

「東日本大震災アーカイブ」基盤構築プロジェクト ラウンドテーブル

技術 WG 議事要旨

- 1 日時 平成 25 年 3 月 21 日 (木) 10:30~12:40
- 2 場所 三菱総合研究所 会議室(大会議室 A)
- 3 出席者 (敬称略):
 - (1) 技術 WG 構成員
岩爪構成員、樫村構成員、嘉村構成員、神崎構成員、高野構成員(座長)、研谷構成員、森山構成員
 - (2) 運用実証・ポータル開発事業者
青木部長 (凸版印刷)、日高主任 (インフォコム)、岩田課長代理 (NTT データ)
 - (3) オブザーバ
 - ①総務省
白石課長補佐
 - ②国立国会図書館
河合次世代システム開発研究室長
 - (4) 事務局 (三菱総合研究所)
前田

4 議事内容

- (ア) 開催にあたって
- (イ) 東日本大震災アーカイブ基盤構築プロジェクトの状況
- (ウ) 運用実証事業結果について
- (エ) ガイドライン案について
- (オ) その他

5 議事

【議題 1 : 開催にあたって】

- 総務省白石課長補佐及び国立国会図書館河合室長より、WG 開催にあたっての挨拶があった。

【議題 2 : 東日本大震災アーカイブ基盤構築プロジェクトの状況】

- 事務局より、資料①「東日本大震災アーカイブ」ソフトウェアの開発状況を説明。

主な意見は以下の通り。

- 構成員 (高野座長)
アクセス数はどの程度か。
- NTT データ (岩田課長代理)
3 月 7 日のオープン当日は 12 万件、オープンした週については、平均 1 日 5 万件程度のページビューであった。
- 構成員 (高野座長)
ユニークユーザはどの程度であったか。

- NTT データ（岩田課長代理）
ユニークユーザ数までは取得できていない。コンテンツへのアクセス傾向としては、TOP の写真及び動画へのアクセスが多かった。

【議題 3：運用実証事業結果について】

- 凸版印刷より、資料②「運用実証事業報告書」を説明。

主な意見は以下の通り。

- 構成員（森山構成員）
ガイドラインを見たが、事例編の中で東北大においてソーシャルタグを付与した取組があったとの記載があるが、取組状況はいかがか。
- インフォコム（日高主任）
ソーシャルタグを付与する仕組みは用意しているが、入れられたものをそのまま公開するのか、査閲するか等の検討が終わっていないため、まだ運用できていない。
- 構成員（森山構成員）
ソーシャルタグ以外に、参加型の仕組みは用意しているか。
- インフォコム（日高主任）
例えば河北新報においてはコンテンツ投稿の仕組みを用意しており、かなりのコンテンツが集まっている状況である。3 末のオープンから 5,000 件程度投稿していただいている。そういった仕組みを用意しているからか、高齢の方については、システムからの投稿だけではなく、コンテンツ自体をお持ちいただけたというケースも存在した。
- 構成員（森山構成員）
双方向型あるいは参加型の仕組みを用いた実証は、今回の運用実証事業では行わないのか。
- 凸版印刷（青木部長）
運用のルールが詰めきれていないため、後年運用で検討させていただく。河北新報については、投稿の仕組みを設けたところである。
- 構成員（研谷構成員）
河北新報のプロジェクトにおいて、既存の新聞記事や写真と震災アーカイブが連携しているようだが、どのように連携しているのか。地方紙では、新聞記事データをどうビジネスにつなげていくのかが課題になっていると聞いている。今回の河北新報の震災アーカイブへの取組が、地方紙における新聞記事の DB 化や地域災害が起きた際のアーカイブ、アーカイブのビジネス化のケーススタディになりえるのかという点に関心がある。この DB はどんなもので、また震災以前からあるものなのか。
- インフォコム（日高主任）
河北新報社については、以前から DB を構築していた。そこから新聞記事の見出しデータを抽出して震災アーカイブの DB に流し込んで、テキストマイニング等を行いアーカイブで見られるようにしている。
- 構成員（研谷構成員）
H.P.で公開されているものか。
- 凸版印刷（青木部長）
有料で公開していたものとなる。
- 構成員（研谷構成員）
新聞紙面はどのように DB 化しているのか。

