジタル化された写真が主であり、権利上問題となるコンテンツの取り扱いが比較的少ない。使用の許諾に関しては、収集時に口頭、メールで許諾を受け、後日文書を郵送し、使用許諾を得る対応を行う。

- 社内保有の写真等は、各支局より郵送にて送付し収集している。
- 各種団体や個人からの収集は、河北新報社のCSR活動を通じた各種団体との交流の中で、提供を呼びかけると共に、団体の横のつながりにより賛同頂ける方を紹介して頂き、連絡を取る等の対応をしている。
- 提供者へは、事前に電話、メールで主旨を伝え、賛同の意思を確認し、直接会って、主旨や使用許諾の説明を行う。
- コンテンツは、提供者へ取りに行く以外に、郵送やメールで送られてくる他、直接会社へ持って頂けるケースやホームページからダウンロードを行うケースもある。
- メタデータ付与の作業性を考え、頂いたコンテンツの住所をコンテンツより特定し、住所事のフォルダ分けを行うと共に、肖像権や個人情報の観点で公開、非公開の選別を行う。

- コンテンツに写り込んでいる被撮影者から、公開することの了解が得られているか、確認が難しいケースがある。また、個人が特定されてしまう情報(表札、ナンバープレート)への対処も課題であり、写っている文字部分を伏せるなどの修正を施すか非公開にする。写り込んでいる人の了解が取れない場合は原則として非公開とするが、不特定多数が出入りするオープンスペースなどで撮影され、特定の個人と識別できない程度であれば公開するというように、報道写真の扱いに準拠して河北新報社が個別に判断していく。
- ボランティア団体のように、複数のボランティアが作業風景を撮影している場合、撮影者である真正著作権者とアーカイブへの提供者が異なることがあり得る。撮影者が、被災地を離れているケースもあり、物理的に許諾が難しい。こういうケースでは撮影者がボランティア団体に写真の所有権を譲渡したものとみなし、河北新報社の責任でアーカイブする。

- 収集したコンテンツには震災に関連しないものが混ざっているケースがあるほか、アーカイブに不適切だったり、同一アングルの写真が多数あったりするため、いただいたコンテンツはアーカイブ前に事前にチェックしている。
- データを取り込む前に、ウィルスチェックを行う。
- 写真で明らかに撮影日時がおかしいものがある。提供者確認を行い、年月日までは、極力特定するように対応する。
- 公開、非公開の判断は、非常に難しく、何処までOKとするのかの判断に迷うケースが多々ある。問い合わせ担当を決めると共に、メンバーで協議し、対応を行っている。
- 提供コンテンツには、動画が提供されるケースがある。アーカイブ対象とするが、内容確認に時間を有することから、本実証調査期間に終わらない可能性もあり、対応を検討する必要がある。

(参考)同意許諾書(暫定版)

河北新報社の許諾書

2012年11月xx日

xxxxxxxx 様

「河北新報震災アーカイブプロジェクト」へのご協力のお祝い

河北新報社編集局デジタル編集部長
八浪英明

拝啓

ますます御健勝のこととお慶び申し上げます。平素は格別のご高配を賜り、厚く御礼申し上げます。

さて、弊社は東北大学災害科学国際研究所の東日本大震災アーカイブプロジェクト「みちのく震録伝(しんろくでん)」などと連携し、市民の皆さまが記録した震災の写真や動画、ブログなどを広く収集・保存し公開する「河北新報震災アーカイブプロジェクト」を進めています。

この秋には総務省「東日本大震災アーカイブ基盤構築事業」のデジタルアーカイブ実証調査に「宮城河北新報プロジェクト」として参加することが決まりました。今後、同省や国立国会図書館なども協力して、アーカイブした写真などを長く保存し、国内外に公開する仕組みを構築し、防災・減災教育なども含めて幅広く活用していく予定です。

また、弊社としても写真投稿サイト「将来への記憶」のリニューアルをはじめ、報道からコミュニティに至る各種サイト・媒体の充実を図り、震災の記録・記憶を次世代にしっかり伝えていく所存です。

つきましてはこうしたプロジェクト展開の趣旨にぜひともご賛同いただき、お千円にお持参の写真や動画などを提供いただきたく、お願い申し上げます。

ご提供いただいた写真・動画などに関しては、アーカイブの趣旨に照らし、できるだけ多くの方に閲覧していただけるよう運用する所存ですが、一方ではプライバシー保護などにも配慮し、「保存はするが非公開」ということもあり得ます。現在、別紙のように使用許諾同意書を作成すべく関係機関と協議中です。書類が整い次第、あらためてお願いいたしますので、宜しくお願いいたします。

敬具

河北新報社東日本大震災アーカイブプロジェクト事務局
(デジタル編集部内)

部長 八浪英明
担当 岩崎真実

電話022-211-1158
FAX022-211-1157
メール: m-iwasaki@po.kahoku.co.jp

東北大学附属図書館の許諾書

平成24年●月●日

各位

東北大学附属図書館長
植木 俊哉
(公印省略)

震災関連資料の電子化及びインターネット公開について (依頼)

日頃より東北大学附属図書館の運営にご支援ご協力を賜り、厚く御礼申し上げます。

当館では、昨年発生した東日本大震災の経験を永く受け継いでいくために、震災関連資料の収集、公開に取り組んでおります。

このたび、貴殿が刊行された資料を、震災の貴重な記録として収集させて頂きました。資料を後世まで永く伝えるとともに、多くの方に資料を閲覧頂くため、ご許可を頂けるものにつきましては貴方で電子化し、当館で作成予定のオンラインデータベースや、「みちのく震録伝」(別紙参照)を用いて、インターネット上にて無償で公開をしたいと考えております。

つきましては、趣旨をぜひともご理解いただき、同封の許諾書に承諾の可否をご記入の上、ご返送いただけますようお願い申し上げます。

なお、誠に勝手ながら、同封の返信用封筒にて平成24年●月●日までにご返送いただけますと幸いです。

記

1. 許諾書
2. 該当資料のコピー (1ページ目のみ)
3. みちのく震録伝について (参考資料)
4. 返信用封筒

□お問合せ先

〒980-8576
宮城県仙台市青葉区川内27-1
東北大学附属図書館総務課情報企画係
Email: kikaku@library.tohoku.ac.jp
TEL: 022-795-5925
FAX: 022-795-5909

(参考)同意許諾書(暫定版)

東北大学の許諾書

No. _____

肖像権承諾書・提供資料使用許諾同意書



東北大学アーカイブプロジェクト

私は、東北大学による東日本大震災アーカイブプロジェクト「みちのく震録伝」(以下、本プロジェクトという。)が行う情報収集活動(みちのく・いまをつたえ隊)に賛同し、本プロジェクトに対して、肖像権・提供資料の利用(二次利用も含む)に同意・許諾します。

●本プロジェクトの目的
本プロジェクトは、東日本大震災の被災地において、震災や被災地の生活の記録や情報の収集をはじめ、住民の方々の現在の暮らしや日頃の考え、未来への思いなど、地域の様々な「欲しい、伝えたい」情報を収集する活動を行うことを目的としています。また、この活動を通じ、被災地を結ぶネットワークづくりを目的とし、その後も「みちのく震録伝」と連携して長く継続できる事業とすることを計画しています。

●許諾の対象となる写真・映像・録音物
東日本大震災に関する復興活動の様子、及び、被災された方や各種関係者へのインタビューなどを収めた写真や映像・録音物など、また被災地の被災者の風景や施設、行事等収めた写真や映像・録音物など。

東北大学 災害科学国際研究所 御中

_____年 ____月 ____日 署名: _____

電話番号: _____

Eメールアドレス: _____

(参考)コンテンツその他

コンテンツ収集事例

- Youtubeにアップされているものを頂いたが、提供者本人のものであった為、問題ないと判断。(河北新報社)
- 地名、建物など目印となるものを撮影し、周辺を撮影することで、整理がしやすくなる。(東北大)
- 道路から見えるものは、撮影OK。敷地内は、許可を取得後撮影する。(東北大)
- HPがあるが、実際は活動を停止している団体がある。逆にHPの更新頻度が少なくても活動している団体もある。WEB上だけで見極めるのは難しい。(河北新報社)
- プレハブ仮設を確認するサイトとして、actizというサイトを活用(河北新報社)
- GPSカメラは、GPSが拾えないケースがある。(東北大)
- 震災時の車両は、ストックされ、持ち主が現れない場合、入札で売却される。その場所を撮影時は、ガードマンよりNGと言われる。(東北大)
- 電話での事前許諾確認時に、他大学からも同じ問い合わせがあり、まとめてくれと言われる。(図書館)
- 許諾について、確認したいと言われるケースは、図書館職員より説明をする。(図書館)
- インタビューを取得時に、機器による録音をおこなっているが、インタビューが終了してから、色々な話をしてくれるケースがある。(東北大)
- 提供して頂いたデータに、提供者の方が写真整理を行う為に、テキストファイルが入っているケースがある。また、提供者の方の業務的なデータが紛れ込んでいるケースもあった。(河北新報社)
- 刊行物で、許諾を取得する対象者が20名のケースがあった。これは、今回対象外とした。許諾書を19名まで取得でき、残りの1名が同意頂けない場合、既に、許諾を取得した方に申し訳ない為である。(図書館)
- MACのPCを使用している方がいた。Windowsと互換性が無い為、専用にHDDを準備し対応した。(河北新報社)
- 写真撮影日付が、明らかにおかしいケースがある。提供者に確認し、対応している。(河北新報社)
- カメラで撮影する際は、カメラの最大解像度で撮影する。後に拡大して見たいケースがある為。(東北大)
- 看板は、保護すべき個人情報ではないので、公開対象とする。(河北新報社)

コンテンツ収集に使用する機器等

- USBメモリ/HDD
- デジタルカメラ
- ICレコーダー
- メール

1.3.メタデータ付与

■メタデータ付与における方法・課題

東北大

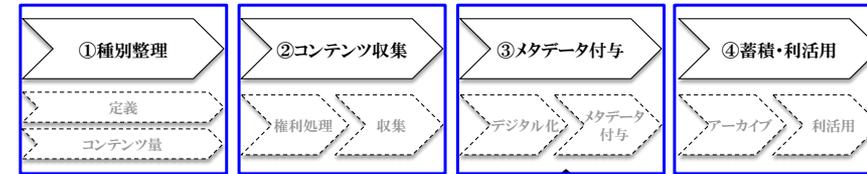
デジタル化

- 東北大附属図書館で収集したコンテンツについて、利用許諾取得後にデジタル化を実施する。
- 文書のデジタル化については、スキャナで読み込む際には解像度は、総務省慣習の「デジタルアーカイブの構築・連携のためのガイドライン」を参考に実際に、試してみると共に、容量を考慮し、300dpiとしている。
- 保存形式は、刊行物の全ページを1ファイルにまとめることができる、pdf形式とする。

- 300dpiのpdf形式は、刊行物の枚数が多いと容量が大きくなる可能性がある。最多枚数の刊行物で試してみて、容量を見ると共に、WEB閲覧用と保存用で分けるなどの検討を行う。

