

「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 24 年 12 月 27 日 (木) 11:00~12:35
- 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
安藤構成員、岩爪構成員、大向構成員、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
岡野課長 (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
尾川課長補佐、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、白石室員
 - (4) 事務局 (三菱総合研究所)
前田、松尾
- 4 議事内容
 - (ア) 第 1 回ワーキングの議事要旨及び指摘事項への対応状況について
 - (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と今後の予定について
 - (ウ) デジタルアーカイブ構築・運用に関する実証調査の現状と今後の予定について
 - (エ) ガイドライン案について
 - (オ) その他
- 5 議事
 - 【議題 1 : 第 1 回ワーキングの議事要旨及び指摘事項への対応状況について】
 - 事務局より、資料①「東日本大震災アーカイブ」第 1 回技術・利活用合同 WG 議事要旨」及び資料②「第 1 回WG 指摘事項への対応方針・対応状況」を説明。
 - 【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と今後の予定について】
 - 事務局より、資料③「東日本大震災アーカイブ」ソフトウェアの開発状況と今後の予定について」を説明。
 - 【議題 3 : デジタルアーカイブ構築・運用に関する実証調査の現状と今後の予定について】
 - 凸版印刷より、資料④「東日本大震災アーカイブ」基盤構築事業 デジタルアーカイブ構築・運用に関する実証調査中間報告」を説明。

主な意見は以下の通り。

《資料 1 から資料 5 について》

- 構成員（嘉村構成員）

実証実験の中でヒアリングやワークショップを実施するとのことだが、オーラルヒストリーを収集する際に、いろいろなやり方があると思われる。全プロジェクトで統一的な方法で実施することを想定しているのか、各プロジェクトで都度検討することとしているのか。
- 凸版印刷（岡野課長）

岩手プロジェクトでワークショップの実施を予定している。そこで現地のNPOの方等が集まっていたいて、どこにどんなコンテンツがあるか等のヒアリングを行うことで、コンテンツを収集しながらオーラルヒストリー収集のための取材へのモチベーションも高められるのではないかと期待している。他の地域もコンテンツ収集のためのヒアリング自体が取材につながっている。
- 構成員（高野座長）

ヒアリングやワークショップそのものの音声データ等がアーカイブに収集されるとよいと思われるが、コンテンツの表に記載がなかったための指摘と思われる。
- 凸版印刷（岡野課長）

ヒアリングとコンテンツ収集を同時に進めていくことを検討しているので、結果は改めてご報告させていただきたい。
- 構成員（神崎構成員）

岩手プロジェクトのP.12地域の表記方法については、どのように記録しているのか。市町村コード等で他の地域でも展開できるようなものを使っているのか、それとも入力者が個別に名づけ入力しているのか。
- 凸版印刷（岡野課長）

現時点ではコード化まではしていない。地域区分について、岩手プロジェクトの実績も踏まえながら、各プロジェクトでやりやすいように検討する。
- 構成員（神崎構成員）

市町村コード等を出発点と決定した上で、進めるのがよいのではないか。
- 構成員（森山構成員）

標準化しないと改めてマッピングする必要があるため、標準化しておいた方がよい。
- インフォコム（日高主任）

岩手で記載している資料区分や住所は、標準化したいのはもっともであるが、他の地域ではこういった情報が入れないものもある。各プロジェクトにどのような情報が集まってくるかがわからなかったため、標準化するためのカテゴリは各プロジェクトに任せて、あがってきた情報を統括事務局でまとめようとしている。
- 構成員（高野座長）

それがまずいのではないかと指摘である。情報をつける場合には標準化したものをつけた方がよい。標準化したものを事務局から配布する必要があるのではないか。標準化ボキャブラリーは用意しようと思えば例はある。
- 構成員（森山構成員）

データの形式については、対応するビューアーや長期保存等も含めて考慮する必要があるのではないか。
- 凸版印刷（岡野課長）

検討させていただいているところである。長期保存といった場合、10年、30年、100年先のどのあたりを念頭に設定するのかという問題もある。

- 構成員（森山構成員）

公開用の形式でしか持っていないと、後々問題になることが多い。標準的なファイル形式は時代とともに変遷するので、将来、ファイル形式をマイグレーションせざるを得なくなることも考慮し、配信用ファイル形式のデータのほか、保存・編集用ファイル形式の元データも準備しておく必要がある。
- 構成員（樫村構成員）

