

第7章 システムの構築・運用について

震災関連のコンテンツを公開し、利用者に提供できるようにするには、震災関連デジタルアーカイブシステムを構築する必要があります。また、システム稼働後には、システムの運用を行う必要があります。

第7章は、システムの構築・運用に必要な作業について説明するものです。「2. 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ています。

1. 震災関連デジタルアーカイブ構築・運用について

どのようなアーカイブを作りたいか、骨格が固まってきたら、必要なシステムの検討に着手しましょう。

この節では、「震災関連デジタルアーカイブ」におけるシステム構築・運用の進め方についてご紹介します。

(1) 震災関連デジタルアーカイブにおけるシステムの構築・運用の流れ

「震災関連デジタルアーカイブ」におけるシステム構築・運用の進め方には、以下の2つの方法があります。

- ▶ 自治体等運用主体となる者がシステムを構築・運用する場合
 - ▶ クラウドサービス等のサービスを利用する場合
- 以下にそれぞれの方法について、進め方を示します。

(a) 自治体等運用主体となる者がシステムを構築・運用する場合

震災関連デジタルアーカイブを構築・運営するためには、アーカイブシステムを構築しなければなりません。自治体等運用主体となるものが自らシステムを構築・運用する場合のイメージは、以下のとおりとなります。アーカイブを構築する際には、NDL 東日本大震災アーカイブとの連携を検討しましょう。NDL 東日本大震災アーカイブとの連携により、横断検索などアーカイブの利用者にとって利便性が高まり、より多くの利用者に利用してもらえるようになります。

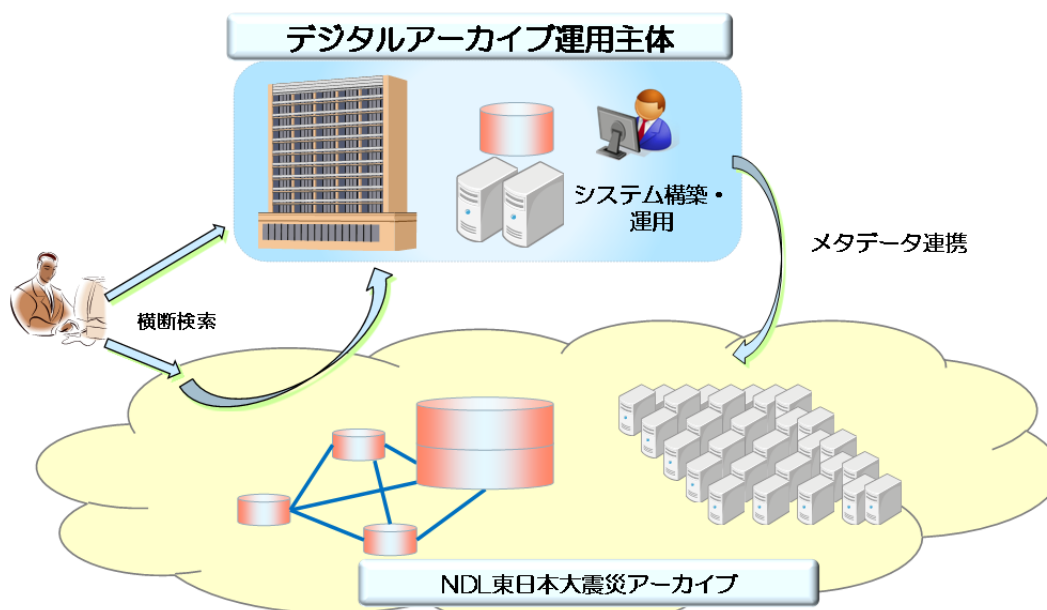


図 7-1 運用主体がシステム構築・運用する場合

自治体等運用主体となる者がシステムを構築し、運用する場合は、次のステップで実施します。



図 7-2 システム構築・運用フロー（運用主体がシステム構築・運用する場合）

システム構築・運用の各プロセスの詳細について、以下に示します。

① システム要件の検討

システム要件の検討では、震災関連デジタルアーカイブを実現するシステムに必要な要求要件を整理し、要求要件を満たすために実装しなければならない機能について検討する必要があります。要求要件の検討では、NDL 東日本大震災アーカイブとの連携の可否についても検討しましょう。

また、システムを実装するために必要となるハードウェア、ソフトウェアについて、検討を行う必要があります。システムベンダー数社に意見を聞くのも良いでしょう（これを、RFI（意見招請）と言います）。

詳細は、「1. (2) 震災関連デジタルアーカイブにおけるシステム要件の検討について」を参照ください。

② システム設計・システム構築

システム設計・システム構築では、要求要件に従い、実装する機能のプログラム開発、試験を実施します。実際にシステム設計、システム構築を実施する業者を選定（調達）するのはこの時期です。業者が決まったら、システム構築業者でハードウェア、ソフトウェア等を構築し、試験環境を準備します。システム連携では、NDL 東日本大震災アーカイブと連携する場合は連携テストを実施しましょう。

詳細は、「1. (3) 震災関連デジタルアーカイブにおけるシステム設計・構築について」を参照ください。

③ システム運用

システム運用では、構築した震災関連デジタルアーカイブシステムについて、運用計画を策定し、システム運用を行います。システム構築業者が実施する場合と、別途運用業者を調達する場合があります。システム運用では、NDL 東日本大震災アーカイブとの連携運用を行いましょう。

詳細は、「1. (5) 震災関連デジタルアーカイブにおけるシステム運用について」を参照ください。

(b) クラウドサービス等のサービスを利用する場合

自らシステムを構築し、運用を行うことが難しい場合や、コンテンツの量がそれほど多くない場合には、クラウドサービス提供事業者が提供するサービスを利用することでアーカイブを運営することができます。

アーカイブサービスを提供している運営主体には、NDL 東日本大震災アーカイブをはじめとして、311 まるごとアーカイブ等があります。クラウドサービス等を利用する場合のイメージは、以下のとおりです。

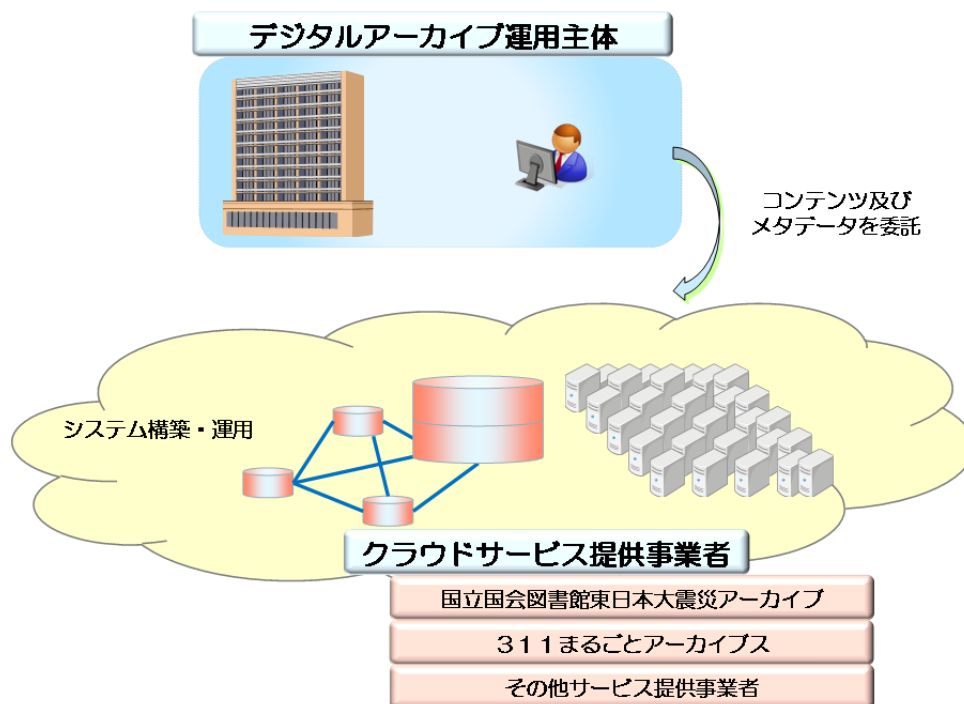


図 7-3 クラウドサービス等のサービスを利用する場合

クラウドサービス等のサービスを利用する場合は、次のステップで実施します。

運用主体者が実施

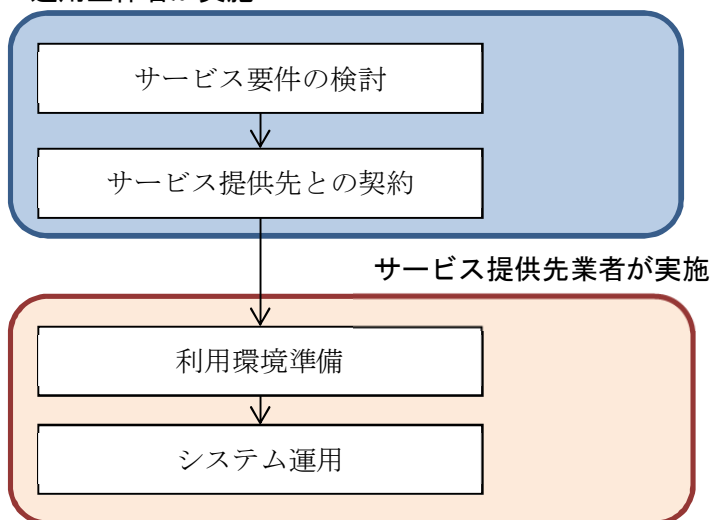


図 7-4 システム構築・運用フロー（サービスを利用する場合）

システム構築・運用の各プロセスの詳細について、以下に示します。

① サービス要件の検討

サービス要件の検討では、震災関連デジタルアーカイブで必要となるサービス要件について検討を行います。サービス提供先により、利用できるサービスの内容が異なりますので、どのようなサービス要件とするかは、デジタルアーカイブのクラウドサービス提供先業者数社に意見を聞くのも良いでしょう（これを、RFI（意見招請）と言います）。サービス提供先によって、様々なサービスの形態があります。

例えば、運営に必要なサービスを全て提供してもらうことにより、運用主体としては、コンテンツとメタデータをサービス提供先へ委託するだけで、震災デジタルアーカイブの運営を行うことが可能です。個性あるシステムを構築したい場合には、ハードウェアやネットワーク等の基本的なサービス提供だけを受け、システムは別途調達するということも可能です。それぞれの運用主体に適したサービスを選択しましょう。その際、後年運用も含めた必要経費等についても検討しましょう。

② サービス提供先との契約

サービス要件が決まったら、サービスを調達し、サービス提供先との契約を行います。

③ 利用環境準備

サービス要件に従い、サービス提供先が準備を行います。

④ システム運用

サービス要件に従い、サービス提供先がシステム運用を行います。

(2) 震災関連デジタルアーカイブにおけるシステム要件の検討について

「震災関連デジタルアーカイブ」は、震災に関係する各種デジタル形式の資料等を収集し、蓄積し、提供することを目的としています。

以下のようなことに留意し、システム要件の検討を行きましょう。

➤ 検索のしやすさ

アーカイブの利用者を想定し、どのような検索機能を用意すべきか検討しましょう。特定の人（研究者や自治体職員等）だけに公開するのか、一般の人に公開するのかわかり、検索キーワードの設定や地図やタイムラインなど検索のための補助機能を用意する必要があるか、など、検索に必要な機能が違ってきます。

➤ 権利関係に配慮したアクセス制御

震災関連のコンテンツ特有の課題として、著作権・肖像権・個人情報保護等、各種権利に配慮したシステム構築を行う必要があります。コンテンツごとに公開／非公開、限定公開、閲覧のみ許可／ダウンロード可等のメタデータを付与する必要があります。ま

た、利用者を限定する場合は、サイト自体へのアクセス制御機能も必要です。

➤ **運用のしやすさ**

この段階で、どんな体制でアーカイブを運用するのかを検討しておくことも重要です。例えば、アーカイブ公開後にコンテンツを増やしたい場合、コンテンツの投稿者が自分で登録できるようにするのか、運用者（自治体職員等）が登録するのかによって、登録のフローや必要な機能が異なります。投稿を許す場合は、登録できるコンテンツのファイル形式や容量のように機械的に判別できるものはなるべく自動化して運用しやすくするなど、システムで対応できることがあります。

➤ **長期保存を考慮**

ハードウェアの寿命やデジタル化技術の進展、デジタル化ツールや再現機器の変化を考慮すると、システムのライフサイクルを考慮した機器更改やデジタルデータのエミュレーション、マイグレーションの計画を立てておく必要もあります。そのために必要な予算についても計画しておきたいところです。

➤ **NDL 東日本大震災アーカイブ「ひなぎく」とのシステム連携の実現**

震災関連デジタルアーカイブは、国立国会図書館の運営する NDL 東日本大震災アーカイブ「ひなぎく」と連携し、各種コンテンツを横断的に検索可能とする場合は、「NDL 東日本大震災アーカイブ「ひなぎく」との連携について」をご覧ください。

(a) **システム機能要件**

「震災関連デジタルアーカイブ」では、最低限、以下に示す機能を実装する必要があります。なお、性能要件については、想定されるアーカイブへのアクセス数を検討し、必要な要件を検討します。