コンテンツ

コンテンツの収集から利活用までのプロセス



※当該プロセス特有のシステムや体制に関する検討についても合わせて依頼予定

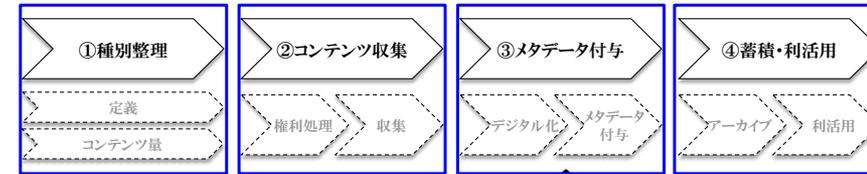
メタデータ付与

- メタデータ付与に際し、実際にSampleで付与をおこなった結果、人の思いこみでメタデータを付与するケースがあった。その為、メタデータ付与においては、見た内容から判断できる内容を付与する方針とし、人の推測を入れない対応とする。
- 写真のケースでは、写真に写っている文字は、メタデータとして付与することで、出来るだけ多くの情報を付加するようにする。
- ツールは、GooglePicasaを採用し、生産性を上げる。
- メタデータ付与は、統制を取る為に、事前に担当者へ説明を行う。
- 質問を受け付ける担当者を設置することで、作業が止まるのを防ぐ。

- 事前に公開、非公開を選別していても、再度、メタデータ付与の段階で、肖像権や個人情報の観点で、公開可能かを質問されるケースがある。ほとんどが、微妙なケースである為、その都度、協議をして決定している。

コンテンツ

コンテンツの収集から利活用までのプロセス



※当該プロセス特有のシステムや体制に関する検討についても合わせて依頼予定

河北新報社

デジタル化

- 河北新報社では、1991年以前の新聞記事をマイクロフィルムで保存している。当実証調査とは別に、デジタル化をおこなっている経緯があり、その状況について、ここでは記載する。
- マイクロフィルムをスキャンし、OCRを使用しデジタル化を実施。(テキスト化)。また、新聞記事紙面イメージを保存する為に、pdf化も同時に実施。
- テキスト化した文字は、旧字体は、新字帯に、漢数字は洋数字に修正することで、見やすくしている。

手法・プロセス

- OCRでテキスト化した結果、連続した文章でテキスト化できない事と誤字も存在し、結果的に人で修正を行う必要がある。テキスト化については、最終的に、人が記事を見てテキスト化する作業を行うことで、対応を実施する。

現状の課題

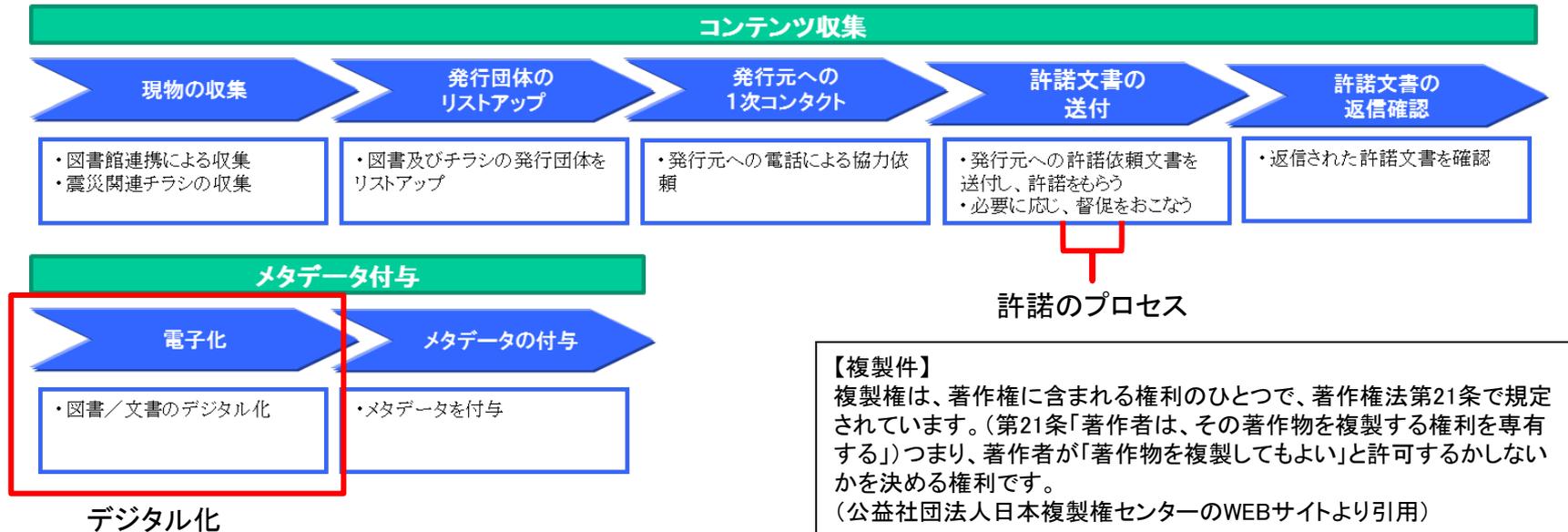
メタデータ付与

- メタデータ付与は出来るだけ、正確に付与する方針で進める。住所、場所、施設名等は調査し、正確に付与する。
- ツールは、GooglePicasaを採用し、生産性を上げる。
- メタデータ付与は、統制を取る為に、事前に担当者へ説明を行う。
- 質問を受け付ける担当者を設置することで、作業が止まるのを防ぐ。
- Keywordは、意味を調べて正確な言葉を設定する。
- Keywordは、一般の人がわかる単語を設定する。
- 周知事項は、ホワイトボードを活用し、周知する。
- 写真に写されている、河川等の地図から見て判断できるものは、撮影場所と合わせて確認し、付与するようにする

- 事前に公開、非公開を選別していても、再度、メタデータ付与の段階で、肖像権や個人情報の観点で、公開可能かを質問されるケースがある。ほとんどが、微妙なケースである為、その都度、協議をして決定している。
- 施設名が時期によって、違うケースがある。また、看板と自治体で名称が違うケースもあることから、施設名は、自治体のサイトを活用し、調査してメタデータとして付与する。
- 仮設住宅には、通称が存在し、WEBで調査したものと自治体サイトで違いがある。自治体サイトを活用し、メタデータとして付与する。

■ デジタル化の例

デジタル化は、図書館のコンテンツについて実施。複製権がある為、許諾の取れたコンテンツより、デジタル化を実施している。



デジタル化の体制

- 専門スキルを持ったスタッフ2名で実施。

デジタル化の手法

- 当初、解像度400dpiでTIFF形式でデジタル化を実施方針であったが、総務省監修の「デジタルアーカイブの構築・連携のためのガイドライン」を参考に実際に、試験的に実施をし、容量を考慮し、解像度を300dpiとしている。保存形式については、刊行物の枚数が多いものを1ファイルにまとめる為に、pdf形式に変更している。一方、刊行物の枚数が多いものをデジタル化した場合の容量については、課題として再度調査を行い、最終決定する。

■メタデータ付与の例

東北大学



メタデータ付与の体制

- 研究員やアルバイトを最大7名体制で作業実施。専門スキルは必要なく、事前のツールの使い方やメタデータを付与するルールをガイドすることで作業可能。
- 作業を支援する担当を1名準備し、Q & Aを受け付ける支援を担当者が必要。
- 特に写真は、人の主観に左右される為、不明点は、メンバーで協議し、ルールを決める等の対応を実施。

タイトル:みちのくいまをつたえ隊写真(投入時に付与予定)
 提供者:東北大学(投入時に付与予定)
 説明:みちのくいまをつたえ隊写真(投入時に付与予定)
 区分:被害、復旧・復興
 区分(Keyword):道路、車、船、精米所、樹木、枯草、建物
 日時:2012/2/29
 種別:写真
 住所(県):宮城県
 住所(市区町村):石巻

メタデータ付与の手法

- メタデータを付与するツールとして、GooglePicasaを使用。
- メタデータ付与においては、見た内容から判断できる内容を付与する方針。人の思い込みを排除し、10万件のコンテンツにメタデータを付与する作業効率を考え、詳細な調査は実施していない。一方、写真のケースでは、写真に写っている文字は、正確にKeywordとして拾う対応はしている。
- 写真のケースは、第三者がタイトル、地域を判断することが説明を振ることは難しい為、付与対象外とし、データ投入する時に、タイトルは、何等かを付与する予定。
- 約50件/時間の生産性である。

河北新報社



メタデータ付与の体制

- 社員OB1名(週3日)+アルバイト4名(常時2名)の体制で実施。専門スキルは必要なく、ツールやメタデータを付与するルールをガイドすることで作業可能。
- 作業を支援する担当を1名準備し、Q&Aを受け付ける支援を担当者が必要。
- 特に写真は、人の主観に左右される為、不明点は、メンバーで協議し、ルールを決める等の対応を実施。

タイトル:河北新報社撮影(投入時に付与予定)
 提供者:河北新報社(投入時に付与予定)
 説明:河北新報社撮影写真(投入時に付与予定)
 区分:記憶
 区分(Keyword):農地、車、流木、津波被害、橋、晴れ、損壊、建物
 冠水、倒壊、がれき
 日時:2011/4/5
 種別:写真
 住所(県):宮城県
 住所(市区町村):仙台市若林区種次

メタデータ付与の手法

- メタデータを付与するツールとして、GooglePicasaを使用。
- メタデータ付与においては、出来るだけ正確に付与する方針で対応。住所や設備名は、写真からインターネットを活用し、出来るだけ調査し、正確な住所、施設名を取得。
- 写真のケースは、第三者がタイトル、地域を判断することが説明を振ることは難しい為、付与対象外とし、データ投入する時に、タイトルは、何等かを付与する予定。
- 約15件/時間の生産性である。

■メタデータ付与の例

河北新報社

掲載日:2011年03月13日, 面名:G30YX0, 記事ID:K20110313G30YX0010

水と食料やっど



ボランティアが炊き出し
熱々みそ汁に長い列

仙台・南材小

震災発生から11日、被災地ではボランティアの炊き出し活動が続いている。仙台市若林区の南材木町小では、地区の商店主や父母らのボランティアが住民への炊き出しを行っている。13日朝は午前6時ごろ、十数人のボランティアがみそ汁の準備を始めた。豆腐と油揚げ、がんもどきが入った熱々のみそ汁をもらおうと、住民が長い列を作った。区が支給したバナナも配った。停電が続く同校周辺はコンビニエンスストアも営業していない。食事の調達と寒さ対策が住民の悩み。朝と夜に実施している炊き出しには毎回、1000～1500人が並んでいる。ボランティアの1人で洋装店経営鹿野成也さん(55)は「行政機関が懸命にやっているのだから、住民も自力で頑張るしかない。二次災害が起こらないよう、電気が一刻も早く復旧してほしい」と話していた。(若林雅人)

仙台市若林区の南材木町小では震災発生からの11日夜から、地区の商店主や父母らのボランティアが住民への炊き出しを行っている。

13日朝は午前6時ごろ、十数人のボランティアがみそ汁の準備を始めた。豆腐と油揚げ、がんもどきが入った熱々のみそ汁をもらおうと、住民が長い列を作った。区が支給したバナナも配った。

停電が続く同校周辺はコンビニエンスストアも営業していない。食事の調達と寒さ対策が住民の悩み。朝と夜に実施している炊き出しには毎回、1000～1500人が並んでいる。