各プロジェクトでデジタル化するプロセスは記載されているが、プロジェクト間で判断の基準や情報共有やデジタル化の方針・指針は共有されているのか。
- 凸版印刷（岡野課長）

基本的には、「知のデジタルアーカイブに関する研究会」のガイドラインを基本にして対応している。
- 構成員（樫村構成員）

プロジェクト間のクオリティにばらつきが出ないように、事務局として全体の方針や品質を決めた上で、対応しておく必要があるのではないかと。
- 構成員（高野座長）

電子書籍でも 600dpi が普通になってきている。何度もデジタル化できるものであればよいが今回しかデジタル化する機会がないものもあると思う。それらを含めて検討して後世に残せるものとして保存すべきではないかと。
- 凸版印刷（岡野課長）

その辺りは検討させていただいている。原則として印刷に耐えられることを判断基準としている。
- 構成員（樫村構成員）

画像のメタデータの記載の中で、景色の写真 96dpi というメタデータがついているものがあつた。経年資料や文書資料等をデジタル化する場合に dpi という概念は必要ではあるが、景色を撮った写真に対して dpi というメタデータがついていることに違和感があるが、その辺は問題ないかと。
- 凸版印刷（岡野課長）

検討する。
- 構成員（樫村構成員）

必要性や根拠がない数字はないほうがいいのではないかと。
- 構成員（研谷構成員）

宮城で google picasa を使っているとの記載があるが、宮城だけで使っているのか。他のプロジェクトではどのようなツールを使っているのか。また、google picasa を採用した理由として生産性をあげるとの記載があるがどのようにメタデータをつけているのか教えていただきたい。
- インフォコム（日高主任）

採用した理由は、宮城プロジェクトはみちのく震録伝のプロジェクトから継続しており、みちのく震録伝で google picasa を利用していた実績があつたためである。メタデータの付与については写真のデータが多く、google picasa に登録することで gio タグがついてくるため、それをダウンロードすることでメタデータを付与している。宮城の場合 12 万件のデータがあるため効率化する必要があり、この方法をとっている。

他のプロジェクトでどのようにメタデータを付与するかは議論になっているところである。

- 構成員（研谷構成員）
ツールの採用について選定・評価等もガイドラインに反映されるとよい。
- 構成員（高野座長）
google picasa を使ってどのような点が便利であったのか、追加してほしい機能等があればガイドラインに反映していただきたい。
- 構成員（安藤構成員）
後年度の運用に不安を残した記載となっているようだが、この辺りに係る方策については検討されているのか。
- 凸版印刷（岡野課長）
後年運用については、検討を開始しているところである。
- 構成員（安藤構成員）
システム開発にあたって、アウトプットが見えていないのではないかと。利活用の目的をもって開発するシステムを検討すべきである。
- 構成員（高野座長）
協力している組織が、実証が終わったタイミングでどのようなことがしたいのか、どの程度の費用等がかかるのか等を、実証が終わった段階で分かるようにしていただきたい。
- 構成員（研谷構成員）
前回の WG でも指摘したが、アーカイブを継続できない場合の対応や東日本大震災アーカイブで継承できるような対応も含めて検討されるのがよい。
- 構成員（高野座長）
コンテンツ収集について、想定件数、実績件数の記載があるが、理想として集めたらよいという目標件数も入っていると、どのくらい仕事が残っているか等が明確になるのではないかと。
- 凸版印刷（岡野課長）
そもそもどのくらいのコンテンツが確保できるか全く手さぐりであったため、岩手では各自治体にどの位のコンテンツがあるのかアンケートをとっている。