システム機能は、以下のように分類されます。

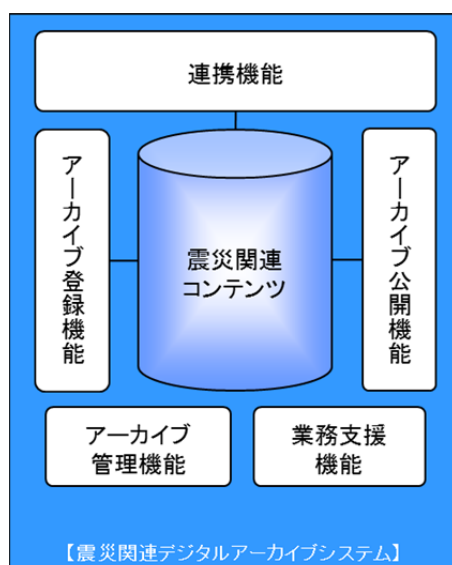


図 7-5 アーカイブシステムの機能分類

運用実証調査では、以下の機能要件をアーカイブに実装しています。

表 7-1 システム機能要件

No.	機能分類	必要となるシステム機能要件
1	基本要件	<ul style="list-style-type: none"> 標準的な Web ブラウザで利用可能なシステムとすること。 文字エンコードとして UTF-8 を使用すること。 国際標準及び業界標準の技術を積極的に用いて構築すること。
2	性能要件	<ul style="list-style-type: none"> 目安として、以下のアクセス頻度及び最大同時アクセス数において、検索実行から応答に要する時間は 5 秒以内とすること。なお、この場合の応答時間には、ネットワーク負荷等の外的要素を含まない。 <ul style="list-style-type: none"> ➤ 1 日平均 10,000 件以上 ➤ 1 時間の平均アクセス頻度 1000 件程度 ➤ 最大同時アクセス数 10 同時アクセス程度
3	コンテンツ公開機能	<ul style="list-style-type: none"> 蓄積された震災関連コンテンツについてインターネットを通じて Web ブラウザを使用し閲覧、利用できること。 日本語版と英語版のインタフェース画面を有すること。日本語、英語以外の言語についても言語設定の追加により対応できる構成とすること。 検索機能としてキーワード検索、項目検索、分類による検索、時間や場所での検索を有すること。

No.	機能分類	必要となるシステム機能要件
		<ul style="list-style-type: none"> ・ 利用者の環境に依存せず幅広く利用可能なシステムを構築すること。 <ul style="list-style-type: none"> ➢ Windows、Mac OS 等 ➢ Internet Explorer、Firefox、Chrome、Safari 等 ・ 検索した震災関連コンテンツはメタデータ文字情報での表示形式に加え、利用者が直感的に利用しやすい形での提供を行うこと。 ・ メタデータでは、標準的な語彙が存在する場合はその語彙を利用することを前提とすること。 ・ 利用者が効果的に震災関連コンテンツを利用するための機能についての検討も行うこと。 ・ 利用者の権限に応じた公開権限を設定できる機能を有すること。権限設定はメタデータ単位、震災関連コンテンツ単位で設定できるものとし、権限の種類については下記を実現するものとする。 <ul style="list-style-type: none"> ➢ 一般公開 ➢ 内部公開 ➢ 管理者公開 ➢ 特定の組織、コミュニティへの公開（学術的・防災・行政等）
4	コンテンツ登録機能	<ul style="list-style-type: none"> ・ 震災関連コンテンツを登録する機能を有すること。 ・ コンテンツ登録は、アーカイブ運用担当者、運用管理者、アーカイブの一般利用者等の利用者種別により実施可能とすること。 ・ 利用者種別により、変更・削除ができる等、権限に応じた機能を提供すること。 ・ 震災関連コンテンツの 1 件単位での個別登録と一括登録の機能を有すること。 ・ 利用者の権限に応じたデータのアクセス管理ができること。 ・ 画像、音声、動画、文書等の幅広い震災関連コンテンツに対応した登録機能を有すること。
5	コンテンツ管理機能	<ul style="list-style-type: none"> ・ 登録された震災関連コンテンツを蓄積、管理する管理機能を有すること。 ・ 震災関連コンテンツを長期的に保存、活用するための機能を有すること。

No.	機能分類	必要となるシステム機能要件
		<ul style="list-style-type: none"> ・ 震災関連コンテンツのバックアップや遠隔地での保存等についてデータの保全性を考慮すること。 ・ 蓄積された震災関連コンテンツの集計機能を有し、震災関連コンテンツの登録状況、種類、データ提供元等の集計情報が確認できること。 ・ 蓄積された全データ及び一部のデータ（メタデータ及びコンテンツ）の一括出力ができること。
6	外部連携機能	<ul style="list-style-type: none"> ・ 外部機関のアーカイブとの連携のため下記全てのプロトコルを実装すること。さらに幅広い連携のため、他のプロトコルの実証についても検討すること。 <ul style="list-style-type: none"> ➤ 検索：Z39.50、SRU、OpenSearch ➤ データ連携：OAI-PMH ・ NDL 東日本大震災アーカイブとの連携機能を有すること。
7	業務支援機能	<ul style="list-style-type: none"> ・ システムの運用を支援する以下の業務支援機能を有すること。 <ul style="list-style-type: none"> ➤ 利用統計機能 ➤ セキュリティ機能、ウイルス対策機能 ➤ 利用者管理機能 ➤ システム監視機能 ➤ サーバ監視機能 ➤ 検索ログ保存機能⁶⁰

(b) ハードウェア・ソフトウェア要件

震災関連デジタルアーカイブのシステム機能要件検討後、又は、システム設計・構築業者が決定し、システム設計完了後、必要となるハードウェア・ソフトウェアについて、以下の事項を検討します。

① サーバ等の設置場所の検討

震災関連デジタルアーカイブのハードウェアをどこに設置するのか検討します。

設置場所としては、以下のようなケースが想定できます。

- ・ システム運用事業者の運用管理センター等に委託し、設置する。

⁶⁰ 震災関連デジタルアーカイブの利用者がコンテンツ検索の際の検索条件にどのようなキーワード等を指定したのかをログとして保存する機能。ログを集計・分析することにより、検索しやすいメタデータ付与方法を検討できる。

- ・ 庁内等のサーバールームに設置する。
- ・ クラウドサービス、ホスティングサービス等を利用する。

② 設置するサーバ等の要件検討

震災関連デジタルアーカイブのハードウェア、ソフトウェア等の要件について検討します。

<システム構成の例>

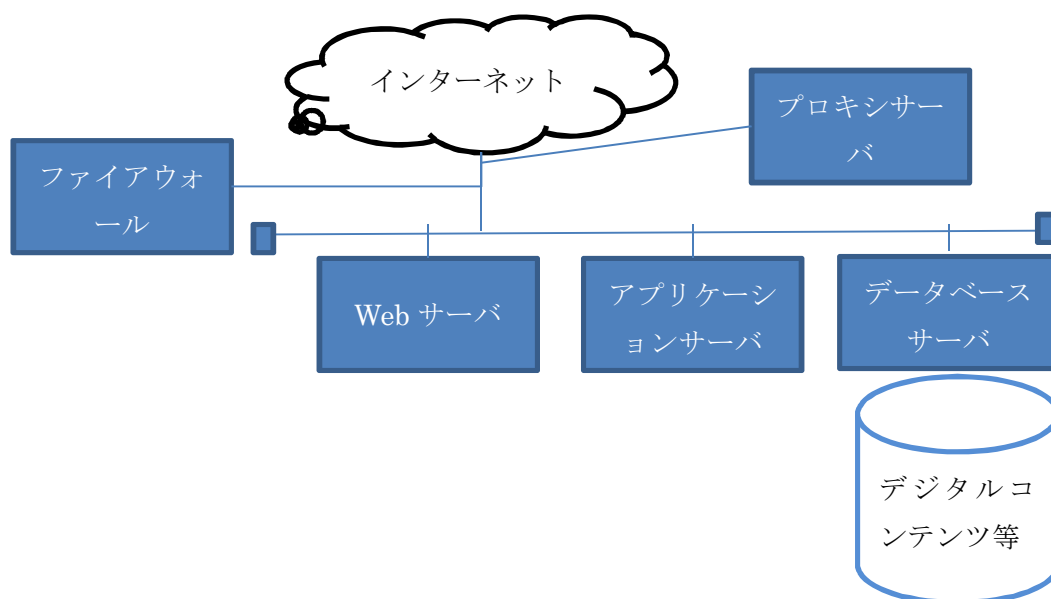


図 7-6 アーカイブシステムの構成例

③ ハードウェア要件の検討

ハードウェアを購入、レンタル、又はリースする場合に検討が必要となります。

既存のサーバ環境を利用する場合は、どの程度の設備が利用可能か、既存のサーバ管理者等に確認し、調整する必要があります。

サーバの維持管理、保守、バックアップ等の検討も併せて実施する必要があります。

④ ソフトウェア要件の検討

ソフトウェアを購入、レンタル、又はリースする場合に検討が必要となります。

Web アプリケーションシステムを基本とし、極力標準的な OS、ソフトウェアを用いることが必要です。

(c) 情報セキュリティ要件

震災関連デジタルアーカイブのシステムとして必要となる情報セキュリティ要件を検討します。震災関連デジタルアーカイブは、インターネットで公開するシステムであるため、一般的な Web システムの情報セキュリティを備える必要があります。また、提供先からコンテンツを受領する場合のウイルス対策等についても設定が必要です。

(d) 規模・性能要件

震災関連デジタルアーカイブのコンテンツは、将来的に拡張できるよう規模要件を決める必要があります。また、利用者に負担とならない応答性能についても記載が必要です。

その他に、以下の要件を検討する必要があります。

(e) 信頼性要件

システムの信頼性をどの程度担保するのか記載が必要です。

(f) テスト要件

システムテストをどのように行うかの記載が必要です。

(g) 移行要件

システムの移行条件の記載が必要です。

(h) 教育・研修要件

教育・研修に関する要件の記載が必要です。

(i) 保守・運用要件

保守・運用に関する要件の記載が必要です。

(3) 震災関連デジタルアーカイブにおけるシステム設計・構築について

システム設計・構築にあたっては、設計・構築を実施するシステム構築業者を調達する必要があります。

システム設計・構築業者の調達では、前項で検討したシステム要件に基づき、調達仕様書を作成し、調達を実施します。

以下にシステム設計・構築で重要となる観点について記載します。

(a) 画面デザインの検討

画面デザインは、その震災関連デジタルアーカイブの顔とも言うべきものであり、重要な決定事項となります。しかし、検討期間が長くなるとシステム設計・開発の遅延にもなるため、早い段階で承認することが必要です。

また、利用者の視点に立った画面デザインや配色を検討する必要があります。⁶¹

(b) 操作性の検討

震災関連デジタルアーカイブは、様々な立場の人が利用するシステムとなるため、操作性については、開発・試験段階でプロトタイプ等を作成し、操作性を確認する等も検討した方が良いでしょう。

また、利用者の視点に立った操作性を実現するため、想定される利用者（ペルソナ）を設定し操作性を検討するなど工夫をしましょう。⁶²

(c) システム設置場所との調整

既存のサーバールーム等に機器等を設置する場合は、早い段階でシステム設置場所との調整が必要となります。電源容量、空調設備、サーバラックの必要有無、ネットワーク帯域の確保等、調整事項を明確にして、調整することが必要です。

(d) 連携機能について

「NDL 東日本大震災アーカイブ」と連携する場合は、運用する国立国会図書館に対し、早い段階で連携の意思を伝え、調整を開始することが必要となります。提供されるサービス内容等は、「NDL 東日本大震災アーカイブ「ひなぎく」との連携について」を参照してください。

(4) 震災関連デジタルアーカイブにおけるシステム試験について

システム試験は調達したシステム構築業者において実施します。システム開発で通常実施される機能試験の他に、運用実証調査では、NDL 東日本大震災アーカイブとの連携試験を実施しました。

⁶¹ 画面デザインの検討方法について詳しく知りたい場合は、「参考 13 震災関連アーカイブサイトのユーザビリティに関する調査報告」を参照。

⁶² 操作性の検討方法について詳しく知りたい場合は、「参考 13 震災関連アーカイブサイトのユーザビリティに関する調査報告」を参照。

表 7-2 NDL 東日本大震災アーカイブとの連携試験

システム試験の観点	試験内容
システム間接続の確認	NDL 東日本大震災アーカイブから、震災関連デジタルアーカイブへネットワーク上での接続が行えることを確認する。
データ連携仕様の確認	NDL 東日本大震災アーカイブとの連携で使用するプロトコル（OAI-PMH）の仕様に基づき、震災関連デジタルアーカイブ側の機能が動作することを確認する。
連携データの確認	震災関連デジタルアーカイブが提供するメタデータが、連携用メタデータスキーマにしたがってマッピングされ、提供されていることを確認する。
データ連携運用の確認	一括収集・差分収集が正しく動作し、NDL 東日本大震災アーカイブ側での想定に基づいた運用が行えることを確認する。
データ連携 (ポータルオープン時のデータ)	NDL 東日本大震災アーカイブのオープン時に閲覧可能なメタデータを順次、収穫する。
NDL 東日本大震災アーカイブシステムから震災関連デジタルアーカイブへのリンク確認	NDL 東日本大震災アーカイブのユーザインタフェースから、各震災関連デジタルアーカイブへのリンクが正しく行われることを確認する。