ボランティアの1人で洋装店経営鹿野成也さん(55)は「行政機関が懸命にやっているのだから、住民も自力で頑張るしかない。二次災害が起こらないよう、電気が一刻も早く復旧してほしい」と話していた。(若林雅人)

タイトル:号外/東日本大震災/水と食料やっど/ボランティア炊き出し/仙台・南材小

提供者:河北新報社(投入時に付与予定)

説明:河北新報社記事(投入時に付与予定)

区分:記憶

区分(Keyword):東日本大震災 復旧する 震災 ボランティア 電気

日時:2011/3/13

種別:記事

住所(県):宮城県

住所(市区町村):仙台市若林区

(参考)デジタル化の解像度と容量

図書館のデジタル化を行うに当たり、解像度による容量を試験的に取得してみた内容である。今後、これを元に解像度決定する予定である。

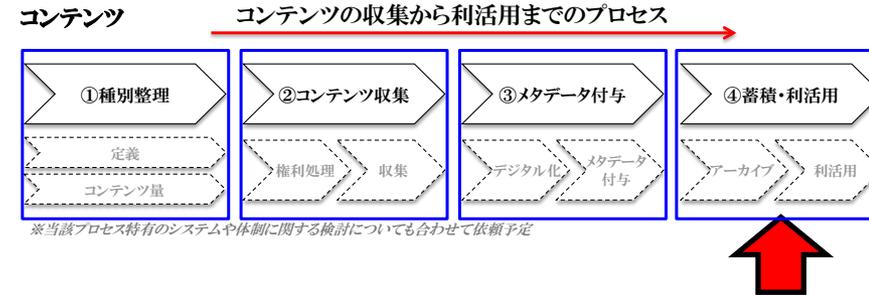
資料タイトル	枚数	用途	解像度	TIFF		PDF		JPG(参考)		イメージ図
				1枚	全体	1枚	全体	1枚	全体	
				KB	GB	KB	MB	KB	MB	
藻塩新商品試食会開催 (チラシ・パンフレット)	1	データ保管用	400dpi	27,058	-	2,419	-	-	-	
		プリンタ出力用(高品質)	300dpi	-	-	1,287	-	-	-	
		プリンタ出力用(一般用)	200dpi	-	-	-	-	-	-	
		WEB閲覧用(200%拡大)	144dpi	-	-	257	-	-	-	
		WEB閲覧用(拡大無し)	72dpi	-	-	135	-	-	-	
宮城県昭和震嘯誌 (古典資料)	747	データ保管用	400dpi	32,574	35.5	1,121	-	-	-	
		プリンタ出力用(高品質)	300dpi	-	-	672	483	583	402	
		プリンタ出力用(一般用)	200dpi	-	-	331	241	291	206	
		WEB閲覧用(200%拡大)	144dpi	-	-	192	143	168	123	
		WEB閲覧用(拡大無し)	72dpi	-	-	86	61.3	72	54.2	
駿府政治録 (古典資料)	101	データ保管用	400dpi	39,249	4.05	1,230	-	-	-	
		プリンタ出力用(高品質)	300dpi	-	-	699	92.1	563	80	
		プリンタ出力用(一般用)	200dpi	-	-	332	47.6	271	42	
		WEB閲覧用(200%拡大)	144dpi	-	-	192	28.9	158	25.7	
		WEB閲覧用(拡大無し)	72dpi	-	-	85	13.7	74	12.5	

(参考)メタデータ付与その他

メタデータ付与事例

- メタデータの単語の意味は、インターネット等を活用し、調べて引用する。(河北新報社)
- メタデータの単語については、一般の人がわかるような単語とし、協議し決定する。(河北新報社)
- 施設が時期によって違うケースや、看板と正式名称が違うケースがある為、自治体のサイトを活用し、正式名称を調べて付与する。(河北新報社)
- 写真に写されている、河川等の地図から見て判断できるものは、撮影場所と合わせて確認し、付与するようにする。(河北新報社)
- 写真に写っている文字は、メタデータとして、付与するようにする。(東北大学)
- 撮影日時が明らかに違うと判断されるケースは、撮影者へ問い合わせる。(河北新報社)
- 看板は、広告なので、そのまま公開する。(河北新報社)
- 仮設住宅等の施設には、通称が存在する。(河北新報社)
- 海辺の並木については、「防風林」、「防災林」、などの判定基準を定めることが難しい為、「松林」「並木」で統一(河北新報社)
- イベントで、プロレスラーが映ったチャリティイベントの写真があったが、後日事務所に確認する。(河北新報社)
- タグ付けメンバーに「被災者=津波被災者」のイメージがあり悩むケースあり。避難所のデータであった為、「避難者」「被災者」の両方を振ることで対応。(河北新報社)
- メタデータ付与時に、肖像権で公開対象として良いかを質問される。微妙なケースが多いが、都度、協議をしながら進める。(河北新報社・東北大学)
- 海辺の平地を見て、「津波」などをつけるケースがあるが、思いこみの可能性がある為、つけないようにする。(東北大)

1.4.蓄積・利活用


 東北大


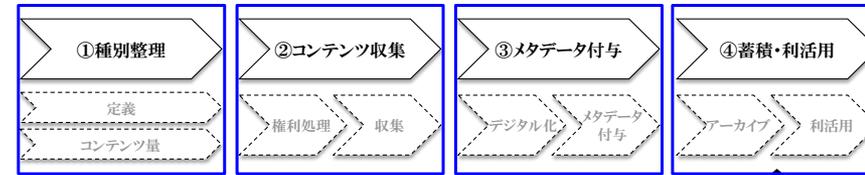
現状地域で想定される利活用

- 後年運用も被災地域を回り収集活動を行うと共に、研究論文等も収集し、保存活動を行う。
- 防災・減災の観点で、コンテンツを研究材料として役立てる。災害科学国際研究所を母体とする「みちのく震録伝」では、歴史的な災害から東日本大震災まで、様々な視点から集められた記憶、記録、事例、知見をもとに、分野横断的な研究を展開し、東日本大震災の実態の解明や復興に資する知見の提供を行っている。収集したコンテンツは、研究題材として、今後発生が懸念される東海・東南海・南海地震への対策に活用する。
- 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を目指す活動を行う。

河北新報社

コンテンツ

コンテンツの収集から利活用までのプロセス

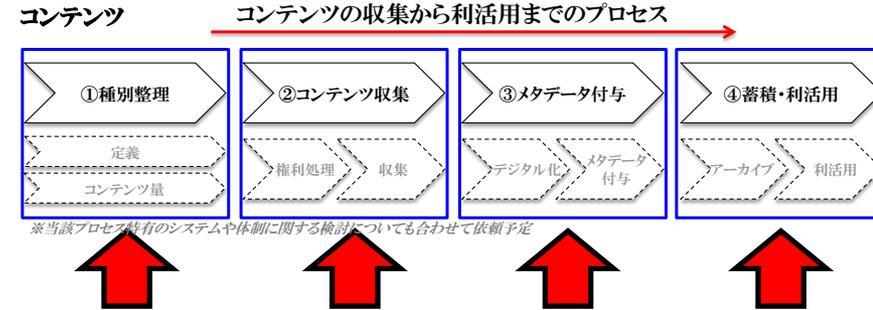


※当該プロセス特有のシステムや体制に関する検討についても合わせて依頼予定

現状地域で想定される利活用

- 地元報道機関として、震災を取材した新聞記事(見出し)を提供し続け、広く公開をし続ける。
- 震災の記憶、記録を後世に残す活動として、継続的にコンテンツ収集、保存活動を行う。東日本大震災は、デジカメ、スマートフォン、携帯電話が普及後、初の大震災であり、膨大な写真、映像が残っていると考える。これらを継続的に収集し、より多くの記憶、記録を保存する活動を行う。
- コンテンツを防災、減災の観点で教育に役立てる。河北新報社は、被災当事者でもあり、「次に備える」防災減災教育に寄与する活動を行う。
- 被災地域のための利活用を念頭に、地域再生の一助として、コミュニティ再生に役立てる。

1.5.その他



 その他特記事項

- 著作権者より通告を受けた場合に、速やかに削除する(Notice & Takedown)の機能及び運用は、重要である。著作権、肖像権をクリアしたコンテンツでも、人の心情、提供者が気づいていない要因、コンテンツから受けるイメージ等により、公開後に削除をお願いされるケースが予想される。

(参考) Notice & Takedownの必要性

当実証調査では、後年運用において、コンテンツを公開後に、著作権者や肖像権者等より通告を受けて、コンテンツを削除又は非公開とするケースが発生すると予測している。コンテンツ収集を行う過程で得られた内容より、発生するケースをまとめる。

【心情の変化】

コンテンツを提供頂き、著作権、肖像権もクリアしたコンテンツについて、写真に写りこんでいる人物の心情の変化により、のちに非公開とするケースが想定される。

(例1)

女兒が、テレビの取材クルーに追いかけて以来、自分が何処かで露出するのを嫌がるようになったケースがあった。

(例2)

3月11日が近づくと、メディアの取材が多くなり、住民の方がうんざりしている。

【写真撮影場所】

写真を提供頂いた中に、撮影場所に問題があるケースが発見された。今回のケースは、提供者が自己申告を行って頂いた為、発見されたが、提供された写真からは、判断つかない情報である。

(例1)

原発の立ち入り禁止区域に入り、撮影をおこなった。

(例2)

建物損壊による立ち入り禁止となっている施設に入り、撮影をおこなった。

(例3)

許可無く、私有地に入り込んで撮影をおこなった。

【写真撮影内容】

写真の公開に関しては、後年運用者が、リスクを持って対応を行うが、撮影内容については、著作権、肖像権の観点以外にも指摘を受けるケースが考えられる。

(例1)

被災した車置き場を撮影しようとしたら、管理者より撮影を拒否された。これらの車は、持ち主が現れない場合、競売にかけられる予定である為、競売後の持ち主に迷惑がかかる。

(例2)

企業の看板や建物が被災している写真について、イメージダウンにつながるとして、後にクレームが上がる。(宮城県沖地震時の実例)

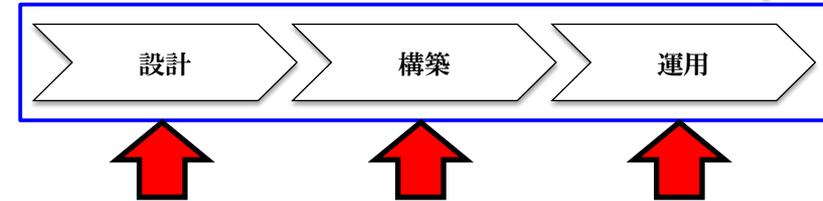
後年運用を行う際は、コンテンツ公開に関しては、リスクを持って行う必要があるが、最終的なリスクヘッジとして通告を受けたら削除(非公開)する機能及び問い合わせ窓口の設置を行くことが、重要である。



2.システム

システム

システム構築のプロセス



方針

- 他アーカイブ団体との交流を通してアーカイブの技術交流を行う。技術交流も視野に入れている為、オープンソースを活用し、開発を行う。
- 災害科学国際研究所保存及び東北大学附属図書館の公開データをより多く集約し、公開させる。
- 1991年からの新聞記事より震災関連コンテンツを抽出する為の分析基盤を準備する。
- 時間、空間、テーマ(Keyword)で構成されたシンプルで見やすい画面を作成する。