- 構成員（高野座長）
記事のイメージとテキストを結び付けて両方保存している。
- 構成員（岩爪構成員）
青森プロジェクトにおいては後年運用について年間数百万円かかるとの記載があるが、他のプロジェクトでの状況はいかがか。また、岩手大学のプロジェクトではサーバの構成やスペックが記載されているが、他のプロジェクトではどの程度のHWを使っているのか。
- 凸版印刷（青木部長）
宮城の2プロジェクトは河北新報、東北大学が事業としているため、コストは運用主体が主として対応している。他のプロジェクトはコスト算出も含めて検討中である。
- インフォコム（日高主任）
システムスペックは各プロジェクトにおいて、コンテンツの収集目標件数を設定していたため、それから見積もりを行った。福島においては1万件を目標としていたが、3万件程度のコンテンツが集まった。しかし、拡張性を考慮したスペックとしていたので現時点で問題は起きていない。各プロジェクトとも、基本的には岩手のスペックに類するものとなっている。
- 構成員（岩爪構成員）
そこまでハイスペックなものではないということか。
- インフォコム（日高主任）
スペックとして重要な検討事項はディスク容量ではないかと考えている。
- 構成員（嘉村構成員）
ディスクとサーバは分けているのか。
- インフォコム（日高主任）
青森、岩手と福島は同一、東北大学は仮想化しているので別サーバとなる。
- 構成員（嘉村構成員）
もしストレージを利用している場合には、最近ではRAID6が普及してきているので、考慮した方がよい。
- 構成員（高野座長）
国立国会図書館とメタデータを連携していても、コンテンツ自体を連携していないもの等について、サーバがクラッシュした場合に今回の事業の結果がなくなってしまうことが考えられる。現時点でコンテンツを保護する仕組み等について、既に対応されているのか。対応されていないのであれば、国立国会図書館に預けるという運用があってもよいのではないか。
- インフォコム（日高主任）
東北大学は遠隔地バックアップを利用している。河北新報も東北大学と同様のものを利用している。青森、岩手、福島についてはディスク及びテープにバックアップを取得しているが、保管場所を検討しているところである。
- 構成員（樫村構成員）
大学でサーバを管理する場合、組織内でのアーカイブの事業の位置づけによって事業がなくなってしまうことがありえる。
- 凸版印刷（青木部長）
大学で管理しているものについては、県レベルで図書館へ移管できないかということを検討している。基本的には継続性のある公共性の高い団体にて、管理していただけるよう動いている。
- 構成員（樫村構成員）

大学の場合、研究者が移り変わるのが自治体に比べると早いと思われるので、なるべく早く対応されたほうがよい。

○ 構成員（高野座長）

図書館は電子的な情報の収集や公開に関しては、得意でない面が見受けられるため、図書館に預けるにしても、その点を踏まえた検討を行った上で行うべきである。また今回の事業で後継団体を作るという話も記載されているが、各県がこの事業に関わってきた人を中心に運用していくというのは自然ではあるが、今回はトライアルで行っており、その結果をガイドラインに取りまとめ、多くの人にアーカイブを構築してほしいというのが本来の目的であったと思われる。そのため実証プロジェクトに関わった人達だけで組織を作ってしまうと、新規の団体が入りづらくなるのではないかと懸念している。後継団体を提案する際には、新規のNPO等が参入できるような仕組みを検討しているのか。

○ 凸版印刷（青木部長）

現状対応している団体に渡すのは暫定的なものであり、団体においては後年運用の中で検討していただくことになる。現時点では時間的に間に合わなかった。

○ 構成員（高野座長）

新規団体からの情報の受入比率は3割を超えなければならない、というような仕組みがあってもよいのではないかと懸念している。

○ 構成員（森山構成員）

図書館はデジタルアーカイブのような事業が、組織に組み込まれていないのが実情である。組み込まれた図書館であっても、継続的に引き継がれる保証がなく、先細りになるケースがある。また自治体は財政難で費用対効果等の指標を設けており、自治体に預けても必ずしも安心できないと思われる。

○ 構成員（高野座長）

図書館の本来業務にデジタルアーカイブは含まれないため、対応してもらうのは難しいと考える。デジタルアーカイブを運用することをミッションとした組織を作らせ、そこを中心に協議会等を組織して、定量的な利用計画を満たす提案を行ったところに任せるとよいのではないかと懸念している。