(5) 震災関連デジタルアーカイブにおけるシステム運用について

システム試験完了後、本格的なシステム運用開始の前には、以下の準備が必要となります。

(a) 運用体制の検討

震災関連デジタルアーカイブのシステム運用主体を決定し、運用体制を決定します。運用体制は、例えば以下のような階層図で表現します。

なお、「参考 1 想定される役割例と本ガイドラインの構成」で示すとおり、運用体制の各担当者は、1人1役ではなく、兼任することや、複数人で担当すること、一部を外部機関に委託することなど、いくつかの形態が考えられます。

<運用体制の例>

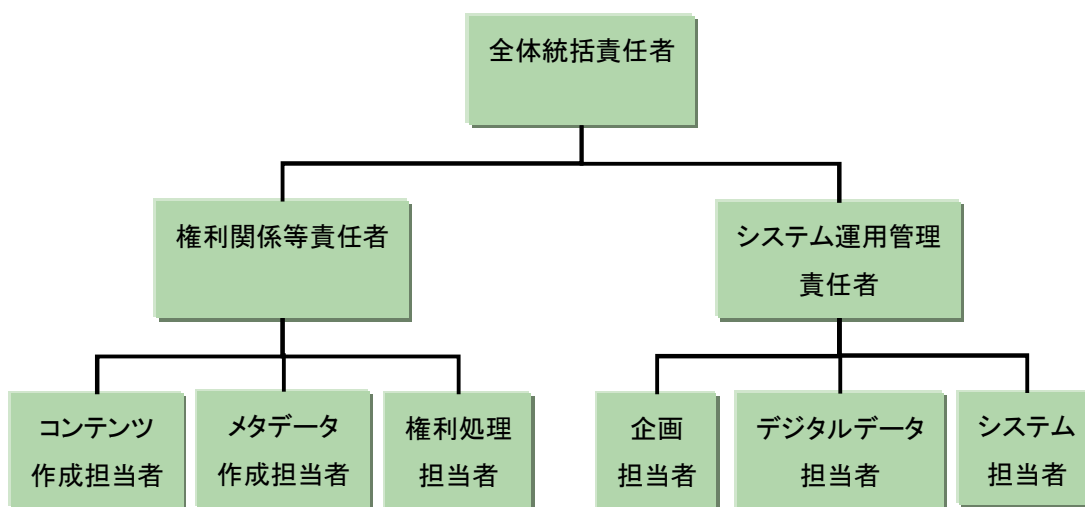


図 7-7 アーカイブの運用体制例

(b) 運用要件の検討

システムの保守・運用については、運用に必要な要件を検討し、運用業者を調達する必要があります。

運用要件としては、以下の項目を決定します。

- ・ システム稼働時間
 - ・ 問合せ対応時間、対応方法など
 - ・ データバックアップ要件（頻度、データ量等）
 - ・ 運用作業報告
 - ・ コンテンツ登録（メタデータ付与を含む）
 - ・ サービスレベル
- 等

(c) システム運用費用の検討

システム運用に必要な費用を以下のように検討し、算定します。

表 7-3 システム運用費用の検討

No.	項目	費用の内容
1	システム保守費用	開発したソフトウェアの保守費用
2	ハードウェア等費用	レンタル・リースの場合のみ
3	ハウジング費用	サーバ設置場所に係る費用（光熱費等含む）
4	システム運用費用	システム運用・監視等に係る費用、及び、問合せ窓口等の費用（主に人件費）
5	コンテンツ収集費用	コンテンツを追加収集する場合のみ（主に人件費）
6	デジタル化作業費用	アナログデータのデジタル化が必要な場合のみ（主に人件費）
7	メタデータ付与費用	コンテンツを追加収集する場合のみ（主に人件費）
8	その他費用	消耗品等の経費

2. 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、震災関連デジタルアーカイブシステムを構築し、運用を行いました。

(1) では、運用実証調査で実施したシステム構築・運用について、全体構成と標準的に実施した構築・運用手順を説明し、(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システム構築・運用について

(a) システムの全体構成

運用実証調査においては、5つの運用モデルプロジェクトでそれぞれ震災関連デジタルアーカイブシステムを構築しました。

ただし、共通の標準仕様については、5つの運用モデルプロジェクトで別々に検討するのではなく、システム設計を開始する前に全体で検討を実施しました。

共通の標準仕様については、以下のとおりとなります。

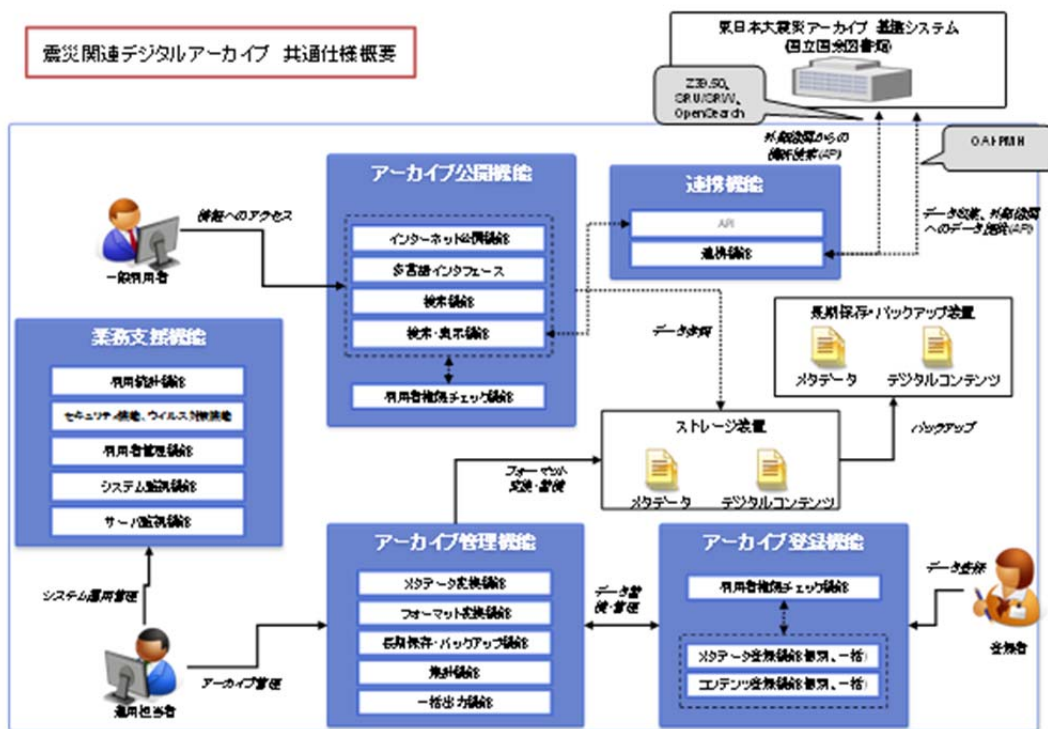


図 7-8 システム全体構成

(b) システム機能要件

共通仕様の機能構成は、以下のとおりとなります。

表 7-4 震災関連デジタルアーカイブの機能構成

機能構成 (大項目)	機能構成 (小項目)
アーカイブ登録機能	メタデータ登録機能 (個別、一括)
	コンテンツ登録機能 (個別、一括)
アーカイブ管理機能	メタデータ変換機能
	フォーマット変換機能
	長期保存・バックアップ機能
	集計機能
	一括出力機能
	利用者権限チェック機能
アーカイブ管理機能	インターネット公開機能

機能構成（大項目）	機能構成（小項目）
	多言語インタフェース
	検索機能
	検索表示機能
業務支援機能	利用統計機能
	セキュリティ／ウイルス対策機能
	利用者管理機能
	システム監視機能
	サーバ監視機能
連携機能	メタデータ連携機能

(c) ハードウェア要件

震災関連デジタルアーカイブ構築において必要となるハードウェア要件について共通に留意した点は、以下のとおりです。

① 規模要件の確定について

一般的に規模要件は、システムに登録されるデータ量から正確に見積もりが可能ですが、震災関連デジタルアーカイブにおいては、コンテンツの収集予定件数だけでは、1件あたりのデータ容量が様々であり、データ容量を見積もることが困難です。

したがって、機器整備を行う上では、将来的に拡張性のあるハードウェアを選定する必要があります。

また、長期保存を考慮し、ファイル形式について保存用デジタルデータと公開用デジタルデータを別々に用意する等の配慮も必要となります。

② データの遠隔バックアップ

東日本大震災においては、多くの企業や自治体でデジタルデータの消失がありました。バックアップデータは、遠隔地に保管する、バックアップシステムを用意する等の配慮が必要となります。

(d) ソフトウェア要件

震災関連デジタルアーカイブ構築においては、基盤となるデジタルアーカイブシステムについて、新規に設計・構築する、又は、市販のパッケージを流用する等の選択が必要となります。それぞれのメリット・デメリットは、以下のとおりです。

表 7-5 開発手法による比較

開発方法	メリット	デメリット
新規に設計・構築する場合	利用者や運用者の要望を充分に取り入れたシステム構築が可能。	一から設計を行うため、開発期間を充分に確保する必要があり、開発費用も大きくなる。
市販のパッケージ等を流用する場合	比較的安価で開発が可能であり、開発期間も短縮できる。	パッケージの制約に従う必要があるため、利用者の要望を充分に取り入れることができないことがある。

運用実証調査では、短期間での開発を行う必要があったため、新規に設計・構築するのではなく、商用パッケージやオープンソースを活用することとしました。

以下に各プロジェクトで使用した基盤となるデジタルアーカイブシステムをまとめます。

表 7-6 各地域のシステム基盤

青森プロジェクト	・商用パッケージを用いた開発
岩手プロジェクト	・商用パッケージを用いた開発
宮城東北大学	・オープンソースソフトウェアと一部商用パッケージを用いた開発
宮城河北新報社	・既存環境を基に、オープンソースソフトウェアと一部商用パッケージを用いた開発
福島プロジェクト	・商用パッケージを用いた開発

(e) 外部連携要件

運用実証調査で構築した震災関連デジタルアーカイブは、NDL 東日本大震災アーカイブと連携し、横断的な検索を可能としています。

運用実証調査では、NDL 東日本大震災アーカイブと連携を実施するため、メタデータ提供・収集用の共通仕様を採用しました。

通信プロトコルについては、OAI-PMH⁶³を選定し、連携用メタデータスキーマを介し

⁶³ OAI-PMH は、メタデータを交換するためのプロトコルの名称のこと。特定のアプリケーションに依存することなく、自動的にメタデータを交換できる利点があります。NDL 東

てメタデータ連携を実現しました。連携用メタデータスキーマは、各プロジェクト共通で定めることにより、NDL 東日本大震災アーカイブと運用実証調査間の連携を統一することが可能となりました。

なお、連携用メタデータスキーマから NDL 大震災メタデータスキーマへのマッピング変換処理は、NDL 東日本大震災アーカイブ側で実装し、利用者からの検索・閲覧を可能としました。

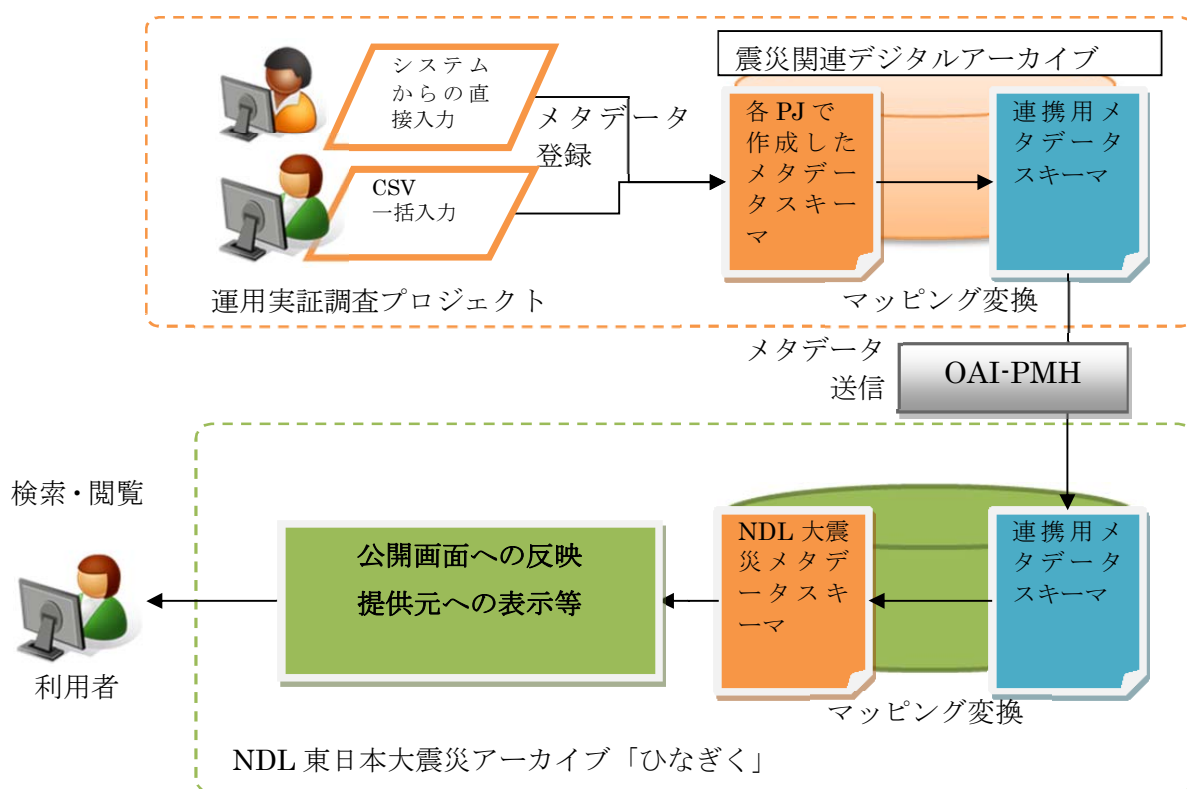


図 7-9 NDL 東日本大震災アーカイブとのシステム連携

(f) 運用要件

① 運用体制について

震災関連デジタルアーカイブにおいては、コンテンツの権利者とコンテンツの利用者間において、何らかの権利処理に係る問題が発生した場合の問合せ窓口を用意する必要があります。