- 新聞記事見出しは、無償提供。新聞記事本文は、有償提供。
- 震災関連新聞記事及び新聞記事単位のKeywordの抽出に、東北大学のテキストマイニング環境を使用する。
- 既存商用システムに影響を与えない。
- 基本機能は東北大学と同じ機能を実装する。
- サーバーは、東北大学に設置し、ネットワークポリシー等は、東北大学に準拠する。
- 震災関連新聞記事見出しと写真等の関連性を表示する公開画面とする。

考慮点

- 10万件のコンテンツの保存に耐えるストレージを準備すると共に、拡張性のある構成とする。
- 開発期間内で作成できる、検索しやすい画面デザインを作成する。考慮点として、仕様書記載のレスポンスタイム(5秒)を満たすことが前提となる。
- テキストマイニングを活用できる可能性があれば、採用する。
- 区分は、Keywordの一部と考え、あえて区分としてメタデータスキーマに分けて定義しない。
- 震災メタデータスキーマは、将来的に拡張されることが予想される為、それを考慮した設計とする。

- ストレージは、年4万件ずつコンテンツが増加する見込みである。
- 震災関連新聞記事(見出し)の提供。後年運用は、震災関連新聞記事を日々抽出し、アーカイブシステムに連携する。
- 震災関連新聞記事を活用した公開画面を検討する。
- コストを抑える為に、機能は東北大学の機能を活用し、画面デザインは、河北新報社独自の画面構成とする。

2.システム

東北大

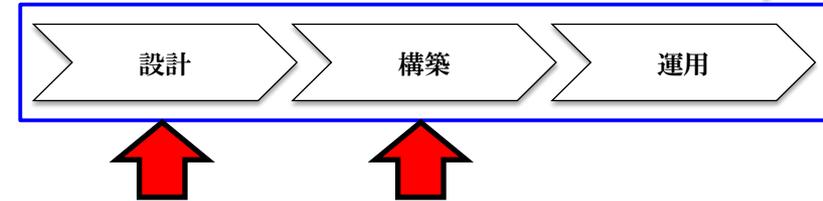
環境(インフラ)

- リソースは、全文検索エンジンのスペックとデータ量、利用者数によりCPU、メモリ、ディスクの使用量が決定されると想定し、動的にリソースを増やせる構成を取る。
- 河北新報社のシステム保守を念頭に、機器構成を減らすことで、保守に係るワークを減らすと共に、サーバーの仮想化技術を使用し、東北大と河北新報社のサーバを構築し、インフラ調達のコストを抑える。
- リソースの状況を確認し、リソース不足となった場合は、リソースの拡張を行う運用が行える環境とする。
- テキストマイニング環境として、震災関連新聞記事抽出及びKeyword抽出を行う環境を準備する。

- リソースは、WEBベースに必要となるリソースを見積もるが、最終的には、収集コンテンツの容量を確認し、決定する必要がある。
- 後年運用を見据えた環境への影響点整理が必要。インフラ環境は、東北大で運用を行う為、河北新報社との運用伝達のプロセスの確立が課題として上げられる。

システム

システム構築のプロセス



環境(ミドルウェア)

- 通常のWEBシステムがベースとなる開発であるが、違いは、全文検索等を行う検索エンジンの実装を行う点にある。検索エンジンには、solrを使用し、全文検索を行わせる。
- OSS製品を活用し、運用モデルにて拡張できるようにする。
- 開発言語は、WEBと親和性の高いJava言語を使用し、運用は、OSのシステムコマンド等を扱いやすい、bshellを使用する。
- 後年運用を考慮し、開発ツールは、東北大で開発しやすいツールを使用する。

- OSSを使用する際のライセンスを確認し、GPLライセンスは、使わない。GPLライセンスは、ライセンスの解釈で難しさがある。GPLライセンスのOSSを動的、静的に組み込んだソースは、全てGPLとして公開する必要がある。

現状の決定状況

今後の想定・課題

河北新報社

環境(インフラ)

- 河北新報社の記事データベースは、商用システムであり、現状運用されているシステムである。新聞記事見出しのアーカイブに際しては、新聞記事データベースシステムへ影響を与えない為に、疎結合とする。
- 新聞記事を震災アーカイブに連携するように、データ連携の口を作成する。
- 新聞記事のテキストマイニングを実現する為の連携の口を作成する。
- サーバーは、東北大に設置する。ネットワークや運用等のポリシーは、東北大に準拠する。

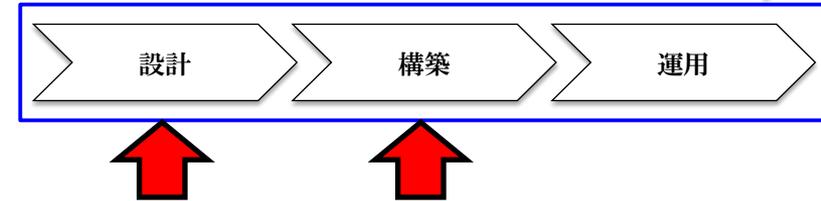
現状の決定状況

- 後年運用を見据えた環境への影響点整理が必要。インフラ環境は、東北大で運用を行う為、河北新報社との運用伝達のプロセスの確立が課題として上げられる。

今後の想定・課題

システム

システム構築のプロセス



環境(ミドルウェア)

- 基本的な機能は東北大で作成した機能を使用する。ミドルウェアに要件は無い為、東北大に準拠する。

- 現状特になし。

東北大

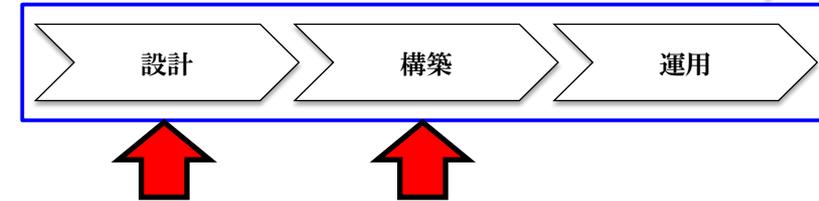
機能

- 仕様書記載の仕様を入れる。
- 東北大学の意見を取り入れた画面デザインとして、後年運用を見据えた画面デザインとする。
- 新聞記事の分析用に震災関連新聞記事の語彙を辞書として登録する。
- 一般からの投稿画面は、内容を公開に備えるかをチェックする運用を入れる。
- 一般からの投稿を考えた場合、投稿時にログインをさせて、メール等でコンテンツの問い合わせを行えるようにする必要がある。一方、一般投稿のユーザー管理は、登録するユーザー数が増えると、運用が大変であることから、GoogleやYahoo等のアカウントを使用する。
- 公開後のクレーム対応として、コンテンツを非公開とする機能を実装する。

- 投稿画面からの内容チェックは、後年運用の体制に反映させる。
- 動画は、種類が多く、WEBで公開を前提とした場合、ストリーミングに向かないものが多い。形式変更を前提として行うが、変更できないものがある為、変更できるものを公開対象とする。
- GoogleやYahooのユーザーを使う場合、システムのアカウントとしてGoogleやYahooへアカウント登録が必要であり、合わせて、サービスも登録する必要がある。

システム

システム構築のプロセス



データモデル(メタデータ)

- メタデータスキーマは、XMLベースであることと、拡張性を持たせる為、データベースへの保持は、XMLで保持する。
- Dublincoreをベースに、独自の定義を追加する。
- 第三者が、収集コンテンツに対して以下の項目を入力させることは、難しい為、必須としない。
 - ✓ タイトル、日付、説明、住所
- 震災メタデータは、将来的な拡張を考慮し、全てXMLで実装する。
- 震災メタデータスキーマをラップするデータアクセス部品を作成することで、開発工数を減らす。
- 区分は、Keywordの一部と考え、あえて区分としてメタデータスキーマに分けて定義しない。

- 区分の定義と付与したKeywordを見て、区分を再度整理し、確認する。

河北新報社

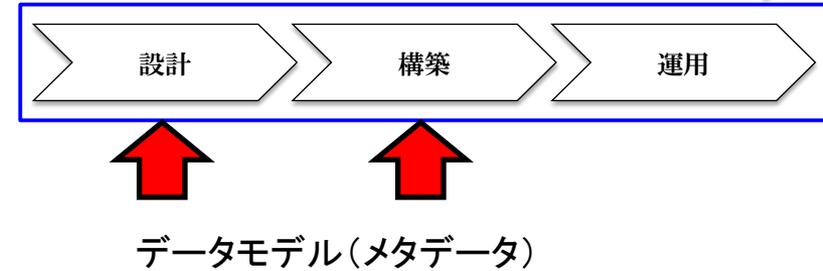
機能

- 仕様書記載の仕様を入れる。
- 河北新報社の意見を取り入れた画面デザインとして、後年運用を見据えた画面デザインとする。
- 収集コンテンツ(新聞記事見出し、写真)に合わせて、表示する画面を変更する。
- 基本的な機能は、期間、コストを考慮し、東北大と河北新報社で共有する。
- 震災関連新聞記事の本文より、Keywordを抽出する為にテキストマイニング環境へ登録する震災関連辞書(新聞記事の震災Keyword)を検討する。
- 一般からの投稿画面は、内容を公開に価するかをチェックする運用を入れる。
- 公開後のクレーム対応として、コンテンツを非公開とする機能を実装する。

- 投稿画面からの内容チェックは、後年運用の体制に反映させる。
- コンテンツには、位置情報が正確に入っていないケースが多々あることが確認できた。コンテンツにMAPを展開する場合、正確な位置にピンをさせない。たとえば、市区町村までしか住所が無い場合、市役所等にピンが刺さってしまう。

システム

システム構築のプロセス



データモデル(メタデータ)

