

メタデータの視点に基づく アーカイブとそのコンテンツ のモデル化

杉本重雄

筑波大学図書館情報メディア系

sugimoto@slis.tsukuba.ac.jp

目次

1. はじめに
 - 素朴な疑問
 - いくつかの基礎概念・モデル
 - アーカイブとそのコンテンツのモデル化
2. 東日本大震災アーカイブのメタデータ
3. メディア芸術データベース – マンガ、アニメ、ゲーム
4. 無形文化財のデジタルアーカイブ化のためのモデル
5. おわりに

1. はじめに

- 背景…素朴な疑問
 - デジタルアーカイブは使いやすいか？
 - メタデータの視点からどのような課題があるか？
- デジタルアーカイブ指向のメタデータに関して議論することを目的として筆者の研究活動から得た知見について述べる
- アーカイブとそのコンテンツのモデル化…メタデータのモデル
 - 東日本大震災アーカイブ…多数の写真と種々の記録
 - メディア芸術データベース（文化庁）…マンガ、アニメ、ゲームのデータベース
 - 無形文化財…無形文化財のデジタルアーカイブ化のためのモデル

いくつかの基礎概念・モデル

- Functional Requirements for Bibliographic Records (FRBR)
 - IFLAによる書誌記述の標準モデル
 - Work, Expression, Manifestation, Item (第1グループ (図書等))
- CIDOC-CRM (Conceptual Reference Model) と FRBRoo (FRBR object oriented)
 - Museumコミュニティで開発された標準モデル (オントロジー)
- メタデータの集約 (aggregation)
 - 何らかの基準に基づきメタデータを集め、まとめる ⇒ 記述対象の集約
- メタデータの1対1原則 (1-to-1 Principle of Metadata)
 - 1つの対象に対して1つのメタデータ記述をとらえる基本モデル

アーカイブとそのコンテンツのモデル化

- モデル化のための基本的な視点
 - コンテンツ組織化のためのデータモデル：デジタルアーカイブを使いやすくするために、コンテンツの特性とその利用目的に合わせて適切にデータモデルを作る必要がある
 - メタデータ集約のためのモデル：従来の所蔵資料単位のメタデータからそれらの集まりを作り、さらにWikipedia等の一般利用者になじみ深いリソースとの連携機能を実現するためのモデルを作る必要がある
- リソースとそのメタデータの見方（本報告における視点）
 - 個体vs.集約： 個別資料とそれらの集まり（まとまり）
 - 有形vs.無形： 具象的（物理的）実体と抽象的（論理的）実体
 - ソース vs. デジタル化： オリジナル実体とそのデジタル化実体

2. 東日本大震災アーカイブのメタデータ

- 写真、ビデオ、文書など、震災に関わる多様な資料
 - 写真が多い ⇒ メタデータ付与の問題、連続写真の判別の問題
 - 資料の組織化とメタデータの付与の課題（限られた時間、予算）
- 欲しい資料が見つかりやすいか？
 - 検索は、基本的に1資料単位（1枚の写真、1件のビデオ等）
 - 資料間のつながりやまとまりはわかりにくい ⇒ 特に写真
- 写真のまとまり（集約）を作り内容へのアクセス性を高める
 - これまでの実験から得た知見：地理情報（緯度・経度）、時間情報、作成者情報を使えばそれなりに意味のある集約を作ることができる
 - 内容（キーワードやタイトル）による集約作りが課題

東日本大震災アーカイブ集約作り

- ひなぎく（みちのく震録伝）で「祭り」と検索した結果、関連した写真が単純に並ぶ

- Aは一連の写真、B, CはAとは異なる場所でとられた写真なので、これらをそれぞれのまとまりとして表示したい

2017/7/22



東日本大震災アーカイブ集約作り実験

	アーカイブ		
	青森	久慈・野田・普代	みちのく震録伝
集約対象資料数	48,338	72,894	96,441
集約数	6,571	8,531	8,188
集約数(要素数>100)	102	189	59
集約数(要素数=1)	1,598	2,025	962

- 地理情報、時間情報、作成者情報のみを利用した集約作成の例
- 比較的簡単な方法でそれなりに集約は作ることができる
 - アーカイブ毎のデータ作成ポリシーやデータ品質の違いへの対応が必要
- 集約約に対するメタデータの作成、意味内容による集約の作成等の課題

3. メディア芸術データベース ー マンガ、アニメ、ゲーム

- メディア芸術データベース(MAD)：文化庁によるマンガ、アニメ、ゲームのデータベース
 - MADにはメディアアートも含まれるが、ここでは除外
 - 1次資料のデジタルアーカイブではないが、マンガ、アニメ、ゲームに特化した1次資料の書誌情報
- Linked Open Dataの技術をベースにMAD（あるいはそれに類したデータベースやアーカイブ）とWikipedia等をつなぐことで、データベースの利用性を高めたい
 - MADは所蔵資料(Item)指向、Wikipediaは作品(Work)中心

書誌情報とWebリソースの結び付け

- 素朴な疑問・問題意識
 - マンガ、アニメ、ゲーム(MAG)の書誌情報のためのデータモデルは？
 - FRBRを基礎にしてMADとWikipediaをつなぐことができるか？
 - Web上のいろいろなMAGリソースをつなぐにはどうすればよいか？
 - ワンピースやガンダム等のMAG横断的なWork的実体をどうとらえるべきか？
 - Work概念でMAGをつなぐことはできるか？
- これまでの研究から
 - Item指向のデータベースを利用して、同一タイトルといった要件でメタデータ集約し、（疑似的な）Work実体の同定は可能
 - MAG横断的なWork的実体はFRBRのWorkとは別物としてとらえる