また、コンテンツのメタデータ付与設定で非公開とすべきものが、誤って公開となってしまう場合、速やかにメタデータを修正する等の運用体制が必須となります。

日本大震災アーカイブがシステム間連携手段として採用しているので、運用実証調査でもこのプロトコルを採用した。詳細は、
<http://www.openarchives.org/OAI/openarchivesprotocol.html> を参照。

また、収集・保存するコンテンツが将来的に増加することを前提に、コンテンツ収集、デジタル化作業を実施するための体制も必要となります。

② 運用支援ツールについて

震災関連デジタルアーカイブにおいては、システム運用について人が実施しなければならない作業を極力少なくするため、運用管理システムを導入することや、コンテンツ管理ツールの充実を図っていく必要があります。

(2) 青森県における実証調査

(a) システム構築体制

青森プロジェクトのシステム構築は、「あおもりデジタルアーカイブ・コンソーシアム」の主要メンバーである以下の2機関で実施しました。

表 7-7 システム構築体制

八戸工業大学	以下の設備等の提供と維持管理 <ul style="list-style-type: none"> ・ハードウェア機器の設置場所 ・電源・空調等 ・ネットワーク環境 ・サーバ起動・停止作業 ・バックアップメディア状況の確認
システム構築事業者	アーカイブサイトにおいて動作する一切のアプリケーションの運用・保守 アーカイブサイトの問い合わせ対応

公開コンテンツに関する異議申し立て等の問い合わせは、あおもりデジタルアーカイブ・コンソーシアム事務局（八戸工業大学内）において受け付け、システム構築事業者において非公開設定への変更対応を行います。

(b) システム構築

青森プロジェクトで構築したシステム構成は、以下のとおりとなります。

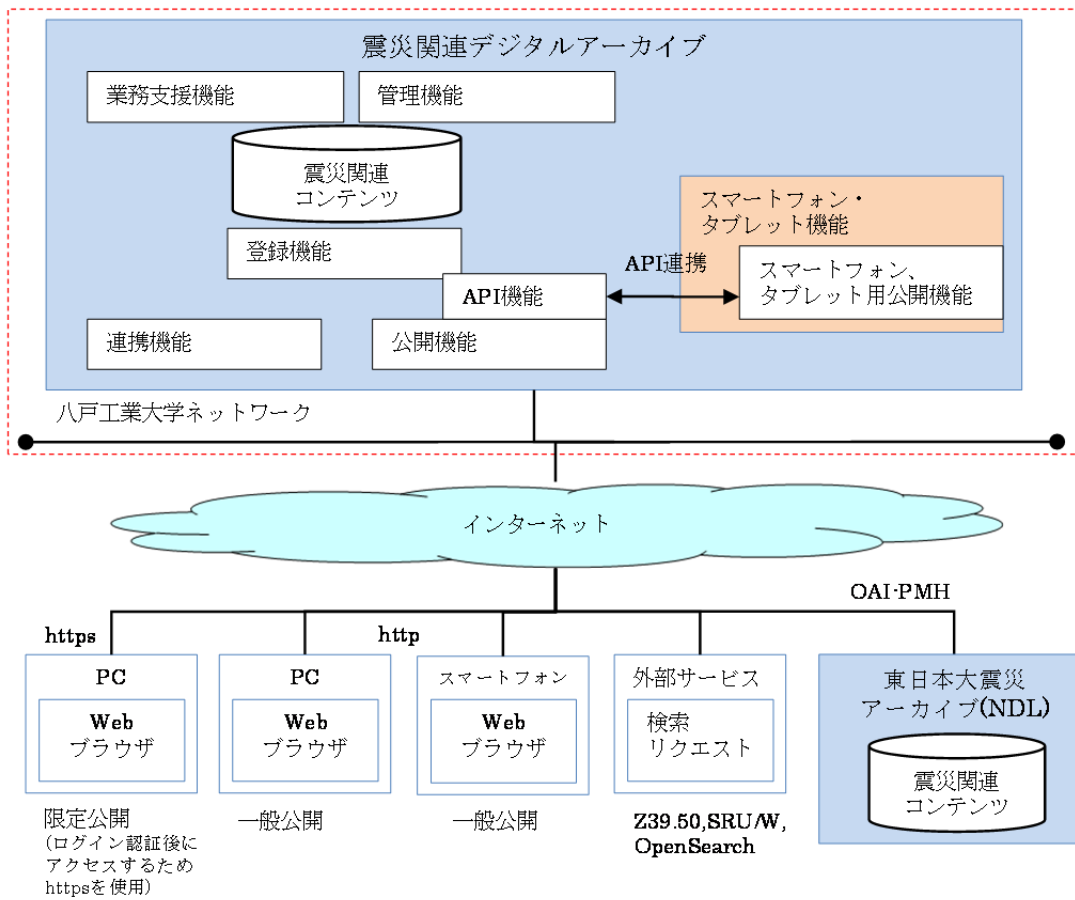


図 7-10 システム構成図

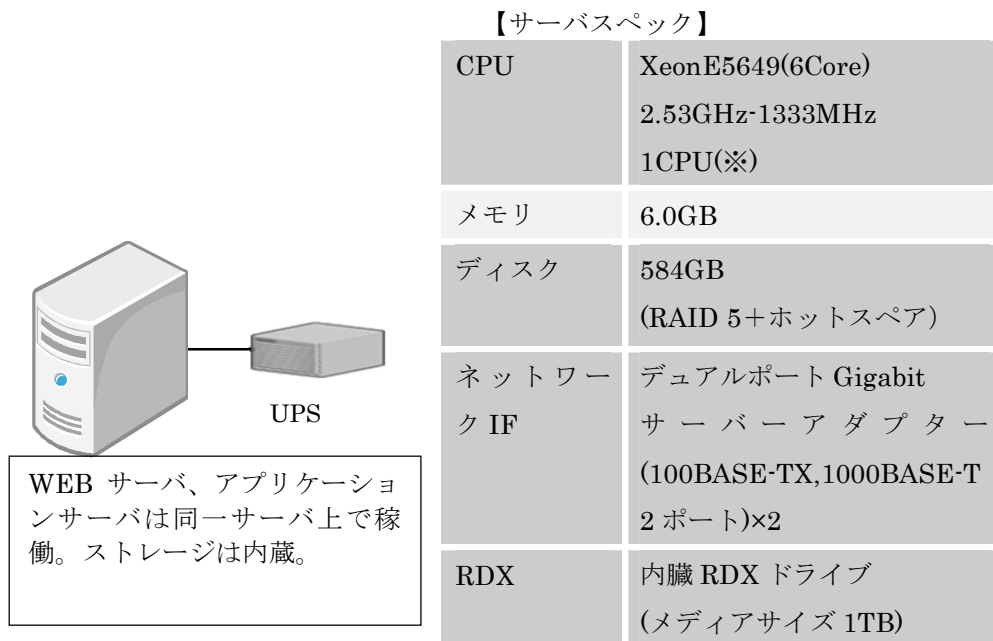


図 7-11 ハードウェア構成図

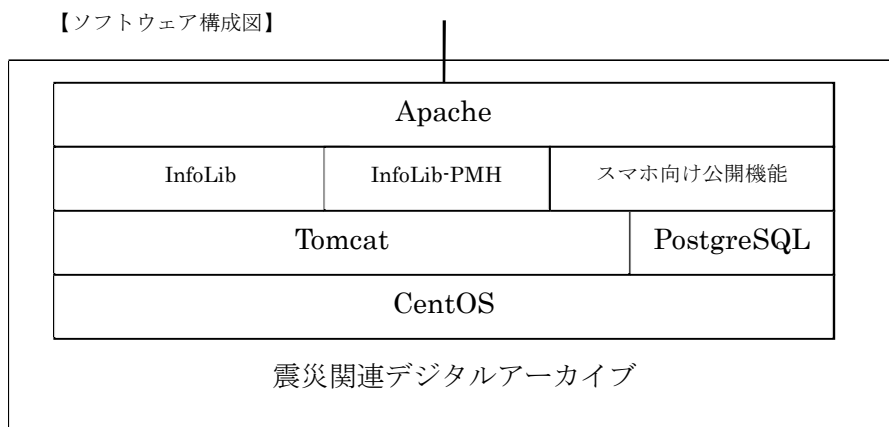


図 7-12 ソフトウェア構成図

① ICT 環境 (ハード)

青森プロジェクトでは、ハードウェア環境を八戸工業大学内に設置することとしました。八戸工業大学は本プロジェクトの共同実施主体であり、該当機関が既存で保有するサーバ室、ネットワーク環境等のインフラを最大限活用することとしました。したがって、セキュリティポリシー及びネットワーク設定については八戸工業大学のポリシーに従うこととしました。

ハードウェアは小規模なサーバ構成としました。これは青森プロジェクトが実証調査期間内に収集する震災関連コンテンツ量があまり多くないという理由からでした。

ただし、今後の震災関連コンテンツの増加に対してスケールアウトが可能な構成であり、拡張性のある構成としました。

バックアップは RDX⁶⁴を使用することで、バックアップ運用におけるテープの入替え作業などが簡単になり、作業者の負担を軽減することを目指しました。

② ICT 環境（ソフト）

収集した震災関連コンテンツを登録、公開、管理する機能、及び NDL 東日本大震災アーカイブとの連携機能は構築期間が短期間でもあるため、実績のあるデジタルアーカイブシステム構築用の商用ミドルウェアを利用した。また、Z39.50、SRU/SRW、OpenSearch、OAI-PMH といった連携プロトコルを標準で実装していることも本ミドルウェアを選定した理由です。

青森プロジェクトでは構築したアーカイブサイトの公開機能をスマートフォン、タブレット端末への最適化を検討するため、スマートフォン・タブレット用の公開機能を開発しました。開発言語は Java とし、アーカイブシステムと同一サーバの Web サーバ、アプリケーションサーバ上で動作します。

ハードウェア機器の監視ソフトウェアとしてオープンソースソフトの Zabbix⁶⁵を採用しました。Zabbix はリモート環境のハードウェア機器リソースをネットワーク越しに監視することができます。

表 7-8 使用ソフトウェア一覧

Cent OS release 6.3
Apache 2.2.22
Apache Tomcat 6.0.13
PostgreSQL 8.3.13
InfoLib-Meta
Zabbix 1.8.15
F-Secure 9.20

③ メタデータスキーマレジストリへの登録

青森プロジェクトでは、設計したメタデータスキーマ⁶⁶をメタデータスキーマレジストリ⁶⁷「MetaBridge」に登録しました。メタデータスキーマをメタデータスキーマレジ

⁶⁴ Removable Disk Exchange system の略。高速で可搬性に優れたバックアップ規格。

⁶⁵ サーバ、ネットワーク、アプリケーションを監視するためのオープンソースソフトウェアのこと。

⁶⁶ メタデータを入力するためのテンプレートのこと。

⁶⁷ メタデータスキーマを蓄積し、広く公開するためのしくみのこと。詳細は、「参考 14 メ

ストリに登録することで、他のアーカイブの設計者が自身のメタデータスキーマを設計する時や、青森県のアーカイブと連携しようとした時に青森県のメタデータスキーマを参照できるようになります。

これにより、アーカイブの相互連携や利用者による横断的な検索を実現できるようになり、アーカイブの利活用が全体として推進されることが期待できます。

ローカル名	ラベル	上位プロパティ	定義域	値域	DCとの関連
abstract	説明、要約	dc:terms:description			dc:description
acc	特許的範囲情報	dc:terms:accessRq	tsa		dc:rights
accessFree	コンテンツへの無償アクセス	dc:terms:accessRq	tsa		dc:rights
accessPeriod	コンテンツへのアクセス期間制限	dc:terms:accessRq	tsa		dc:rights
accessRestriction	メタデータアクセス	dc:terms:accessRq			dc:rights