- 東北大のデータモデルに準拠する。
- 新聞記事は商用である為、見出しを公開対象とし、記事本文はアーカイブ対象としない。

- 区分の定義と付与したKeywordを見て、区分を再度整理し、確認する

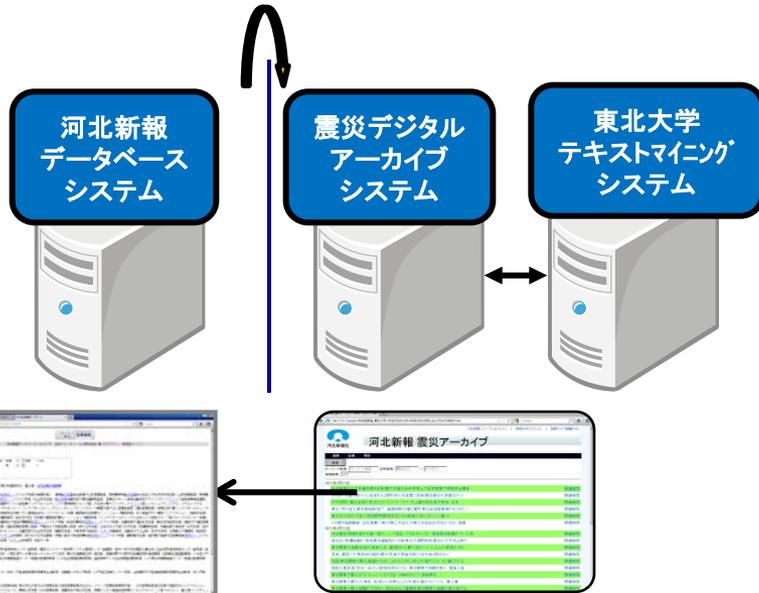
現状の決定状況

今後の想定・課題

(参考)河北新報社のICT環境整備

- 河北新報社では、新聞記事を保存、公開するシステムがあり、有料サービスとして行っている。今回、震災関連新聞記事をアーカイブし、公開するに当たり、公開の範囲、形態について検討を行っている。
 - 新聞記事の公開
 - ✓ 新聞記事は、商用である為、本文を含めて全てを無償提供は無理であるが、新聞記事見出しは、無償で公開可能とする。
 - ✓ 震災関連記事は、記事本文を人が見て判断するのが、適切である。現在の運用に新聞記事より震災関連記事を特定する運用を追加する。
 - ✓ 震災関連の新聞記事は、1991年よりデジタル化されている。過去記事本文を全て読まないといけない為、膨大な作業量となることから、東北大学のテキストマイニングを活用し、過去新聞記事より、震災関連記事を特定する。
 - 河北新報社社内システム
 - ✓ 河北新報社の新聞記事データベースは、商用システムであり、アーカイブの観点で用途が違うシステムであることから、商用システムに影響無いシステム構成とし、震災アーカイブに震災関連新聞記事を連携する手法とする。

震災記事連携



震災デジタルアーカイブより、クリックすることで、河北信報データベースより本文を読むことができる。(有償サービス)

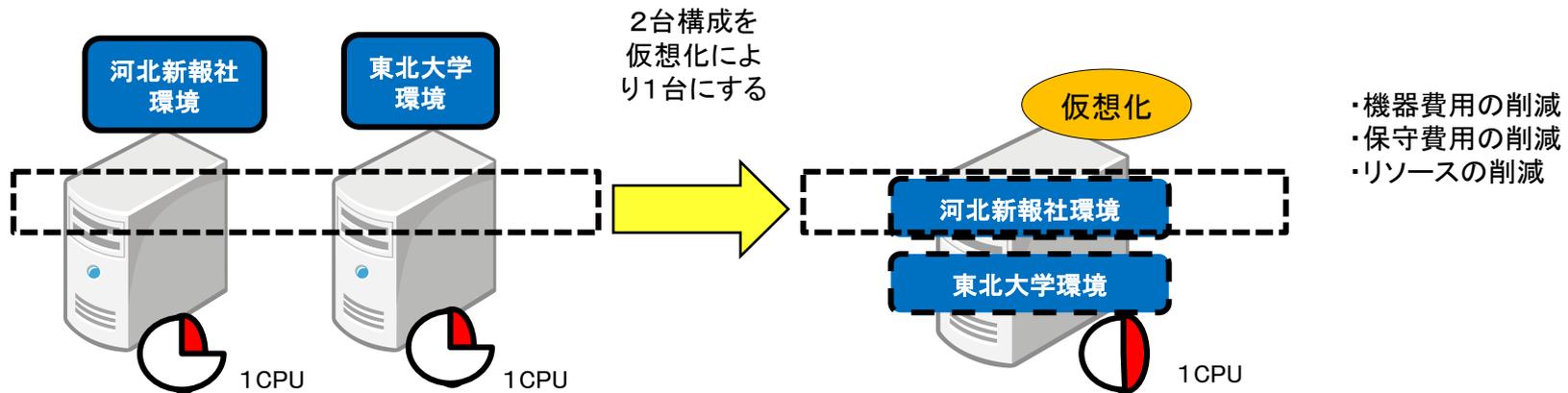


震災デジタルアーカイブより検索でき、無償公開

河北新報データベースより有償公開

(参考) 東北大学のICT環境整備

- 東北大学では、河北新報社のシステムの保守運用を行うと共に、東北大学及び河北新報社の後年運用における収集を想定したデータ容量やアーカイブシステムの使用率を考慮する運用を求められる。
 - 機器の仮想化技術の採用
 - ✓ 河北新報社、東北大学でそれぞれ機器を準備した場合、機器費用が2倍となることから、仮想化技術を使い機器費用をおさえる構成とする。
 - ✓ 後年運用を見据えて、仮想化技術を使い、リソース(CPU、メモリ、ディスク)を不足時に動的に割り当てられる構成とする。



・一般的に、CPUやメモリは、フルに使われている状態ではないケースが多い。仮想化によりまとめることで、リソースを効率的に使える(余剰がない)ことで、機器費用を抑えることができる。

(参考)区分とKeywordの実装の考え方

- 区分とKeywordの実装では、コンテンツ表現及び検索の為に使用する語彙として考え、メタデータスキーマの実装では、同じものとして実装する。
 - ✓ 区分: 一般的に分かりやすい語彙
 - ✓ Keyword: コンテンツに付随する、固有の情報
 - ✓ メタデータスキーマでは、subjectとして実装

区分
被害
法律
経済
行政
防災・減災
救援・救助
暮らし
社会福祉
教育
地震
津波
医療
土木・建築
産業
物流
交通・港湾
情報・通信
芸術・文化財
文芸
復旧・復興
原発



区分: 被害、復旧・復興
 区分 (Keyword): 道路、車、船、精米所、樹木、枯草、建物



メタデータスキーマでの実装例

```
<dcterms:subject><rdf:Description rdf:about="http://shinrokuden.irides.tohoku.ac.jp/tags/code"><rdf:value>災害</rdf:value><rdf:value xml:lang="en">1</rdf:value></rdf:Description></dcterms:subject><dcterms:subject>
<rdf:Description rdf:about="http://shinrokuden.irides.tohoku.ac.jp/tags/code"><rdf:value>復旧・復興
</rdf:value><rdf:value xml:lang="en">1</rdf:value></rdf:Description></dcterms:subject>
<dcterms:subject><rdf:Description><rdf:value>道路</rdf:value></rdf:Description></dcterms:subject>
<dcterms:subject><rdf:Description><rdf:value>車</rdf:value></rdf:Description></dcterms:subject>
<dcterms:subject><rdf:Description><rdf:value>船</rdf:value></rdf:Description></dcterms:subject>
<dcterms:subject><rdf:Description><rdf:value>精米所</rdf:value></rdf:Description></dcterms:subject>
<dcterms:subject><rdf:Description><rdf:value>樹木</rdf:value></rdf:Description></dcterms:subject>
<dcterms:subject><rdf:Description><rdf:value>枯草</rdf:value></rdf:Description></dcterms:subject>
<dcterms:subject><rdf:Description><rdf:value>建物</rdf:value></rdf:Description></dcterms:subject>
```

3.体制・保守運用

体制

運用フローの中で体制を検討すべきポイント

東北大



収集

保守・運用

利活用

- 被災地域の収集担当として14名名のアルバイトを雇い、事務局スタッフとして2名が活動。収集担当者へ、事前に説明会を行うと共に、定期的に事務局スタッフによるヒアリングや電話対応によるサポートを実施。事前の専門スキルは必要ないが、地域住民とのコネクションを持っている収集担当は、活動しやすい傾向にある。
- 図書館保存のコンテンツについては、収集対象をリスト化するのに、職員1名(兼務可能)が対象を特定。事前の許諾確認として電話による意思確認を2名のアルバイトで実施。その後郵送にて許諾書を送付する担当として1名のアルバイト。電話による対応は、マニュアルを作成し、事前研修が必要。その他、事務局スタッフとして1名が対応し、必要に応じて図書館職員も対応。

- 東北大「みちのく震録伝」プロジェクトで保守・運用を実施。他のサーバーの保守・運用も実施していることから、合わせて研究員が保守・運用を行う。要員は、1名で実施予定。また、河北新報社のアーカイブシステムも機器及びシステムは、東北大で実施。
- コンテンツの問い合わせは、コンテンツ提供者ではなく、東北大が対応する。

- 体制として、東北大「みちのく震録伝」プロジェクト活動に組み込み、利活用を行う。みちのく震録伝では、防災・減災の研究を継続して行うプロジェクトであり、その一助として活用する。
- みちのく震録伝の研究成果は、公開しており、その一助として活用する。

- 収集に伴う要員は、専門スキルがないアルバイトで対応可能であるが、作業を支援するスタッフの設置が必要。目的、許諾といった問い合わせは、専門スキルを持つスタッフが対応をする体制が必要。
- 他のアーカイブ団体も同じような活動を行っており、収集先から見ると、何度も同じような許諾のお願いをされると受け取られる。これについても、専門スタッフが対応を行い、説明を実施する。

- 要員/体制を含め、みちのく震録伝の活動費より捻出される為、特に問題はない。

- 要員/体制を含め、みちのく震録伝の活動費より捻出される為、特に問題はない。
- 研究者の専門用語をKeywordとして充実させることで研究者が使いやすくなる。専門用語は、分野により違いがあり、研究者が振るような語彙は、一般的には使わない語彙であることから、利活用の課題として、後年運用で整理を行う。

現状の決定状況

今後の想定・課題

体制

運用フローの中で体制を検討すべきポイント



河北新報社

収集

保守・運用

利活用

- 収集は、社員が業務と兼務し5名体制で実施。河北新報社のCSR活動より得られたコネクションが大きく寄与している。趣旨を記載した趣意書を準備すると共に、趣旨について説明できるような教育が必要である。

- サーバーの保守・運用は、東北大にて担当し、コンテンツの保守・運用は、河北新報社で実施する。
- コンテンツの問い合わせは、コンテンツ提供者ではなく、河北新報社が対応する。

- CSR活動の一環として、対応を実施する。地元紙として、継続したCSR活動の体制の中で、コミュニケーションを取りながら、利活用を検討する。

- 後年運用も収集を行うに当たり、社員が継続して収集活動を行う体制を検討する必要がある。通常業務との兼務となるが、CSR活動と合わせて検討を実施する。

- 後年運用の体制について検討が必要である。コンテンツの保守・運用は、収集したコンテンツの登録以外に、通告を受けたら、削除する運用が考えられる。頻度としては多くないと考え、社員の兼務が可能である。
- サーバーの保守・運用に関して、東北大と協議が必要であり、どのような形で運用負担を行うかを決定する必要がある。

- 震災アーカイブは、広く行われている事業では無い為、技術者が限られている。後年運用におけるシステムの改修時は、東北大との共同研究の枠組みの中で知見を借りながら、機能改修を行う予定である。

現状の決定状況

今後の想定・課題

資料 4-4

福島プロジェクトの中間とりまとめ

2013.1.23

1.0.福島プロジェクト全体概要

福島の地域的特徴としては、地震、津波によって被害を受けただけでなく、原発被害が発生して、相双地方などが避難指示区域等になったことがあげられる。このような多重(地震、津波、原子力、風評等)被害によって、役場移転等が発生し、情報が散逸することもあった。そこで、相双地方を中心として、被災地域が求めるアーカイブ機能を実現するとともに、散逸した記録、資料等の収集・発信について実証する。