○ 構成員（研谷構成員）

有償か無償かの切り分けはどのように分けられているのか。

○ インフォコム（日高主任）

新聞記事は全て有償である。収集されたコンテンツについては無償となっている。

○ 構成員（高野座長）

河北新報はもともとそういったコミュニティサイトを持っており、提供を依頼していた。それと自社の記事を結びつけ、経営的な判断を行っているのではないかと懸念している。

○ 構成員（研谷構成員）

アーカイブの運用にあたって、全て無償とした場合に援助や補助金によって運用することになるが、河北新報の事例は、有償にした場合に利益でどの程度運用費をカバーできるのか、お金を払う人がどの程度いるのかということの参考事例になるのではないかと懸念している。今後実証的に見ていくことも必要ではないかと懸念している。有償のコンテンツをどれくらいの人を買ったのか、可能であれば今後河北新報から報告をいただきたい。災害が起きた際には地方紙がこういった取組を行うと思われるので、継続性という観点から、地方紙との連携を議論すべきではないかと懸念している。

○ 構成員（嘉村構成員）

5つのプロジェクトは、今後、統合するような予定はあるのか。

- 凸版印刷（青木部長）
現時点ではそれぞれで運用する想定であるが、情報共有の機会を設ける想定である。
- 構成員（森山構成員）
組織が別々であるとずれが生じてしまう。それはキーワードや分類にしてもそうだが、それによって、検索できないということが生じてしまうことがある。また、キーワードや分類で使用する言葉は時代と共に変わっていくため、キーワードや分類のメンテナンスを五つのプロジェクトで連携して行う必要もあるのではないかな。

【議題4：ガイドライン案について】

- 事務局より、資料③「ガイドライン案一式」（はじめに～第4章まで）を説明。

主な意見は以下の通り。

- 構成員（嘉村構成員）
P.92 にクラウドの利用について記載されているが、日本の会社であってもサーバの設置国によって、問題が発生する可能性がある点を記載しておいた方がよいのではないかな。
- 構成員（森山構成員）
東日本大震災において公的機関がソーシャルメディアを使って発信した情報を東日本大震災アーカイブにアーカイブしてはいいかがか。救援等で大きな役割を果たした事例もある。
- 事務局（前田）
ソーシャルメディアについてはP.47に記載があるため、そちらに公的機関を追加する。
- 構成員（森山構成員）
被災地でないために自分たちは関係ないと思っている自治体が多いかもしれないが、支援自治体として震災直後に現地へ入り、活動実績を記録し保持している場合が実は多い。そういった支援自治体が行った活動の記録も対象となることを強調してもよいのではないかな。
- 構成員（研谷構成員）
地元のケーブルTVや自衛隊等について、データのDVDがパッケージとして出ている。それらは表3-1、表3-2いずれにも記載がないと思われるので、どこかに記載されてはいいかがか。
- 構成員（森山構成員）
双方向型、参加型の仕組みが注目されている社会の中で、震災関連のアーカイブでも、そうした機能を設けられると、さまざまな方の関心を高めていく上で有意義ではないかな。
- 構成員（高野座長）
アーカイブを利用者に開放する仕組みがないと、よいコンテンツがあってもだんだん利用されなくなっていってしまう。
- 構成員（森山構成員）
新しいコンテンツが入ってこない、利用は右肩下がりとなってしまう。
- 総務省（白石課長補佐）
今後の課題として検討する。
- 事務局より、資料③「ガイドライン案一式」（第5～第7章まで）を説明。