図 7-13 メタデータスキーマレジストリに登録された青森県震災関連デジタルアーカイブのメタデータスキーマ

④ 特徴とする機能

▶ スマートフォン・タブレット用公開機能

アーカイブサイトの公開機能、キーワードからの検索、地図から検索をスマートフォン・タブレット端末用に最適化した機能（以下、スマートフォン用画面という。）を開発しました。構築したスマートフォン用画面とアーカイブサイトの関連は以下のとおりです。

スマートフォン用画面は、アーカイブサイトの検索用 API にリクエストし、その結果を整形し画面表示する機能を有し、独自の検索モジュール等は保持しません。これによってスマートフォン用画面からもアーカイブサイトと同レベルの検索が可能になりました。

「メタデータスキーマレジストリ」を参照。

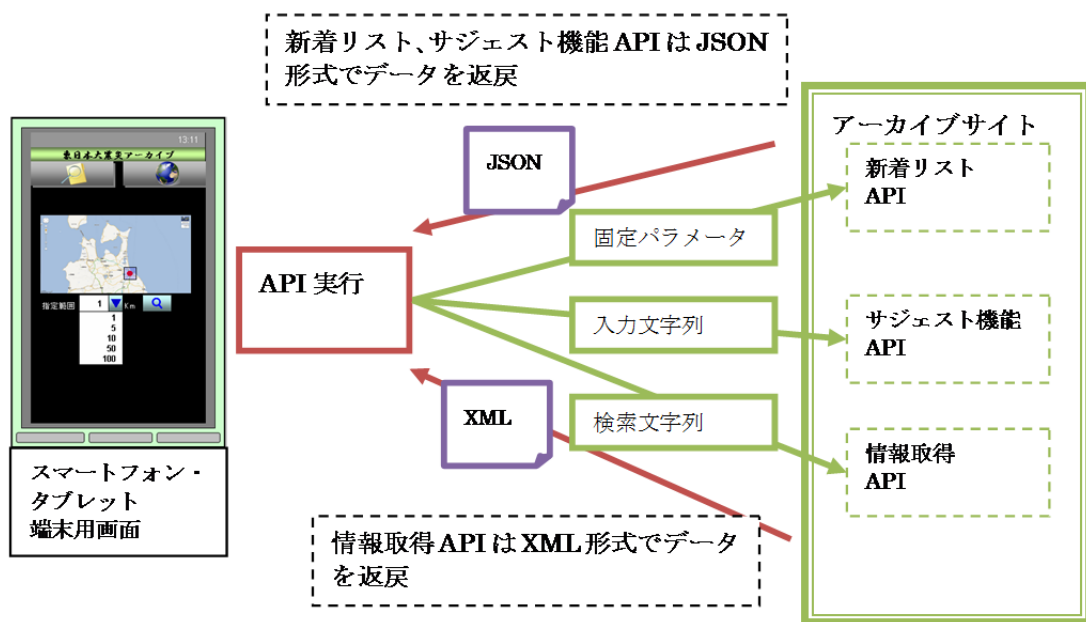


図 7-14 アーカイブサイト機能とのデータ入出力の関係図

▶ 幅広いデジタルコンテンツを Web 配信

青森プロジェクトでは、収集が予想される幅広いコンテンツフォーマットを配信できるようなシステムを構築しました。

写真などの画像ファイルは JPEG、PNG、GIF、BMP、TIFF といったファイルフォーマットが予測されました。これらのファイルフォーマットに対応し、Web 配信用のファイル変換を可能としました。本アーカイブシステムとして、対応しているコンテンツフォーマットは以下のとおりです。

表 7-9 Web 配信用ファイル変換

種別	形式	拡張子	配信用変換後拡張子
静止画	JPEG	jpe	.jpg
		jpeg	
		jpg	
	GIF	gif	
		giff	
	PNG	png	
	TIFF	tif	
		tiff	
BMP	bmp		
AI	ai		

種別	形式	拡張子	配信用変換後拡張子
動画	WindowsMedia (動画)	wmv	.mp4
		avi	
		asf	
	QuickTime (動画)	mov	
		qt	
	MPEG (動画)	mpg	
		mpeg	
		ts	
		m2t	
		m2ts	
		mp4	
Flash Video	flv		
	f4v		
音声	WindowsMedia (音声)	wma	.mp3
	QuickTime (音声)	aiff	
	MPEG (音声)	mp2	
		mp3	
	WAVE	wav	
	Flash Video	f4a	
f4b			
その他	PDF	pdf	.pdf

また文書ファイルは主に Microsoft の Office 形式、ODF 形式、PDF 形式が収集されることが予測されました。これらのファイルを配信用フォーマットとしては PDF フォーマットに変換して配信しています。映像、音声はそれぞれ MPEG4、MP3 に変換、配信可能としました。

➤ マトリクス表示

青森プロジェクトでは震災前～復興までを時系列で比較できるようなアーカイブを目指しました。これを実現するため、復興フェーズ（震災前、震災直後、復旧、復興）とコンテンツ分類をマトリクスとして表現したコンテンツ閲覧機能を構築しました。

横軸に復興フェーズ、縦軸に分類を配置することで、収集されたコンテンツの分布を直感的に表現できるようにしました。各マトリクスには該当コンテンツが表示されます。



図 7-15 マトリクス表示

(c) システム構築における課題

① スマートフォン・タブレット端末の対応について

青森プロジェクトでは、スマートフォン、タブレット端末対応用の公開機能を構築しましたが、今後開発される端末の機種によっては、うまく表示できないなどの不具合が発生することが予想されます。

iOS 端末⁶⁸に関しては画面サイズが固定されており、将来的にも多様なバリエーションが発生するとは考えにくいと思われます。しかし、Android 端末⁶⁹に関しては様々な画面サイズが既に発売されており、アーカイブの利便性を高めるためにどの範囲まで対応するかは、今後の課題となります。

またコンテンツの表示に関して、画像、文書の表示は各端末のブラウザで表示することが可能ですが、映像・音声ファイルは端末ごとに搭載されるソフトウェアに依存してしまい、端末によっては正常に表示されない可能性も考えられます。

⁶⁸ Apple 社が開発しているモバイル OS (iOS) を搭載した端末のこと。

⁶⁹ モバイル用オープンソース・オペレーティングシステム、ミドルウェアである Android を搭載した端末のこと。

(d) システム運用

① 運用体制

青森プロジェクトのアーカイブ運用体制は、以下のとおりとなります。

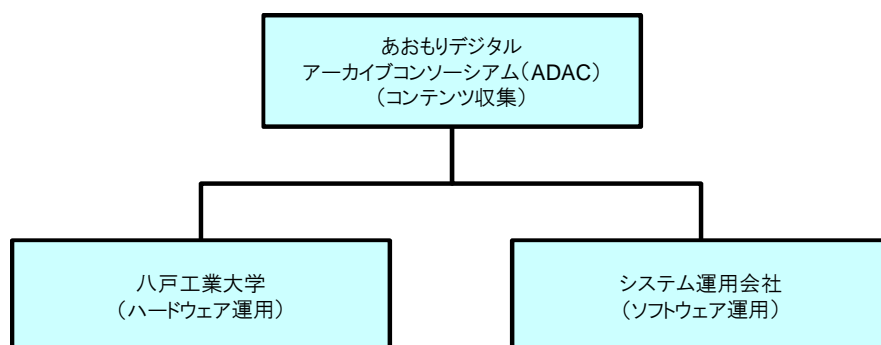


図 7-16 システム運用体制

② 役割分担

アーカイブ運用における役割分担は、以下のとおりとなります。

表 7-10 役割分担

No.	組織名	役割分担
1	あおりデジタルアーカイブ・コンソーシアム (ADAC)	・運用主体 ・コンテンツの収集・保存・公開 ・問合せ窓口
2	八戸工業大学	・ハードウェア運用・保守
3	システム運用会社	・ソフトウェア運用・保守

③ システム設置場所

サーバは、八戸工業大学のサーバ室へ設置し、サーバ及びネットワークの運用ポリシーは八戸工業大学のポリシーに従うこととしました。

(e) システム運用における課題

アーカイブ運用にあたり、八戸工業大学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、システム開発会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ますが、これらの資金的手当ての目途が立っておらず、後年運用の主体となるあおりデジタルアーカイブ・コンソーシアム(ADAC)と費用負担の方法を検討します。

(3) 岩手県における実証調査

(a) システム構築

岩手プロジェクトにおいて構築するシステムの全体像は以下のとおりです。なお、システム構成およびソフトウェア構成に関しては、青森プロジェクトと同様です。

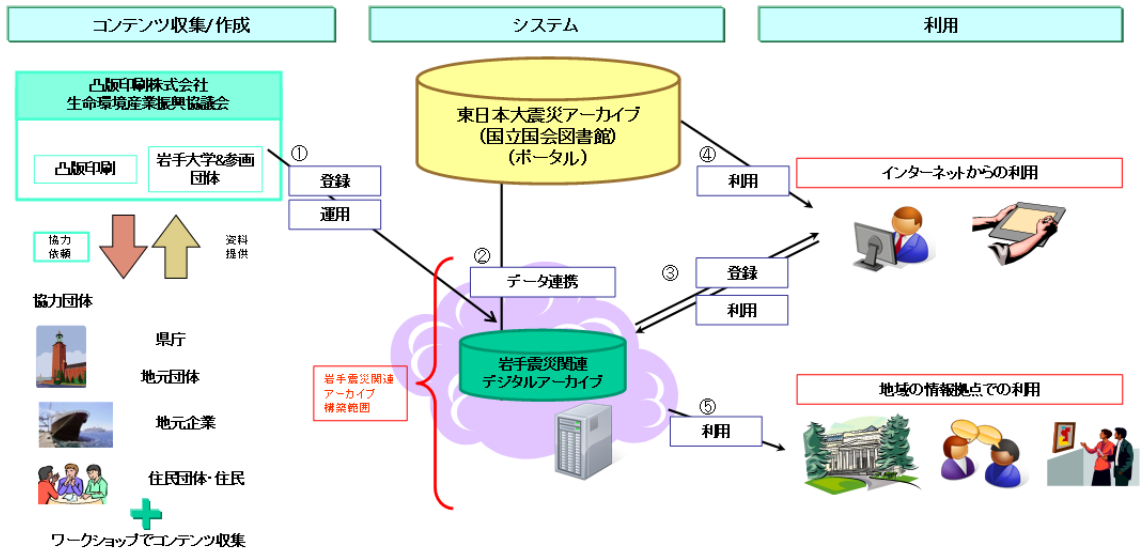


図 7-17 岩手プロジェクトにおけるシステムの全体像

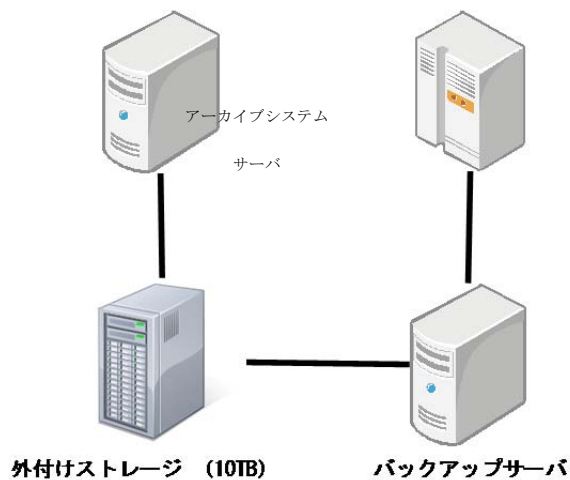


図 7-18 ハードウェア構成図

(b) システム運用

① 運用体制

アーカイブ運用体制は、以下のとおりです。

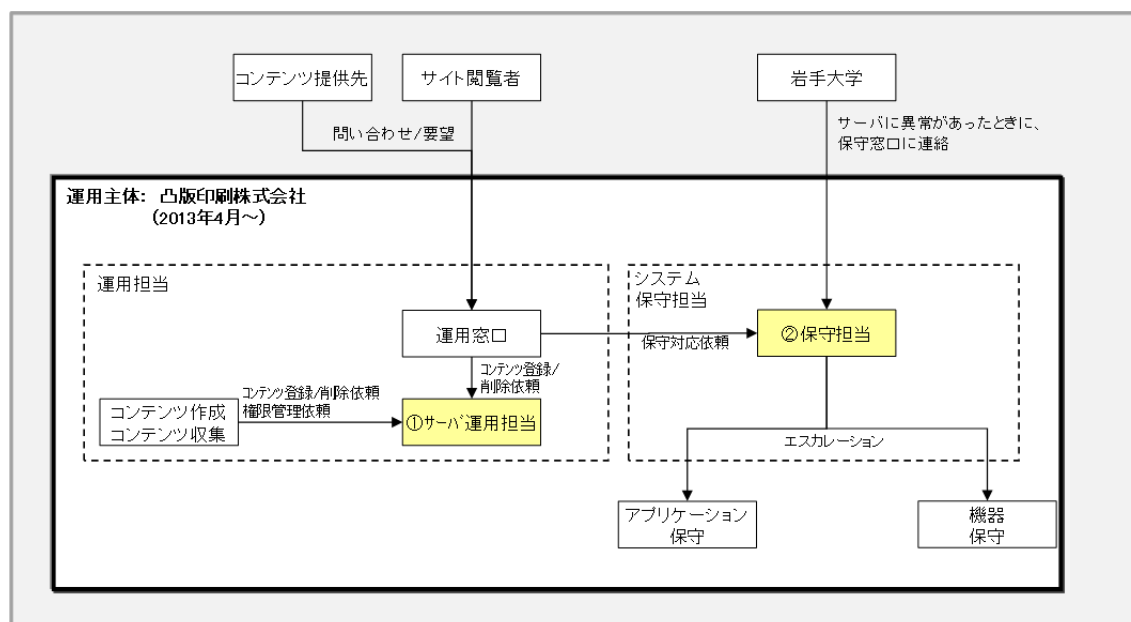


図 7-19 システム運用体制

② 役割分担

アーカイブ運用における役割分担は、以下のとおりとなります。

表 7-11 役割分担

No.	役割分担	作業内容
1	運用窓口	・ 問合せ受付・回答
2	サーバ運用担当	・ コンテンツの登録・削除
3	保守担当	・ システム監視 ・ 障害発生時の切り分けとエスカレーション