《資料 6 から資料 7 について》

- 構成員（神崎構成員）
何がくるか分からないためかもしれないが、メタデータスキーマの中で権利の記述 (dcterms:rights 権利に関する情報(リテラル))が文字列となっているが、API を使う側から考えると自由に使えるものだけを利用したいという要望は高いと思うので、それが簡単に分かるような記載、例えばクリエイティブコモンズなど標準化されている権利を利用し、CC ライセンスの URI を記述するようにすべきではないかと。
- 国立国会図書館（白石様）
本日配布している連携メタデータスキーマが古い版であると見受けられる。最新版では dcterms:rights を構造化表記とし rdf:value と dcterms:description を用意し、前者に定型的な文字列、後者に自由な文字列を入力できるようにし、さらに dcterms:license を追加し、ここにクリエイティブコモンズ等の URI の情報を入れられるようにしている。どのような利用条件や権利条件が存在し、具体的にどのような値がくるか NDL では想定できなかったため、値を入れられる場所をとりあえず用意しているという状況である。
- 構成員（高野座長）
具体的に情報共有していただき、なるべく標準 URI を入力し、自由記載欄に記載が集中しないようにしていただきたい。

《資料 8 から資料 9 について》

- 構成員（高野座長）
制度・運用検討委員会の中では、権利関係についての結論を得るのは難しいということか。
- 凸版印刷（岡野課長）
第 1 回目の結論としてはそうなる。
- 構成員（大向構成員）
行政文書等、権利関係について統一的に決着をつけることができるドメインもあるのではないか。その点も結論は出さないのか。
- 凸版印刷（岡野課長）
現時点ではそのような結論は出ていない。
- 構成員（高野座長）
資料 5 では個別に対応されているようなので、その辺を含めてご対応を検討されてはどうかか。
- 構成員（神崎構成員）
権利関係については権利の保護に力点が置かれすぎている気がする。もっと広く利用できるような形にすべきではないか。
- 構成員（高野座長）
震災アーカイブ由来のデータであることを記載すればよいというような許諾の取り方もある。許諾を求める際に、利用の自由度が高いものを最初に求めて、それで拒否されたら自由度を低くするという方法もいいのではないか。
- 構成員（安藤構成員）
開発途中のシステムで館内限定データというフラグがあったが、本資料ではどこが該当するのか。
- 国立国会図書館（白石様）
連携メタデータスキーマの項番 R87、88、89 が該当する。NDL 東日本大震災アーカイブメタデータスキーマでは、デジタル化資料の NDL 館内限定公開等当館独自の利用制御をする必要があるため、別の項目で利用条件を管理している。運用モデル実証側からメタデータを受け取る際は、R87、88、89 の値を基に、NDL 側の利用条件の値に変換して格納する。
- NTT データ（岩田課長代理）
配布した連携メタ一覧は運用実証から東日本大震災アーカイブに提供されるメタデータスキーマであり、東日本大震災アーカイブシステムのメタデータはもう少し項目が追加されている。このメタデータ自体は原本として保持することになるが、震災のメタデータは変換したものを管理することになる。
- 構成員（大向構成員）
NDC が項目に含まれているが、現実的に本実証事業で取得可能なのか。
- 国立国会図書館（白石様）
モデル実証側で各県の図書館が震災文庫に取り組んでおり、場合によって入る可能性がある。
- 構成員（森山構成員）
NDC については、それ自体の継続的なメンテナンスの保証があるという点で、他に代

わるものは見当たらないが、一般の付与者にはなじみが薄いという問題はある。

- 構成員（大向構成員）

type を書くというところがあるが、type に網羅性があるのか等も検討の必要があるのではないか。
- 国立国会図書館（白石様）

資料種別については DC・NDL で定めている資料種別の範囲で付与する方向で運用モデル実証側と合意しているため、ある程度統制できると考えている。主題の分類については神戸大の震災文庫を参考に統制されたものをつけられないかを検討したが、現在運用モデル実証事業者側で検討しているところである。資料種別であればある程度統制することが可能であるが、主題については統制することは難しい。主題をどう付与するかについては各プロジェクトにお任せしているのが現状である。
- 構成員（高野座長）