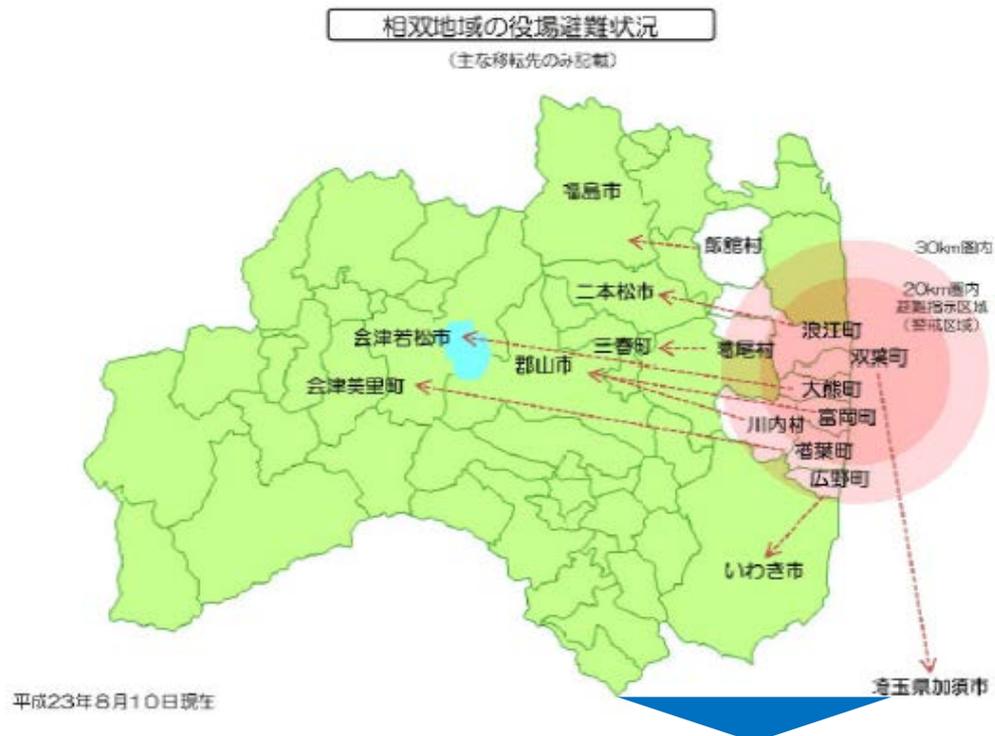


1.0.福島プロジェクト全体概要

(主な特徴)

福島県沿岸部に所在する相双地方及びいわき地方は、地震・津波のみならず、原発被害や、原発被害に伴う風評被害を受けた。さらに、緊急避難区域指定により、役場も移転せざるをえず、住民も移転や離散を余儀なくされた。このため、未だ生活復旧の段階である地域も多い。また、福島県沿岸部では、被災直後は避難所は転々と移動し、役場も今年度に入っても避難区域指定解除によって再移転するなど、震災以降、被災住民のみならず自治体を取り巻く環境もめまぐるしく変化を続けている。これらを踏まえて、福島プロジェクトでは、相双地方、いわき地方を中心に取組むものとする。

東日本大震災に関する災害の記録や伝承等の収集・保存は、人手不足により被災地域の尽力だけでは、ままならない状況であり、関連する資料・情報が毀損したり、散逸したりしつつある。また、①地震・津波被害、原発被害・風評被害といった多重被害を受けたこと、②被災後に自治体や住民を取り巻く環境の変化が継続的に発生したこと、③風評被害など二次的被害を受け続けていることの三点を背景とした、地域の取組や記録は、東日本大震災アーカイブにおける他県と異なる特徴である。



福島県(特に沿岸部)は原発被害によって、緊急避難区域の拡大、解除など、他の被災県よりも、状況が著しく変化した。これによって、行政文書や、地域メディアなどによる震災関連資料は、何度も修正されたり、上書きされたりしたことが、聞き取りにより明らかとなった。

また、現在も関連情報の管理・提供は、人手不足により必ずしも安定していない。さらには、関連情報の管理が十分でないケースもある。

そこで、このような状況を改善するために、本実証事業では、現時点で残存する行政文書やメディア保有情報等を中心に、整理を開始した。

(参考)アーカイブ格納コンテンツに関する仮説

福島県において収集すべき情報

広域での継続的な避難に関する情報

福島県では他県と同様に震災被害や津波被害を受けたが、これに加えて、原発被害を受けた。原発被害は広範に及ぶだけでなく、被災直後に緊急の避難を要することとなり、その後も長期間の避難生活を余儀なくされる。

この広域での継続的な避難に関する情報は、自治体が震災直後から住民向けに発信した情報を収集することにより、体系的に収集可能である。

風評被害に関する情報

福島県では原発被害と同時に、風評被害を受けた。このため、自治体自らが放射線量検査や広報活動を行った点に特徴がある。これらの風評被害及び風評被害対策に関する情報は、自治体のプレスリリース等情報を収集することにより、体系的に収集可能である。

福島県における情報の散逸状況

福島県では、津波被害を受けた沿岸部の自治体や、原発周辺の自治体(特に小規模の自治体)において、震災直後から数ヶ月間の情報が散逸している可能性が高い

安否確認、避難指示、放射線量情報、避難所情報等、自治体から住民へ緊急配信した情報は多量であり、さらには震災直後から十分な体制が整えられないまま、緊急に対応した可能性も高いため、包括的に情報の保存までを行う余裕がなかったと思われる。

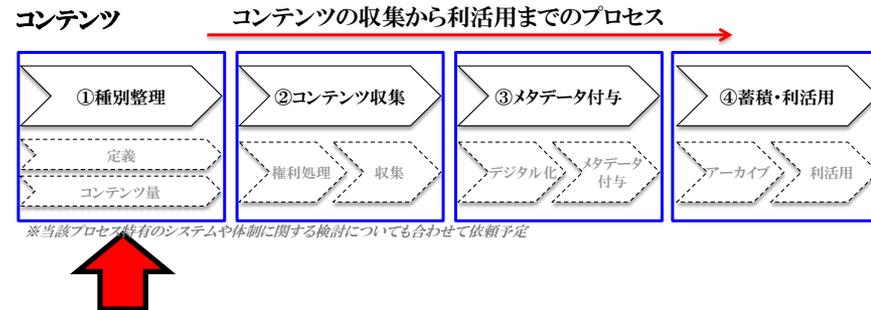
したがって、上記情報は保存されずに散逸しているものも少なくないと考えられる。

庁舎移転や全住民避難を実施している自治体は、様々な情報が散逸している可能性が高い

福島県においては、自治体の庁舎が津波や震災によって使用できなくなった例はないため、直接被害により震災資料が散逸することはなかったようであるが、原発事故の影響を受け、警戒区域等に指定され、立ち入りできなくなった等の理由により、庁舎移転や全住民避難を余儀なくされたケースがあった。したがって、文書の保管場所やサーバー・ディスクスペースの確保の問題から、震災以降、日々更新された情報が保管されずに散逸している可能性がある。

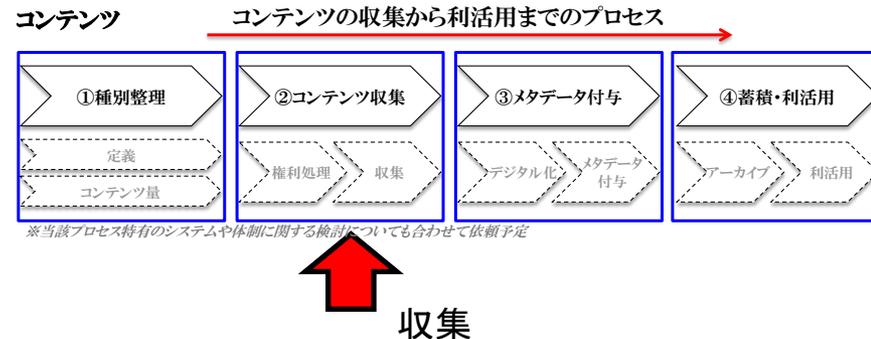
そこで、自治体などを対象に、市民向け情報やプレスリリースを中心にコンテンツを収集する。

1.1.コンテンツ収集動向



No.	内容	想定種別	想定点数	実績種別	実績点数
1	地方自治体によるプレスリリース	文書	3,500点	PDF,Word,Excel等	約3,200点
2	地方自治体による広報誌、回覧文書	写真	約1,000点	PDF,Word,Excel	約14,000点
3	地方自治体等のホームページなど住民向けコンテンツ	テキスト	約15,000点	htm,html	約10,000点
4	地方自治体等所蔵の画像データ	写真	約1,000点	JPEG	約5,000点
5	災害情報(国土交通省等)	文書	約10,000点(参考)	PDF,Word,Excel	収集交渉中
6	メディア・インフラ企業等の災害の記録類	文書	約1,000点(参考)	PDF,Word,Excel	収集交渉中

1.2.コンテンツ収集における方法・課題



権利処理

- 散逸・滅失する情報の保存を優先するため、原則として「デジタル化された公開情報」の収集を行っている。現時点までに収集した情報としては、
ホームページ情報
報道機関・住民向けのプレスリリース
広報誌
住民回覧情報
等があり、権利処理にあたってセンシティブな対応を要するコンテンツは比較的少ない。
- コンテンツの使用許諾等については、依頼状で趣旨を共有し、相対の協議で承諾を得た上で、別途、文書を取り交わすこととしている。

【体制】

- 権利処理等について協議する。アーカイブ、システムともに一定の知見を有するものが担当。学識経験者を交えて、コンテンツの収集を実施。

【管理】

- コンテンツ収集先に対し、通常業務に差支えない範囲での協力を依頼。受領したデータについてはセキュリティ・メディアにて保管。

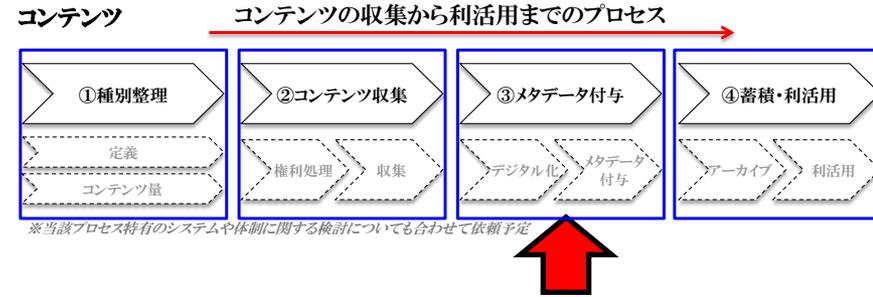
【収集方法】

- 収集先を訪問し、事業の概要の説明等を実施の上、総務省からの事業協力依頼文の発送を受けて再度訪問。記録メディアにてデータを受領。CMSにてホームページを構築している場合、データをダウンロードして受領することもある。

- ①自治体や企業が発行している「震災の記録」等の刊行物において、新聞の画像を引用しているケース、②市町村のホームページで「写真館」として住民から集めた画像を掲載しているようなケース、③市町村職員が撮影した写真に市民が映り込んでいるようなケースなどにおいては、コンテンツの提供者と著作権、肖像権等を有する者が異なる。
- 上記の場合において、対応方針を統一する必要がある。
- 例えば、一部の自治体からは職員が撮影した写真を提供され、写っている人物に関しては、問題があれば本人に許諾を取る旨の回答をいただいているが、処理には相当の時間を要することもありえるため、個人を特定するような情報の場合には、当初は非公開とする可能性もある。