③ システム設置場所

サーバは、岩手大学のサーバ室へ設置し、サーバ及びネットワークの運用ポリシーは岩手大学のポリシーに従うこととしました。

(c) システム運用における課題

今後、アーカイブの運用主体について、後年運用を安定的に継続させるため、岩手県全域で震災関連デジタルアーカイブ事業支援を構築する体制づくりを目指し、岩手県を

中心とした公的機関がサーバを共有管理し、各自治体が地域で運用できるモデルへの移行を働きかけていきます。

(4) 宮城東北大学における実証調査
(a) システム構築体制

宮城東北大学のシステム構築体制は、以下のとおりとなります。

表 7-12 システム構築体制

要員	人数	役割
管理者	1	作業状況の確認
プログラム開発者	5	アプリケーションの開発テスト、データ投入
インフラ担当者	1	OS、ミドルウェアのインストール及び設定

(b) システム構築

東北大学プロジェクトで構築したシステム概要図は以下のとおりとなります。

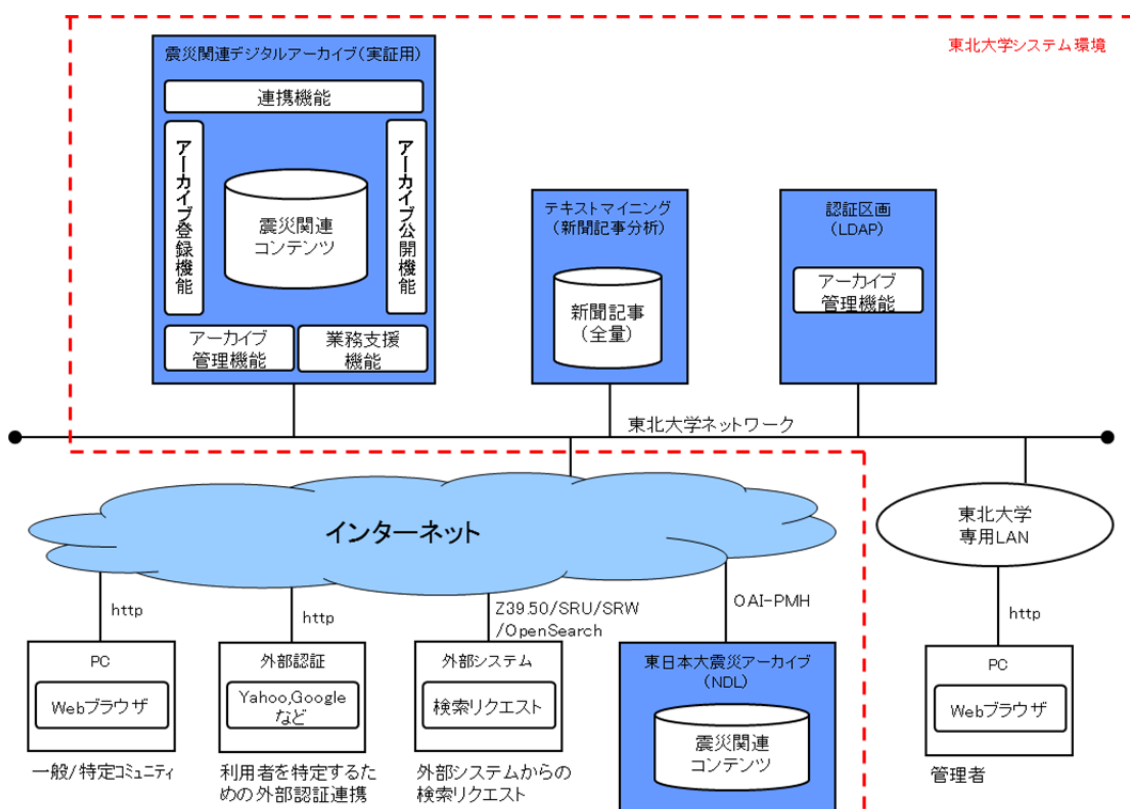


図 7-20 システム構成図

① ICT 環境（ハード）

東北大学プロジェクトでは、河北新報社のシステムの保守運用を行うとともに、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が必要でした。そこで、機器の仮想化技術を採用し、ICT 環境を構築しました。

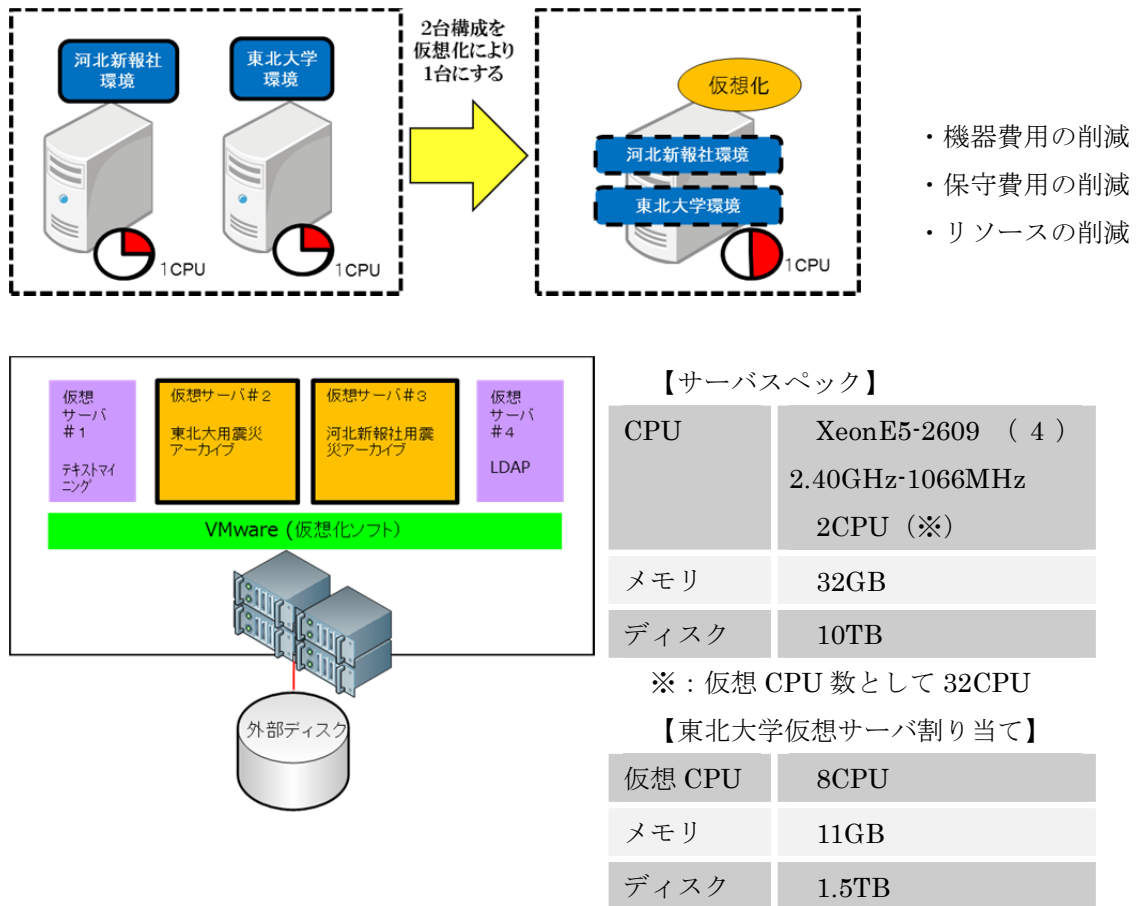
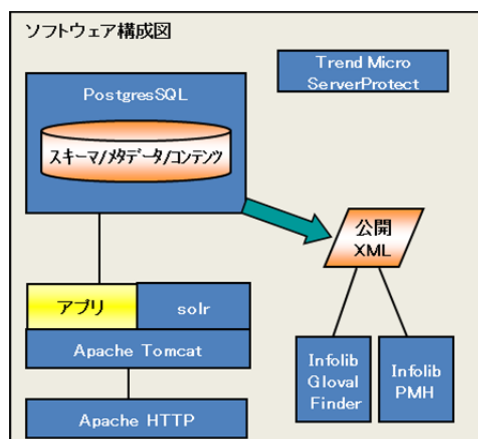


図 7-21 ハードウェア構成

② ICT 環境（ソフト）

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、NDL 東日本大震災アーカイブとの連携機能については、実績のある商用システムを導入することで、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と

しました。



【オープンソース一覧】

RHEL6
Apache Tomcat 6
Apache HTTP 2
PostGreSQL 9
PostGreSQL 8
Seasar2
Solr 3

【商用製品】

Infolib Global Finder
Infolib PMH

図 7-22 ソフトウェア構成

③ 特徴となる機能

➤ キーワードを重視した検索

本プロジェクトのアーカイブシステムは、一般的なアーカイブシステムと実装する機能はほぼ同一です。一方で、今回コンテンツを検索するには、「時間」、「空間（位置情報）」、「テーマ（検索語彙）」の3つの組み合わせによる検索となります。東北大学では、特にキーワードの検索が中心になると考え、実装を行いました。

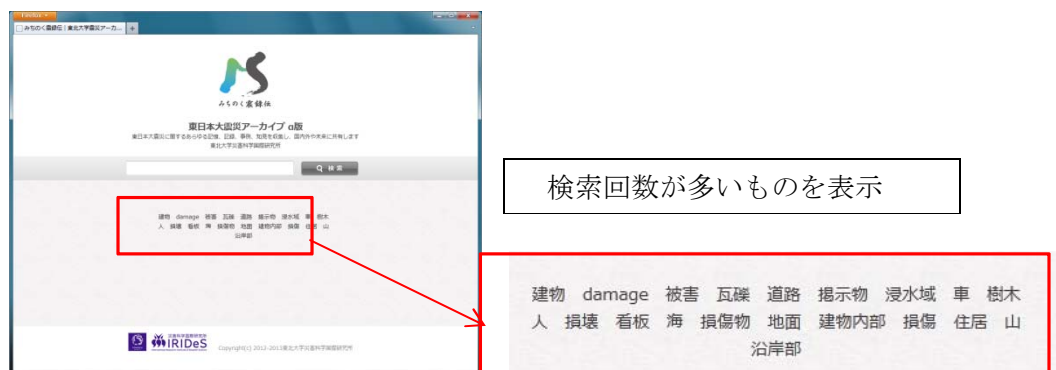


図 7-23 キーワードを重視した検索



検索したコンテンツのキーワードを集計し、表示することにより、他のキーワードを探しやすくする。

図 7-24 よく使うキーワード

また空間情報（位置情報）は、写真に GPS 情報がある物については、そこから情報を取得し、位置情報を表示するようにしています。



地図の表示は、GoogleMAP を使用している。

図 7-25 位置情報の表示

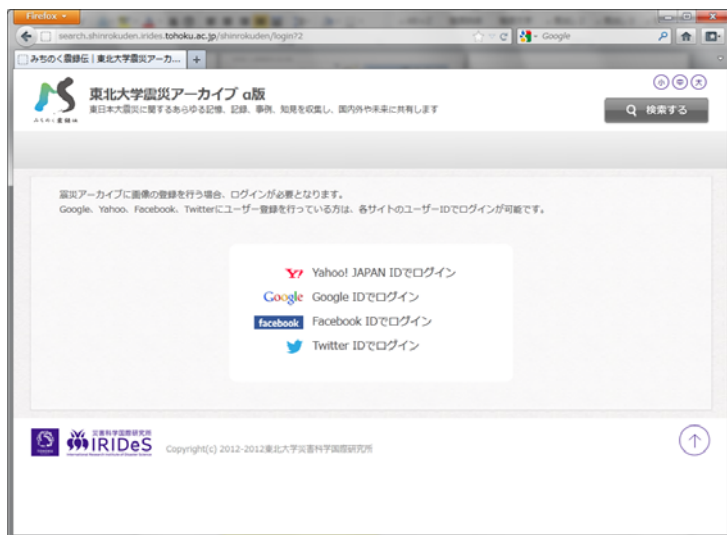
一方、検索機能の課題として、時系列に並べる機能が不足していることが挙げられます。データを相当数入れてみると、条件を絞った検索後の時系列の状態を見たいとなるとの意見が得られましたが、現状では機能上そのような閲覧はできません。現在の機能にどのように時系列での閲覧を組み込むかは、今後の課題となります。