集まってきたものを見て統制できるようであれば統制をお願いするようなフィードバックが必要ではないか。
- 構成員（大向構成員）

個々のプロジェクトを見ているとキーワードはカンマ区切りでデータが入るとの記載があるが、どこに入るのか。値が複数ある場合に、メタデータ要素を繰り返して値を入力するのか一つの要素内にカンマ区切りで入力するのか等といった、細部のルールが明確になっていないように見える。こういった細部の調整がメタデータ連携の際に重要になるのでは。
- 国立国会図書館（白石様）

dcterms:subject に入る。具体的には R32 が該当する。当館からは、基本的に一つの要素無いにカンマ区切りで入力していただくのではなく、要素を繰り返して入力してほしいと運用モデル実証側に要望している。ただし、運用モデル実証側には MLA 機関の職員のようにメタデータの作成スキルに長けた人員がいるわけではないうえ、メタデータの入力の手間もかかるため、入力について注文を付け過ぎないように留意している。
- 構成員（大向構成員）

検索をどのようにしたいのかにもよるが、メタデータはきっちり決まっているが、個々のワークフローが固まっていないように見えてしまう。
- 構成員（森山構成員）

入力記述形式がまちまちだと集約したときに問題が発生するということか。
- 構成員（大向構成員）

まちまちであるのであれば、それはそれでその項目を全文検索とすればよい。そうではなくキーワードを統制したいのであれば、ワークフロー側で統制しないといけない。方向を決定する必要がある。
- 構成員（高野座長）

統制してワークフローに載せるというのが、難しいという事例でもある。

【議題4：ガイドライン案について】

- 事務局より、資料⑤「ガイドライン案①②⑥（基礎編）」を説明。

主な意見は以下の通り。

- 事務局（前田）

構成や粒度についてご意見があればメーリングリストでご連絡いただければと思う。

作成できたものから送付させていただく。また、ワーキングについては全3回とさせていただいていたが、ガイドラインに係るワーキングを追加で開催させていただきたいと考えている。

- 構成員（高野座長）

2月中・下旬に1回と3月中旬に1回開催することになるか。フィードバックをお願いする。
- 構成員（樫村構成員）

デジタルデータ化のガイドラインについて様々な方法を示しているものになっているが、ガイドラインという名前が適切なのか疑問である。デジタル化には様々な方法があるため、ガイドラインというより参考書、ガイドといった位置づけにして、いろいろな方法を提供する今の内容でいいと思う。これはしてはいけないというものはまとめてもよいかもかもしれない。
- 構成員（大向構成員）

一方でベストプラクティスが記載されていた方がよいという観点もあるので、その辺りも切り分けて記載できるとよいのではないか。運用実証に基づくベストプラクティスプラス情報源というのがよい。
- 構成員（樫村構成員）

まとめ方として深入りすると色の管理やデータの真正性等を含めて考慮する必要があるため、データの多様性を考えるとガイドという形でまとめておいた方がよい。
- 構成員（嘉村構成員）

デジタルデータ化の段階でも権利問題を含めて検討する必要があることが記載されているのがよいのではないか。デジタル化する際に、ガイドライン⑥も読むように、と、誘導する形でもよい。
- 構成員（樫村構成員）

他のガイドラインを参照させる場合にもどこを見る必要があるかを指すとよい。
- 構成員（安藤構成員）

実証の結果が東日本大震災アーカイブに反映されるのか。対応を検討していただきたい。
- 構成員（高野座長）

先に一次リリース分が公開されたが、見ようと思っていたら閉鎖されてしまった。構築中のアーカイブの最新バージョンがいつでも見られるとありがたい。
- 構成員（森山構成員）

メタデータ作成のガイドラインについて、必須項目等を記述していく必要がある。
- 構成員（岩爪構成員）

後年度運用に関してガイドラインの他に相談できるような窓口(ガイダンス)があるとよい。

【議題5：その他】

- 事務局より、資料⑥「東日本大震災アーカイブ」基盤構築プロジェクト ラウンドテーブル開催要綱改定版」を説明。

主な意見は以下の通り。

○ 事務局（前田）

WGについて今後2回開催する件について、ご了承いただいたため年明けに早々に調整させていただく。

以 上