- 福島プロジェクトでは自治体から収集しているのは、公開情報が多いが、写真等の画像データもあり、個別の許諾を取得するには、自治体や住民の現状に鑑みれば、ある程度の時間を要せざるをえない。この場合は、公開までに一定の時間を要することもある。
- 音声データや動画データについては、個人が特定されている情報が含まれるかどうかについて、すべてのデータを把握する必要があるが、自治体や機関がこれらの把握を迅速に実施できない可能性があり、収集したコンテンツの詳細な全情報を記録するには、一定の時間が必要となろう。
- 福島県内の多くの自治体において2011年3月11日以降にホームページのリニューアルが行われており、リニューアル以前の情報については既に参照できなくなっているものもある。また、コンテンツの上書き更新により古い情報が残っていない場合も散見される。

1.3.メタデータ付与における方法・課題



デジタル化

- 福島プロジェクトにおいては、住民向けに公開された情報を収集の中心としているため、データの残っていない広報誌や文書類の一部についてデジタル化が発生する。
- 文書のデジタル化にあたっては、総務省監修の「デジタルアーカイブの構築・連携のためのガイドライン」にて参照している「国立国会図書館資料デジタル化の手引き2011年版」の①保存用画像の項において「原資料に対して300～400dpi」との記載がある。この記述内容を踏まえ、データのスキニングの効率化を考慮した結果、白黒文書の解像度を300dpiとしてスキニングを行い、PDFファイルとして保存することとした。
- デジタル化作業については、一般刊行物以外については、文書の管理及び取扱いを簡素化するため、文書の保管場所において行い、文書の持ち出しは行わないこととしている。

- 収集したコンテンツには、インターネット放送局の番組データがあるが、この番組がハイビジョン番組で1番組あたりのサイズが25GB程度あり、アーカイブで閲覧する際のダウンロード時間等を考慮するとこれを分割する必要がある。場合によっては画質を落とす必要があるが、データの分割に当たっての1ファイル当りのサイズや再生時間の基準を協議する必要がある。
- 紙文書のデジタル化については、紙文書に付箋等がついていることが多く、ソーターによる大量高速でのスキャンが難しい。このため、トレイに1枚ずつセットし、スキャンし、結果を確認し、ファイル名を付与して保存するプロセスとなる。現実的には文書1枚当たり、1～2枚となり、相当程度の作業時間を要する。

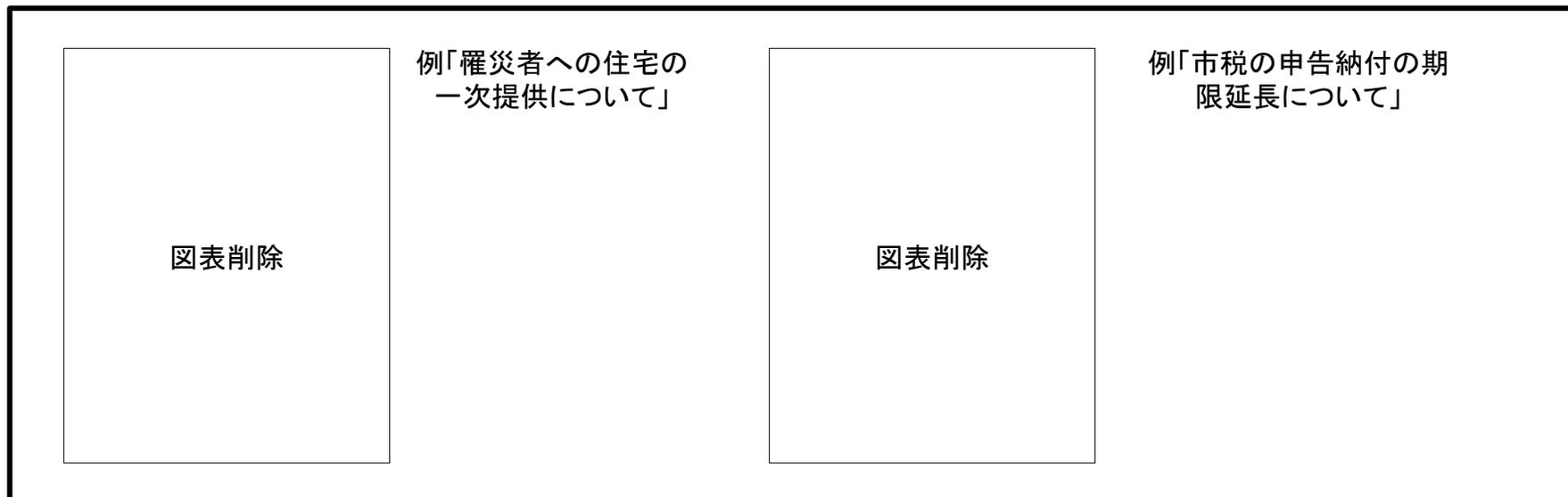
メタデータ付与

- 福島プロジェクトでの収集情報は文書データが比較的に多いため、メタデータの付与に当たっては、内容属性については文書の内容を判断しながら付与する必要がある。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する。
- メタデータ付与作業にあたっては、エクセルシートにコンテンツのファイル一覧を作成し、そこにメタデータの記入を行い、最後にアーカイブシステムへ一括で取り込む。
- メタデータ付与については、実験的に数千単位で実施した上で、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する。

- アーカイブ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、すべてのコンテンツに緯度・経度情報を付与する方針ではあったが、自治体の広報誌のような市内のさまざま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ない。
- また、特定の位置情報が含まれないケースもある。このような場合にどのスポットを緯度経度情報として登録するか、収集先との協議が必要。
- 単純に一つ一つのコンテンツに緯度経度情報を付与するだけでなく、コンテンツの内容を分析し、水道、電気等の復旧状況や支援物資の配送状況、避難者の移動等を個別に位置情報と紐づけて管理することで新たな有意義なコンテンツを形成することが可能となる。このような分析が可能となるようなコンテンツの格納方法も検討したい。

1.3.デジタル化の例

福島プロジェクトにおいては、自治体の紙文書の保存については緊急性が必ずしも高くないため、デジタルデータの保存を優先する。



デジタル化の体制

- 福島プロジェクトでは、方針に照らして、デジタルデータの収集を中心にしているため、今年度のデジタル化作業は多大ではない。デジタル化が必要な場合においては、コンテンツ収集担当者が作業を実施。

デジタル化の手法

- デジタル化に当たっては、文書については300dpiとし、写真については400dpiとした（国立国会図書館 資料デジタル化の手引き 2011年版）の①保存用画像の項において300dpi～400dpiと規定されており、文書については作業効率を優先して300dpiとし、写真については画像の鮮明度を優先して400dpiとした。）
- デジタル化に当たっては、検索時の内容の把握のしやすさを考慮し、文書1枚を1コンテンツとするのではなく、内容的にひとまとまりの文書を1コンテンツとした。

1.3.メタデータ付与の例

図表削除

提供者: 檜葉町
 提供者の分類: 行政・自治体・独法
 タイトル: 常磐線津波被害
 作成者: 檜葉町災害対策本部
 撮影日: 2011/3/15
 キーワード: 常磐線、津波
 分類1: 行政
 分類2: 記録写真
 分類3: 津波・震災被害
 資料種別: 電子資料
 サイズ: 720 × 480
 解像度: 96dpi

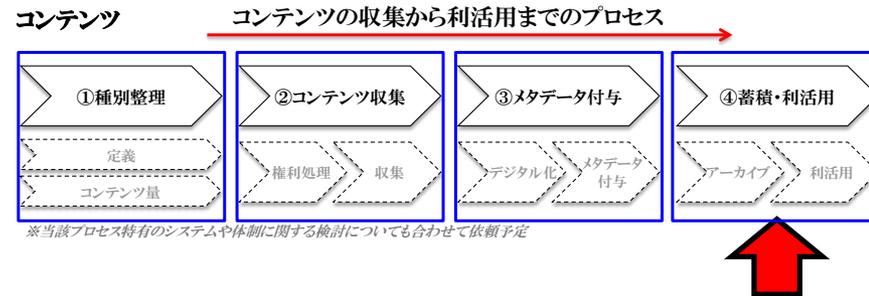
メタデータ付与の体制

- メタデータの属性値については、ポキャブラリが、ある一定数以上集まるまでは、記載内容の把握、分類、体系化に関するスキルが必要となるため、メタデータ付与作業を一定数以上行う必要がある。
- 当事業における収集対象コンテンツ数が10,000件であるため、20%～30%程度を少人数で実施する必要がある。一定のポキャブラリが確保でき、単純作業によってメタデータ付与が可能との判断ができた時点で分業体制を整える。

メタデータ付与の手法

- 国際的な標準語彙をできるだけ利用する方針とした。
- 標準語彙として、Dublin Core, FOAF, PREMISを採用。また国内のメタデータ語彙としてDCNDLを採用。
- 震災アーカイブの記述に適した国際的な標準語彙セットを探したが、適当なものが無い。参考となるメタデータスキーマがあるとメタデータの設計作業が効率的になる。
- 分類体系として他の運用モデルプロジェクトでの成果を参考にする予定。
- Webアーカイブに必要なメタデータ語彙に関して追加可能性あり。
- 定期的な収集による差分管理とそのメタデータ記述についても検討する。

1.4.蓄積・利活用

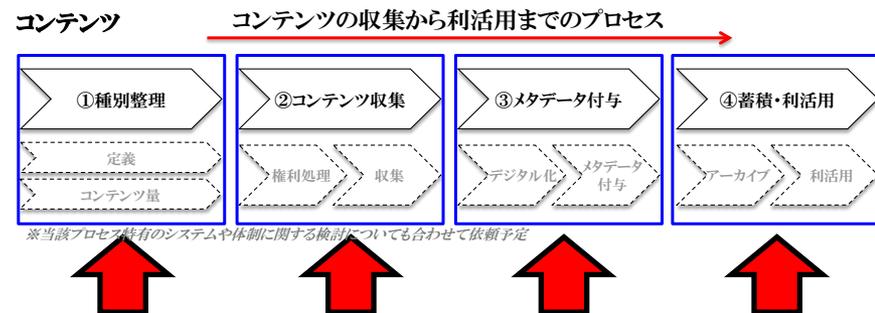


現状地域で想定される利活用

- 自治体そのものが避難する、もしくは、自治体住民が全避難を行うような災害はこれまでにないため、今後、自治体の災害対策計画策定の参考資料として利用。
- 災害発生後、国からの指示や指針がない中で、全市残留を決定した自治体、全町避難を決定した自治体、それぞれがあるが、これらの自治体の対応を整理し、検証することによって、避難マニュアル等の作成に活用。
- 福島県内における各市町村が抱える課題に関して、震災発生直後からの変遷を自治体横断的に把握することによって、広域災害の発生時や、広域災害の対策計画策定時に参考資料として利用。
- 現在避難中の自治体において、サーバー要領の確保ができずに古いデータを保存できない場合等に、データのバックアップ先として利用
- 各自治体ではホームページのリニューアルが行われているが、リニューアル前の情報を外部から閲覧可能な状態で保管しておくための倉庫として利用
- 県内の各自治体の放射線対策等の情報を掲載し、対外的に情報を配信し、風評被害対策として利用
- 学術機関での災害リスクマネジメントの研究資料として利用