- 構成員（神崎構成員）
第5章についてタイトルはメタデータを利用するためのガイドというよりは付与するためのガイドラインではないか。また、冒頭のメタデータの説明に関する部分において、データについてのデータという意味が分かりづらい。データという用語はコンテンツに改めて、「コンテンツそれ自身を説明するデータ」といった表現にすべきである。
- 構成員（神崎構成員）
P.100にて使用する文字コードの問題もあると記載されているが、対応方法が記載されていないため、読む側が迷ってしまう。削除するか、あるいは「文字コードは原則としてUTF-8とする」と記載してもよいのではないか。また、参考資料6にてマッピングを行っているとは記載されているが、意図が不明確であり、マッピングしたということを本文に記載する必要はないのではないか。参考資料6自体はあっても構わない。
- 構成員（神崎構成員）
メタデータスキーマを設定する必要があると書かれているが、メタデータスキーマという言葉は読者にとってハードルが高い。また、ここまでやったのにさらに作業が必要なのかと、負担感を与えることになってしまう。スキーマという用語は事例編にもあまり出てきていない。ルールを決めておく必要があるということだけ記載すればよいのではないか。事例を参照しながら説明してもよいのではないか。
- 構成員（研谷構成員）
全体的な話として、本ガイドラインは全てを1冊で開示するのか。
- 総務省（白石課長補佐）
基本的にはまとめるが、章毎に見られるようにするか、web化する等も含めて見せ方を検討する必要があると認識している。冊子とすることは特に検討していない。
- 構成員（樫村構成員）
メタデータに関して重要だと記載されているが、それがアーカイブにどのように活かされるのかの記載がまるでないため、記載された方がよいのではないか。また、デジタルアーカイブのイメージは人によって異なるため、その点も定義しておくともよいのではないか。
- 総務省（白石課長補佐）
P.98に記載しているがそれでは不足しているか。
- 構成員（樫村構成員）
デジタルアーカイブとメタデータの関連性が記載されている必要があるのではないか。デジタルアーカイブとは何かという記事を参照情報として紹介してもよい。
- 構成員（高野座長）
P.106に「個人名は全て不明とする」という記載があるが、基本的に残せる情報は全て残した方がよいと考えており、事例としてはあまりふさわしくないのではないか。こういった事例をガイドラインに載せることにより、その事例に固定化されてしまう懸念がある。
- インフォコム（日高主任）
個人情報の取得にあたるため、慎重に対応した。しかし、継続的な情報の管理という意味では不十分な点もあった。
- 構成員（神崎構成員）

参考資料 7 の P.20 にその点に関する注意点も記載されているが、こういった情報はガイドライン本文に記載すべきではないか。メタデータ入力的前提として必要な情報があるはずであり、そういった内容はガイドライン本文に記載すべきである。参考資料 7 を見ると、こういった前提となる情報と入力する際に必要な情報、技術的な情報が混同して記載されているため、扱いを分けるべきではないか。

- 構成員（嘉村構成員）
参考資料 4 について、メタデータスキーマの説明ではない。
- 構成員（高野座長）
参考資料 4 は削除してもよいのではないか。
- 構成員（研谷構成員）
全体的なフローと詳細な記載や、ベストプラクティスとそうでないものがまとめて記載されており、1冊になってしまうと読みづらいものになってしまうのではないか。
- 総務省（白石課長補佐）
冒頭の部分でもう少し説明させていただくようにする。
- 構成員（研谷構成員）
読者によって読みたい部分は異なるため、読みたい人によってどこを読めばよいか分かるようなものがよいのではないか。
- 構成員（高野座長）
目的別にどこを見ればよいのかが記載されているとよいのではないか。
- 構成員（森山構成員）
詳細版だけでなく、1枚で図示されているような簡易版がないと、容易には入り込めない。
- 総務省（白石課長補佐）
入門又は概要版のようなものを検討する。
- 構成員（高野座長）
自治体が事例とするようなガイドラインとするのであれば、ひなぎくとの連携やサブシステムについて強調しておく、ひなぎくが活用されるようになるのではないか。
- 構成員（研谷構成員）
震災アーカイブプロジェクトを行うにあたって、プロジェクトに携わる方それぞれの役割があると思われるので、それらをもう少し洗い出して、それをモデルとして示すのが重要ではないか。そうすることで次の災害があった際に、こういったプロジェクトがすぐに立ち上がるようになるのではないか。

【議題 5：その他】

主な意見は以下の通り。

- 事務局（前田）
利活用 WG では有志の皆様で ML を使ってハッカソンやアイデアソンを実施したり、ガイドラインを web 化したりということを検討している。
- 構成員（高野座長）
ML ベースで対応されるのであれば、問題がなければこちらの WG も追加していただ

けるとありがたい。

○ 事務局（前田）

ガイドラインと今後のことについては、ラウンドテーブルに諮り決定事項を皆様にお伝えする。

以 上