➤ 一般投稿におけるログイン ID

宮城東北大学プロジェクトでは、一般からの投稿を可能とする投稿画面を作成しまし

た。投稿画面は、投稿者が特定できるように、ログイン操作をさせた後、投稿できる仕組みとしました。

ログインは、Yahoo、Google、Facebook、Twitter のユーザ ID でログイン出来るようにすることで、ユーザ管理を簡略化しました。



Yahoo、Google、Facebook、Twitter には、各 API がありユーザ名、メールアドレスを取得可能となっている。コンテンツ投稿者へ問い合わせを行う場合は、メールで問い合わせを行うことが可能である。

図 7-26 ユーザ登録の連携

(c) システム構築における課題

本プロジェクトのシステムに対して、以下の 3 つの課題がありました。

① キーワードのつながりの充実

検索のしやすさを求める場合、関連するコンテンツが探せることが重要であり、利用者から、関連するコンテンツを探せる機能が必要との意見を利用者の方からいただきました。関連性については、システム的には、メタデータ項目を使用し、表示することとなりますが、どの項目同士で関連性を持たせるか等の検討は、難しいテーマであり、今後検討が必要となります。

② 同義語機能の充実

メタデータの課題として「キーワードのばらつき」があります。これは、「がれき、ガレキ、瓦礫」などのように、同じ内容を指すが、違う語彙が付与されているケースとなります。対応策として、同義語の定義をシステム的に実装することで解決できますが、一方、同義語の定義を作成するのは、時間を要するため、今後の検討課題となります。

③ 位置情報の取得

東日本大震災の特徴として、地域により被害の状況が違えることが挙げられます。その

ため、コンテンツの位置情報が重要になります。研究の目的で利用する場合、コンテンツに正確な位置情報があると、より使いやすいコンテンツとなります。これに対応するために、GPS 機能付カメラを採用し、写真の属性より位置情報を抽出し、住所情報への変換を系統的に実施しました。ただし、GPS 機能が無いカメラで撮影された写真の場合には、正確な位置情報が取得できませんでした。

(d) システム運用

① 運用体制

アーカイブ運用体制は、以下のとおりです。

表 7-13 システム運用体制

役割		団体
運用主体		東北大学
	コンテンツ	東北大学
	システム	東北大学

② 役割分担

アーカイブ運用における役割分担は、以下のとおりです。

表 7-14 役割分担

No.	組織	担当
1	東北大学	<ul style="list-style-type: none"> ・運用主体 ・コンテンツ収集 ・システム保守・運用

③ システム設置場所

サーバは、東北大学のサーバ室へ設置し、サーバ及びネットワークの運用ポリシーは東北大学のポリシーに従うこととしました。

(e) システム運用における課題

アーカイブ運用主体と利用者間の対話の中から継続的に成長するシステムとするため、コンテンツの公開に伴う一次窓口は、東北大学で実施することとなりますが、コンテンツの内容については、二次問合せ先として、みちのく震録伝に窓口を設置することとしました。

今後は、運用実証調査でコンタクトをとった他機関とも連携し、コンテンツの相互運用を行えるようにしていくこととしました。

(5) 宮城河北新報社における実証調査

(a) 体制

宮城河北新報社プロジェクトでは、コンテンツ、システムに関しては、上述のとおり東北大プロジェクトと連携し、構築しました。

宮城河北新報社プロジェクトに参加する事業者は下表のとおりとなります。東北大プロジェクトとの違いとして、株式会社河北新報社が参画しており、マスメディアに關与する事業者として震災関連資料の収集を行いました。

表 7-15 システム構築体制

企業・団体名	主たる役割
日本アイ・ビー・エム株式会社	宮城河北新報プロジェクトにおける運営事務局 デジタルアーカイブシステムの設計・構築 実証調査
株式会社 河北新報社	震災資料収集 実証調査支援
東北大学災害科学国際研究所	宮城東北プロジェクトにおける運営支援
日本総合システム株式会社	デジタルアーカイブシステムの設計・構築
株式会社テクノプラン株式会社	デジタルアーカイブシステムの設計・構築
インフォコム株式会社	デジタルアーカイブシステムの設計・構築
株式会社 ラネックス	デジタルアーカイブシステムの設計・構築

(b) システム構築

宮城河北新報社プロジェクトでは、以下の図に示すようなシステムを構築しました。

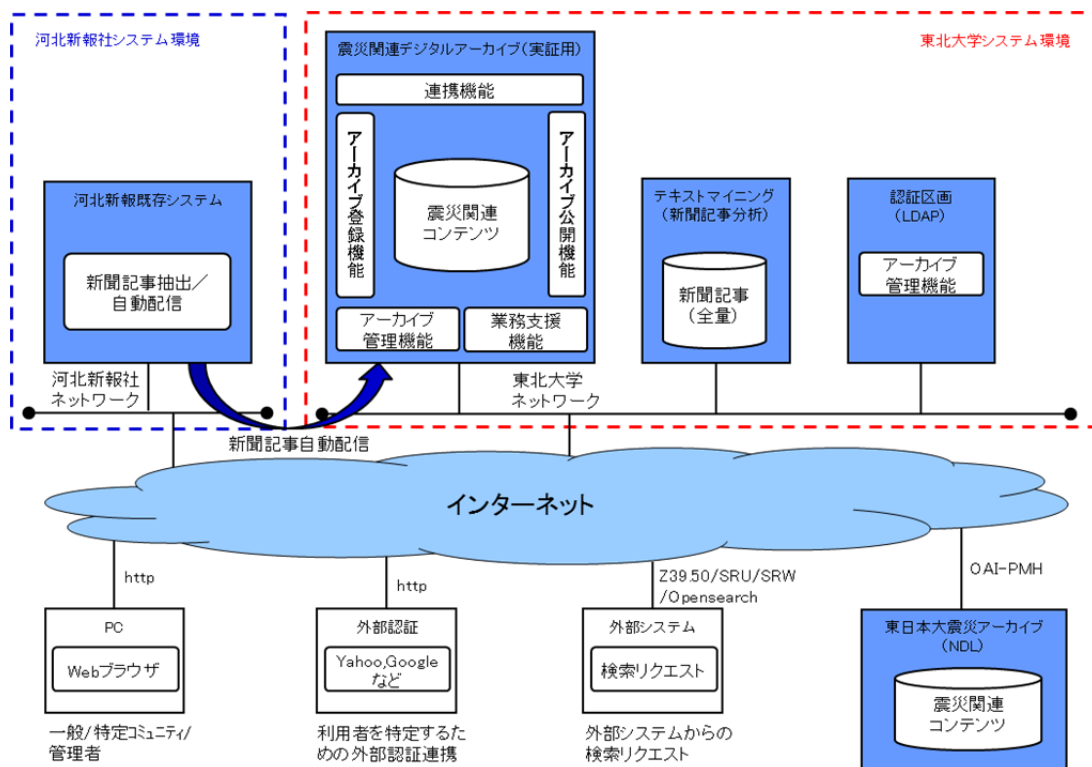
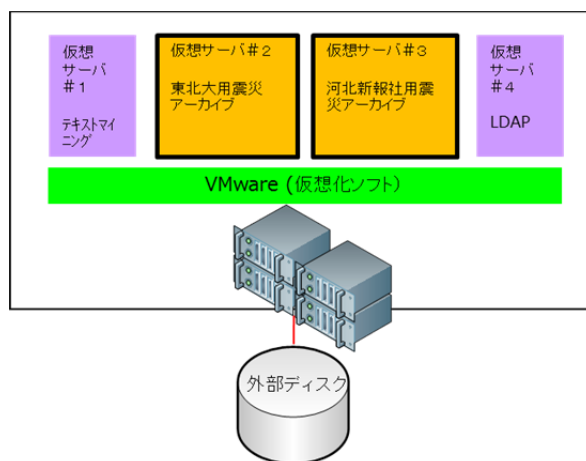
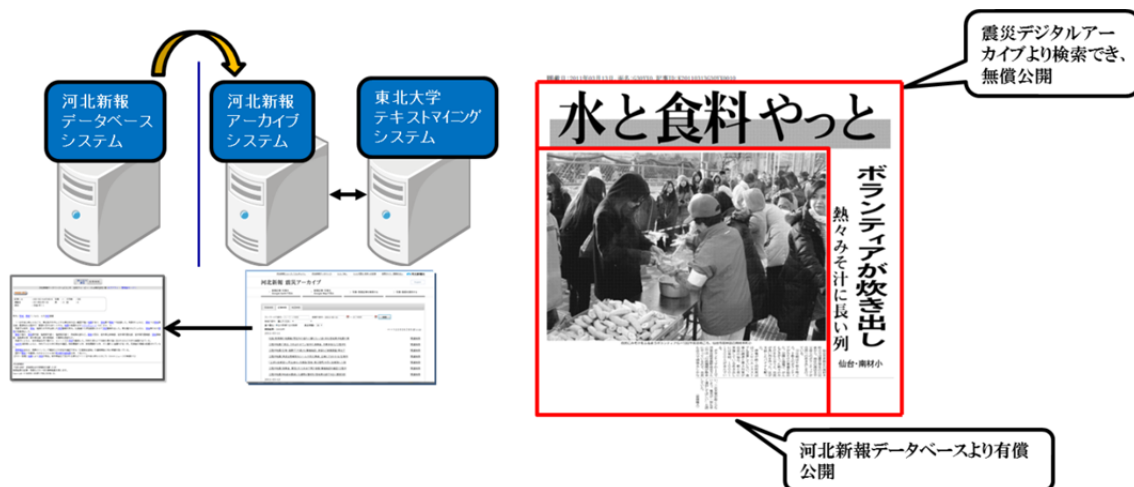


図 7-27 システム構成図

① ICT 環境（ハード）

宮城河北新報社では、後年運用を継続的に実施可能とするため、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としました。



【サーバスペック】

CPU	XeonE5-2609 (4) 2.40GHz-1066MHz 2CPU (※)
メモリ	32GB
ディスク	10TB

※：仮想 CPU 数として 32CPU

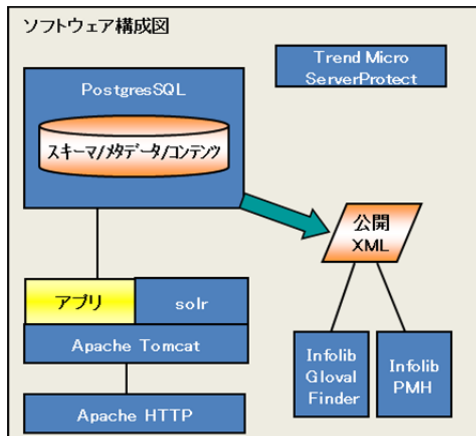
【河北新報社仮想サーバ割り当て】

仮想 CPU	8CPU
メモリ	11GB
ディスク	1.5TB

図 7-28 ハードウェア構成図

② ICT 環境 (ソフト)

後年運用において、システム保守を依頼する東北大学のシステムと同様、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、NDL 東日本大震災アーカイブとの連携機能については、実績のある商用システムを導入することとしました。



【オープンソース一覧】

RHEL6
Apache Tomcat 6
Apache HTTP 2
PostGreSQL 9
PostGreSQL 8
Seasar2
Solr 3

【商用製品】

Infolib Glocal Finder
Infolib PMH

図 7-29 ソフトウェア構成

③ 特徴とする機能

➤ 新聞記事と写真のつながり

河北新報社では、アーカイブシステムは、検索したコンテンツにメタデータであるキーワード、掲載日、地域で関連する新聞記事や写真を検索できるシステムとしました。

河北新報 震災アーカイブ

キーワード 「津波」 検索した結果

地震の写真

地震の記事

写真

新聞記事見出し

図 7-30 新聞記事と写真のつながり

➤ GoogleMAPによるコンテンツ数表示

収集したコンテンツで、住所情報を持っているコンテンツは、GoogleMAPに件数表示を行いました。表示は、市区町村ごとに件数を集計し、表示することとしました。

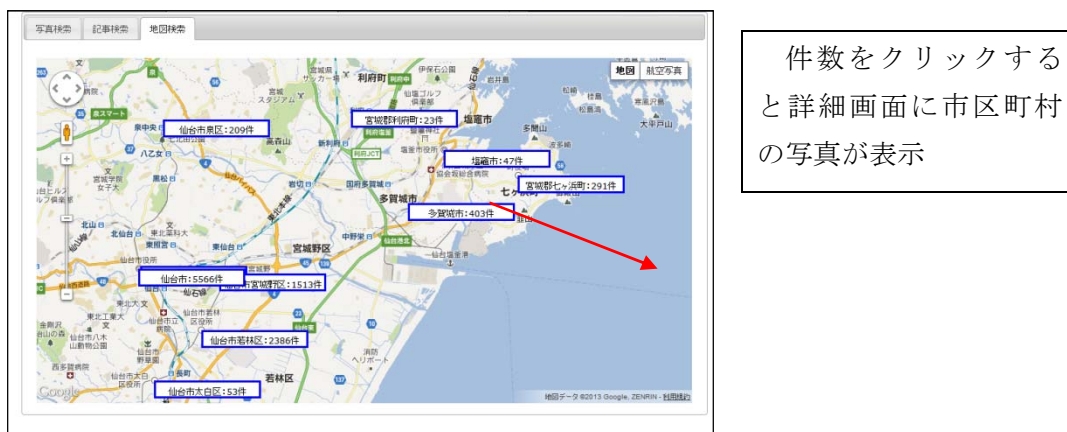


図 7-31 地図によるコンテンツ数表示

➤ GoogleEarthによる表示機能

コンテンツの利活用のアイデアとして、GoogleEarth、MAPへコンテンツを表示することとしました。

東日本大震災は、津波の影響もあり、沿岸部の地域事に事象が違うことが特徴であり、新聞記事や写真を選定し、地図上に表現し、どのようなことが起きているのかを伝える画面を作成しました。