※現在、コンテンツ収集先として広報等の担当課とコンタクトを取っているところ、利活用方法を考える上では、災害対策本部等へのヒアリングが必要となるので、利活用方法の検討は継続的に実施する必要がある。

1.5.その他



その他特記事項

- 福島県(特に沿岸部の自治体)に関しては、原発被害を受けており、個人が特定される可能性がある情報については特に慎重に扱う必要がある。
- 自治体が公表している放射線量の情報は、公表情報ではあるものの「ネガティブ」情報となり得る可能性がある。これらの情報のアーカイブにあたっては、その取扱いを適切に判断する必要がある。
- プレスリリース等には、企業からの寄付金の情報も掲載されているが、企業によっては企業名の公表を望まないケースがあるため、アーカイブ上は一旦すべて非公開として扱う。

2.システム

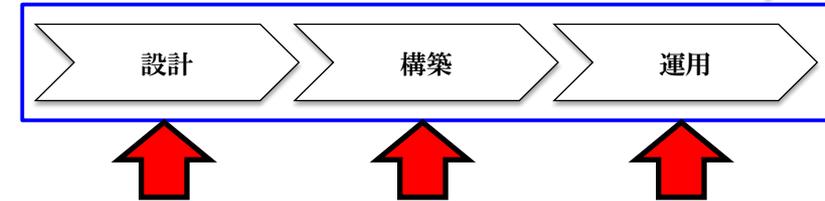
運用モデルの特性

方針

- 現存する資料を収集することを主眼に活動を行っているので、多様なデジタルコンテンツのWeb配信に対応できるシステム開発を行う。
- 収集したデジタルコンテンツは、長期保存に適したファイル形式に変換する。他方、公開という観点で、Web配信に適したファイル形式への変換も行う。(実証実験期間中は収集した元ファイルもオリジナルファイルとして保存する。)
- ウェブアーカイブ機能を実現する。これは自治体訪問時にサーバ容量の問題でデータ削除を行わざるを得なかったというヒヤリング結果に基づき、構築することとした。
- 遠隔地バックアップの観点から、奈良先端大学院大学へ実証環境を整備することとした。

システム

システム構築のプロセス



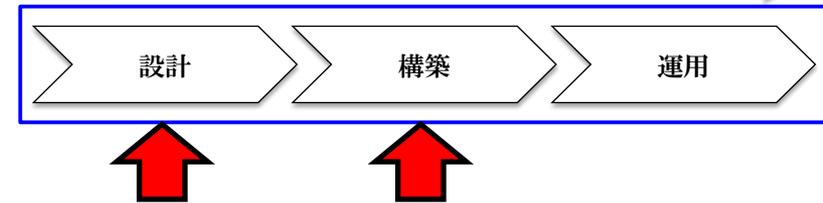
考慮点

- 多様なデジタルコンテンツ(文書、画像、音声、動画等)をスムーズにWeb配信できるシステム構築を行う。
- 1つのデジタルコンテンツに対して、配信用フォーマット(ただし配信用フォーマットは長期保存に適しているものとしている)、元ファイルフォーマットが存在することとなるので、その管理が出来る仕組みを構築する。
- ウェブアーカイブ機能を実装するにあたり、実証期間内に実現するため、先行事例のあるOSS(オープンソースソフトウェア)を採用した。
- データのスムーズなバックアップを実現するために、実績のある商用ソフトウェア(NetVault)を採用、導入した。

2.システム

システム

システム構築のプロセス



環境(ハード)

環境(ソフト)

- サーバ機器を他の運用モデルプロジェクトの設置場所から遠隔地に設置。(奈良県:奈良先端大学)
- システムの想定利用頻度(最大1日100人程度)、システムダウン時に要求される復旧時間等から大容量で安価なSATAのディスクをRAID5構成で採用しサーバ機器のコストを削減。
- 長期保存の観点からバックアップ機器は重要と考え、LTO4オートローダ(テープ・カートリッジ容量:9、物理容量:7.2TB)で多くの世代のバックアップを可能とする。
- サーバ機器のCPU、メモリは実証機関内に収集する1万件のメタデータの検索とコンテンツの配信を想定し小規模構成でも対応可能と判断(1CPU、MEM16GB)。サーバ機器の冗長化については、震災関連デジタルアーカイブに要求される稼働率は、業務系システムと比較して高くないと判断し、機器の冗長化は行わない。但し、サーバ機器内で低コストで可能な部品の冗長化については考慮した。(電源の二重化、HDDのRAID5構成、LANの二重化)
- 必要最低限の停電対策として1500VAのUPSを設置し、停電時にシャットダウンを行う設定とした。

- テープ媒体での遠隔地保存も想定する。そのためHDD等でのバックアップ運用は選定外とした。
- 後年運用のデータ増加を考慮し拡張可能(CPU、メモリ)な機器を選定。

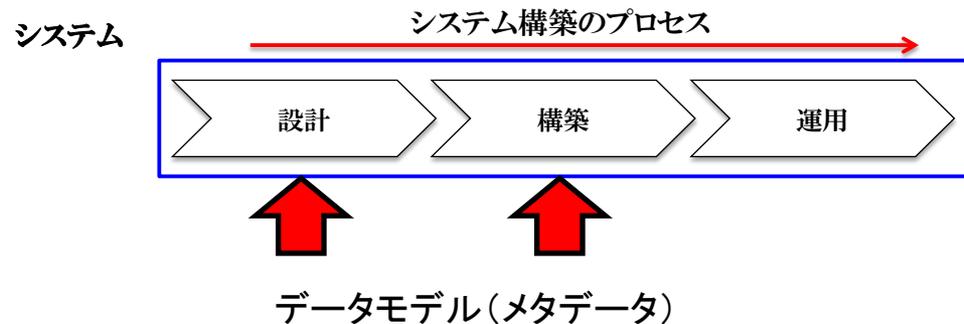
- ウェブアーカイブ用のソフトウェアとして、Internet Archive等で使用されているオープンソースソフトウェアWayback Machineを採用。
- 各種プロトコルの実装では、アーカイブ領域で多くの実績を有する商用ソフトウェアを選定。
- バックアップは重要であるため、多くの実績を有する商用ソフトウェアを採用。
- ウィルス対策ソフトウェアを導入
- 高速XML検索エンジンの採用により利用者向けサービスの高速化を図る。

- 今後のメタデータの項目追加等の拡張性を考慮し、XMLベースでのデータ保持を実施。
- ウェブアーカイブの機能によりコンテンツの自動収集の検討。

現状の決定状況

今後の想定・課題

2.システム



機能

- 国際的な標準語彙をできるだけ利用する方針とした。
- 標準語彙として、Dublin Core, FOAF, PREMISを採用。また国内のメタデータ語彙としてDCNDLを採用。
- 震災アーカイブの記述に適した国際的な標準語彙セットを探したが適当なものはない。参考となるメタデータスキーマがあると作業は楽になる。
- NDLとのメタデータ連携のために連携メタデータへの変換機能を実装。
- 連携メタデータはRDF/XML形式で表現する。
- 連携プロトコルとしてOAI-PMHを採用。

- 仕様書記載の機能を全て実装。
- 検索性の向上のために、検索キーワードのサジェスト機能、同義語辞書の登録機能を実装。
- マッシュアップ技術を活用した地図(Google Map)や年表(TimeLine)と連携した利用者サービスを提供。
- 震災関連コンテンツはWebに適したファイル形式に変換して配信する。合わせて配信形式を長期保存に適した形式を採用し震災関連コンテンツの長期保存の目的も兼ねる。元ファイルはオリジナルファイルとして保存する。
- 画像、動画、音声、文書等はそれぞれ配信形式を決定。内閣府「電子公文書等の管理・移管・保存・利用システムに関する調査報告書」の長期保存フォーマットを参考とし一般的なアーカイブに適用可能なフォーマットを採用。また、長期保存のためのメタデータを付与した。
- 幅広いデジタルコンテンツのフォーマットに対応するために、ファイル変換のためのフリーソフト(ImageMagic, wvText, docx2txt, ffmpeg, iconv等)を組み合わせて利用。
- Webページのアーカイブ機能を実装。日々更新されるWebページを履歴を含めてアーカイブする。

- 分類体系として他の運用モデルプロジェクトでの成果を参考とする予定。
- Webアーカイブに必要なメタデータ語彙の追加可能性あり。定期的な収集による差分管理とそのメタデータ記述について検討する。

- 今後のメタデータの項目追加等の拡張性を考慮しXMLベースでのデータ保持を実施。
- ウェブアーカイブにて収集したWebページはISO標準のWARC形式での保管を検討中。

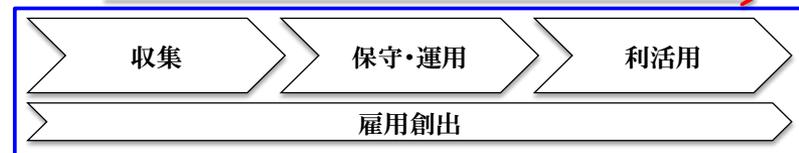
現状の決定状況

今後の想定・課題

3.体制・保守運用

体制

運用フローの中で体制を検討すべきポイント



収集

保守・運用

利活用

現状の決定状況

今後の想定・課題

- 今年度事業においては、「デジタル化された公開情報」を優先的に収集するため、紙文書についての収集作業はひとまず次年度以降に拡大することとする。
- コンテンツの収集に当たっては、実効性、効率性を考慮すると、ある程度専門知識を有する、地域の学識経験者等が、より主体的に関与することが望まれる。

- 今年度事業においては、散逸しつつあるデータの保存を最優先とし、「デジタル化された公開情報」を中心に収集した。ただし、文献として貴重な紙文書情報も自治体等に保管されているため、これを来年度以降に収集し、デジタル化する必要がある。
- 紙文書のデジタル化については1分当り1枚程度が限度であり、膨大な時間を要する。他方、地元の雇用創出には寄与可能である。

- 国立国会図書館関西分館と連携を有する奈良先端科学技術大学院大学のサーバールームにアーカイブシステムのサーバーをハウジングし、リモートアクセスによって保守・運用管理を行う。

- 来年度以降のコンテンツの収集に関しては、紙文書が多くなる。紙文書の収集にあたっては、既にシステムが整備されており、メタデータのボキャブラリもある程度整備され、メタデータ付与作業がマニュアル化されていれば、比較的安価にアーカイブ可能と考えられる。

- 【激甚災害を想定した災害対策】
- 災害対策マニュアルを作成する際に、複数自治体にまたがる広域避難ルートや支援物資運搬ルートを決定する資料として活用。
 - 原発周辺自治体での情報管理、情報発信、避難指示等のマニュアル作成に利用。
 - 自治体機能そのものが移転する規模の災害発生時のマニュアル作成に活用。
- 【学術研究】
- 学術機関での研究にあたって、一次資料として利用。

- 来年度以降、アーカイブの利活用促進のため、アーカイブそのものの広報活用やアーカイブを利活用した研究の事例紹介等を行う必要がある。このような活動を予算を措置した上で学術機関等において実施することも一案であろう。