4. 無形文化財のデジタルアーカイブ化のためのモデル

- 無形文化財のデジタルアーカイブ
 - 写真やビデオ等の記録物のデジタルアーカイブ
 - モーションキャプチャ等の計測データのデータとそれをもとに作られる仮想現実イメージ等
 - これらは無形文化財そのものではなく、無形文化財の実演を記録したもののデジタル化
- 「無形文化財＝文化的価値の高いモノを作り出す技能」ととらえるたうえでメタデータのためのデータモデルを考えるべき
 - 1対1原則で考えてみる

Digital Space

aggregated-as/member-of

デジタルアーカイブ内の
無形文化財コンテンツ

デジタル
コンテンツ

デジタルアーカイブの
コンテンツのメタデータ

records/recorded-by

記録物

represents / represented-by

実演

無形文化財の実体と直接
関連する実体

Instantiation-of / instantiated-as

無形文化財
(技能・技芸)

技能・技芸保持者
(Agent)

performed-by/performs

Physical Space

Digital Space

デジタルアーカイブ内の
無形文化財コンテンツ

aggregated-as/member-of

aggregation

デジタル
コンテンツ
Motion/Still Image

デジタル
コンテンツ
Motion/Still Image

デジタル
コンテンツ
Description/Text

デジタル
コンテンツ
...

records/recorded-by

記録物
Motion/Still Image

記録物
Sound/Speech

記録物
Description/Text

記録物
...

Curated Objects

represents/represented-by

実演

実演

実演

Instantiation-of/instantiated-as

Physical Space

無形文化財
(技能・技芸)

技能・技芸保持者
(Agent)

performed-by/performs

Digital Space

デジタルアーカイブ内の
無形文化財コンテンツ

aggregated as/member of

aggregation

現在の無形文化財デジタルアーカイブはこの部分のみ

Motion/Still Image

Motion/Still Image

Description/Text

...

records/recorded-by

記録物

Motion/Still Image

記録物

Sound/Speech

記録物

Description/Text

記録物

...

Curated Objects

represents/represented-by

実演

実演

実演

Instantiation-of/instantiated-as

Physical Space

無形文化財
(技能・技芸)

技能・技芸保持者
(Agent)

performed-by/performs

Digital Space

デジタルアーカイブ内の
無形文化財コンテンツ

aggregated as/member of

aggregation

デジタル
コンテンツ
Motion/Still Image

現在の無形文化財デジタルアーカイブ

デジタル
コンテンツ
...

ネット上におかれたデジタルデータのみならず無形文化財
そのもの、その実演記録を適切に記述したメタデータが必要

Web上には有用な無形文化財に関するリソースがあるので
それらをリンクづけることが重要

instantiated-by/instantiated-as

Physical Space

無形文化財
(技能・技芸)

技能・技芸保持者
(Agent)

performed-by/performs

5. おわりに

- 現在のデジタルアーカイブ（あるいはMLA）のメタデータは、所蔵資料の書誌情報、いわば、提供者として提供できる資料に関わる一般的・基盤的な情報の域を出ていない。
- ここで上げた例が対象とする資料は異なるものであるが、提供されるメタデータを基礎として、提供者視点とは少し違った視点からのアクセスを可能にすることを目的としている。
- デジタルアーカイブのコンテンツは、オリジナルのリソースのコピーであるデジタルデータであることが多く、メタデータの記述対象がぼやけてしまうことがあり、メタデータの1対1原則に基づくモデル化が重要であると考えている。

参考資料(Webページ等)

- FRBR
 - IFLA Functional Requirements for Bibliographic Records, <https://www.ifla.org/publications/functional-requirements-for-bibliographic-records>
 - Library Reference Model (LRM), <https://www.ifla.org/publications/node/11412>
- CIDOC-CRM
 - CIDOC Conceptual Reference Model, <http://www.cidoc-crm.org/>
- FRBRoo
 - IFLA: Definition of FRBRoo: A Conceptual Model for Bibliographic Information in Object-Oriented Formalism, <https://www.ifla.org/publications/node/11240>
- メタデータの1対1原則
 - Dublin Core Metadata Initiative: Glossary/One-to-One Principle, http://wiki.dublincore.org/index.php/Glossary/One-to-One_Principle
- 国立国会図書館東日本大震災アーカイブ（ひなぎく）：<http://kn.ndl.go.jp/>
- 文化庁メディア芸術データベース：<https://mediaarts-db.bunka.go.jp/>

参考資料(筆者の研究室の関連論文)

- 横山雄哉, 積 佑典, 三原鉄也, 永森光晴, 杉本重雄. 「シンプルなメタデータが付与された東日本大震災アーカイブの写真資料のための時空間情報を利用したコンテンツ集約手法」, 情報処理学会第79回全国大会, 12F-02, 2017.3
- Mihara, T., Nagamori, T., Sugimoto, S. “Creating A Work Entity Dataset of Console Games Using Wikipedia”, Poster, JADH 2017, 2017.9 (to appear)
- Kiryakos, S. et al. "Towards a Conceptual Framework for Superworks“, Poster, JADH 2017, 2017.9 (to appear)
- Wijesundara, C., Sugimoto, S., Narayan, B., Tuamsuk, K. “Bringing Cultural Heritage Information from Developing Regions to the Global Information Space as Linked Open Data: An Exploratory Metadata Aggregation Model for Sri Lankan Heritage and its Extension”, Proc of A-LIEP 2016, pp.117 - 132 , 2016.11
- Wijesundara, C., Monika W., Sugimoto, S. “A Metadata Model to Organize Cultural Heritage Resources in Heterogeneous Information Environment”, draft, 2017.7