図 7-32 Google earth による表示

(c) システム構築における課題

① コンテンツの地図への展開

コンテンツ数が多いと地図にピンを指してもピンだらけとなり、見るに堪えない状況となることが確認できました。また、コンテンツの住所情報は、コンテンツに GPS 情報が無いことから、手作業で入力しており、町字まで入力することが困難でした。そのため、市区町村ごとの件数表示としました。しかし、件数表示が重なりあって、見づらいつの意見や地図は、拡大しないと使えないという意見もあり、今後の検討課題となっています。

(d) システム運用

① 運用体制

アーカイブ運用体制は、以下のとおりとなります。

表 7-16 システム運用体制

役割	団体
運用主体	河北新報社
コンテンツ	河北新報社
システム	東北大学

② 役割分担

アーカイブ運用における役割分担は、以下のとおりとなります。

表 7-17 役割分担

No.	組織	担当
1	河北新報社	・運用主体 ・コンテンツ収集
2	東北大学	・システム保守・運用

③ システム設置場所

サーバは、東北大学のサーバ室へ設置し、サーバ及びネットワークの運用ポリシーは東北大学のポリシーに従うこととしました。

(e) システム運用における課題

東北大学へシステム保守・運用を委託することとし、運用費用を削減することとしました。また、公開に伴う問い合わせ先として、河北新報社デジタル編集部が担当し、今後も東北大学と密接な連携を取りながら、問題解決を共有していくこととしました。

(6) 福島県における実証調査

(a) 体制

アーカイブシステムの運用、保守等の体制については以下の2機関で対応します。

表 7-18 システム構築体制

奈良先端科学技術大学院大学	以下の設備等の提供と維持管理 ・ハードウェア機器の設置場所 ・電源・空調等 ・ネットワーク環境 サーバ起動・停止作業 バックアップテープ状況の確認
システム構築会社	アーカイブサイトで動作する一切のアプリケーションの運用・保守 アーカイブサイトの問い合わせ対応

(b) システム構築

福島プロジェクトで構築したシステムの概要図は、以下のとおりとなります。

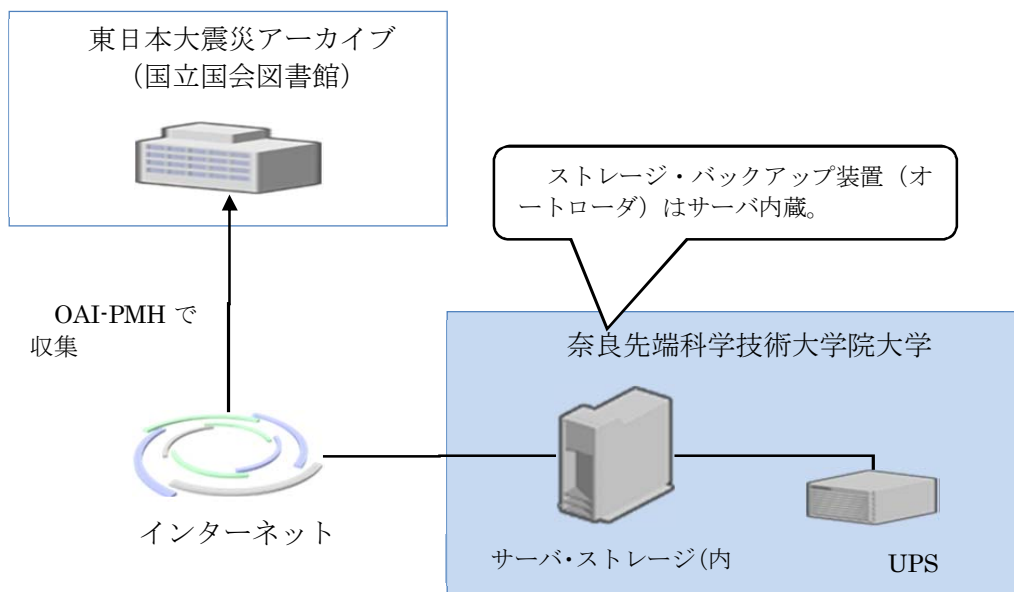
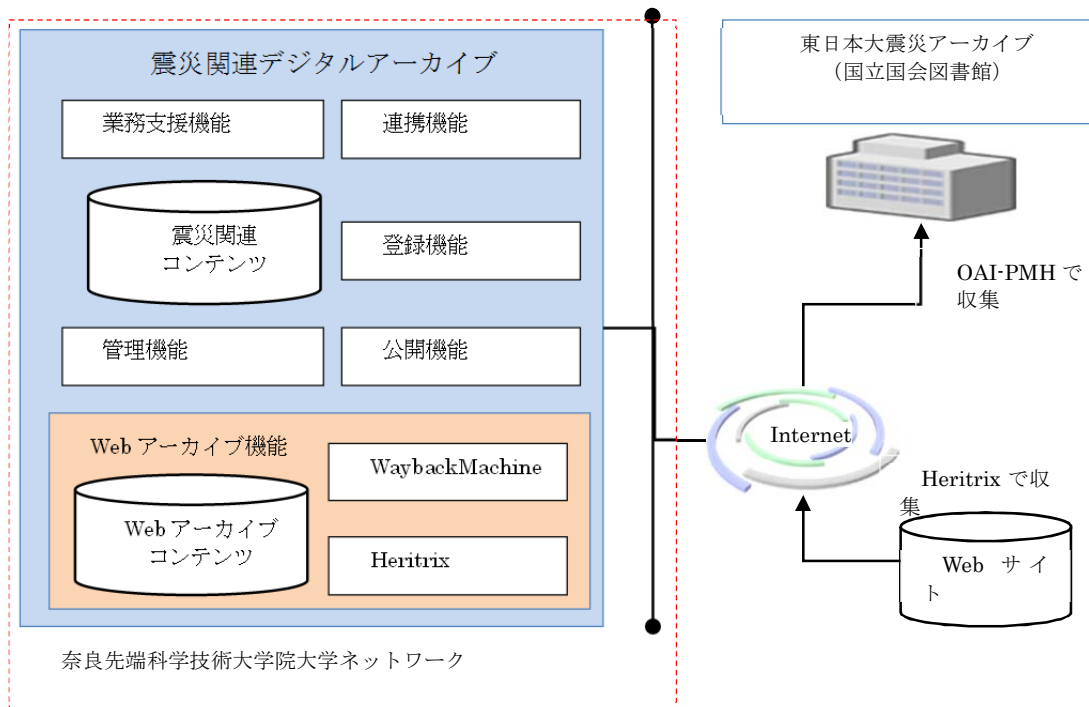


図 7-33 システム構成図

① ICT 環境 (ハード)

福島プロジェクトでは行政文書を含め、散逸する可能性の高い震災関連コンテンツ、おもにデジタルコンテンツを保存する前提でハードウェア環境を構築する必要がありました。また自治体のサーバがディスク容量の問題で、格納しているデータを削除せざるを得ない状況があるため、格納庫としての位置付けも想定しました。

また福島では ICT 環境を奈良先端科学技術大学院大学に設置することで遠隔地運用を

行いました。これにより、今後新たな災害が東北地区で発生した場合でも、同時に被害にあう確率が低くなります。

また奈良先端科学技術大学院大学の ICT 環境は「曼荼羅システム」と呼ばれ、日本最高峰の情報処理環境であり、信頼性に優れています。

※曼荼羅システム

<http://itcw3.naist.jp/itc-intro/itc-j.html>

② ICT 環境（ソフト）

収集した震災関連コンテンツを登録、公開、管理する機能、及び NDL 連携機能は構築期間が短期間でもあるため、実績のあるデジタルアーカイブシステム構築用の商用ミドルウェアを利用しました。NDL 連携機能においては Z39.50、SRU/SRW、OpenSearch、OAI-PMH といった相互連携プロトコルを標準で実装していることも本ミドルウェアを選定した理由です。

Web アーカイブを実現するために実績のあるフリーソフトウェア Wayback Machine を使用しました。本ソフトウェアは Internet Archive (<http://archive.org/index.php>) においても使用されているものです。

Wayback Machine は保存された WARC ファイル (Web アーカイブ用ファイルフォーマット) を検索、閲覧できる仕組みですが、WARC ファイルを作成する機能を保持していません。このため、同じく Internet Archive で実績のある Web サイトクロールソフトウェア Heritrix を合わせて導入しました。

構築した本プロジェクトのアーカイブサイトのトップページを構成するために、MovableType OpenSource 版を導入しました。本ソフトウェアは日本国内で最も普及しているといわれている商用 CMS ソフトウェアのオープンソース版です。

ハードウェア機器の監視ソフトウェアとしてオープンソースソフトの Zabbix を採用しました。Zabbix はリモート環境のハードウェア機器リソースをネットワーク越しに監視することができるため、遠隔地にハードウェア機器を設置している福島プロジェクトでの使用に適しています。

表 7-19 使用ソフトウェア一覧

Redhat Enterprise Linux 6 Server
Apache 2.2.22
Apache Tomcat 6.0.13
PostgreSQL 8.3.13
InfoLib XX
Wayback Machine 1.6.0
Heritrix 3.1.1
Movable Type 5.2.3
MySQL 5.1.66
Zabbix 1.8.15
F-Secure 9.20

③ 特徴となる機能

➤ 一次コンテンツ(ドキュメントファイル)から文字列抽出

福島プロジェクトでは自治体から収集した文書資料が数多く存在します。これらの資料はメタデータのみならず、資料本文の情報をテキスト抽出し、検索可能としました。対象となるファイル形式は PDF 形式 (テキスト埋め込み型) としました。ただし Office 形式等でも配信用に PDF 変換できたファイルについては、同様にテキスト抽出対象となります。

ドキュメントファイルからのテキスト抽出を行うことで、検索データのインデクス作成時間が本文抽出を行わない場合に比べて、長くなるという課題があります。

➤ 関連コンテンツ表示機能

検索結果のコンテンツ表示画面から関連コンテンツ一覧を表示させる機能を設置しました。関連コンテンツは表示中のメタデータの **subject** 項目 (分類) 値を使用して、データベースを検索します。検索結果は非同期で表示するようにしました。これにより利用者には内容が似ているコンテンツを辿ってサイト閲覧が可能となることを目指しています。

➤ Web アーカイブ機能

Wayback Machine と Heritrix を組み合わせることで Web アーカイブの仕組みを構築しました。Heritrix で WWW 上の Web サイトを収集することで、随時更新されている Web サイトのスナップショットを保存していくことができます。本機能を運用することで、実サイトでは消去されてしまった内容を保存し、時系列で閲覧することができるようになります。

(c) システム運用

① 運用体制

アーカイブ運用体制は、以下のとおりとなります。

表 7-20 システム運用体制

役割	団体
運用主体	東日本大震災アーカイブ福島協議会
コンテンツ	東日本大震災アーカイブ福島協議会
システム	システム運用会社

② 役割分担

アーカイブ運用における役割分担は、以下のとおりとなります。

表 7-21 役割分担

No.	組織	担当
1	後年運用協議会（準備会）	・ 運用主体 ・ 継続的に運用可能な運用主体の検討
2	奈良先端科学技術大学院大学	・ サーバ監視
3	システム運用会社	・ システム運用

③ システム設置場所

サーバは、奈良先端科学技術大学院大学のサーバ室へ設置し、サーバ及びネットワークの運用ポリシーは同大学のポリシーに従うこととしました。

(d) システム運用における課題

今後のコンテンツの収集に関しては、紙文書が多くなることが予想されます。紙文書の収集・保存については、既に運用実証調査でデジタル化、メタデータ付与のためのシステムが整備されており、比較的安価にアーカイブ可能と考えられます。今後は、継続的にメタデータのボキャブラリ整備、マニュアルの改訂等を行い、より効率的な運用を行えるようにしていくことが必要となります。

今後の運用では、紙文書についてはデジタル化後、提供者に返却又は、デジタル化後に廃棄処分とすることを前提に収集する予定です。ただし、運用上、保管が必要と考えられる文書がある場合には、県立図書館等に寄贈することを検討します。