

データベース横断検索のための 民族学標本資料情報の Dublin Core による記述

山本 泰則* 中川 隆†

概要 人文科学系のデータベースを構造のちがいを越えて、また分野横断的に検索するために、すべてのデータを Dublin Core メタデータで記述するという方法をとってきた。しかし、Dublin Core は、博物館の標本資料のような物理的な実体を記述するには、必ずしも適しているわけではない。本稿では、国立民族学博物館の民族標本資料を例に、その情報を Dublin Core へ変換方法を提案し、その有効性と課題について議論する。

Description of Ethnographic Museum Objects by Dublin Core for Cross-Domain Search

Yasunori YAMAMOTO * Takashi NAKAGAWA †

Abstract We are investigating the cross-domain retrieval system for studies in humanities where the cross-database search is realized by converting the fields of each databases to the Dublin Core metadata elements. But the Dublin Core is not necessarily suitable to describe the physical objects like in the museum. This report proposes an example to map the information of ethnographic objects in the National Museum of Ethnology to the Dublin Core, and discusses the feasibility of the mapping for cross-database search in humanities.

1 はじめに

人文科学の研究で扱う資料は、史料所在情報や全文データ、博物館のモノ資料、論文目録、図書目録、画像データ、動画データなど多種多様であり、しかも所在が分散している場合が多い。

このような資料の中から必要な情報を効率よく発見するために、人文科学系の大学共同利用機関である、国立歴史民俗博物館、国文学研究資料館、国際日本文化研究センター、国立民族学博物館および、大阪市立大学学術情報総合セ

ンター、京都大学東南アジア研究所、東京大学史料編纂所が参加して、各機関がもつ合計 30 以上のデータベースを統一的に検索・利用する方法について共同研究をおこなってきた [1, 2, 3]。

その方法の要点は、

- 各種データベースのデータ項目を Dublin Core メタデータ [7] に変換することで、データベースの構造のちがいを吸収する
- 情報検索プロトコルとして国際標準の Z39.50 をもちいることで、データベースごとに異なる検索手順を統一するというものである。

このような枠組みの中で、いかにして求める情報を発見できるかは、オリジナルデータベースの情報を Dublin Core メタデータでどれだけうまく記述できるかにかかっている。しかし、

* 国立民族学博物館 文化資源研究センター

† 国立民族学博物館 情報管理施設

* Research Center for Cultural Resources, National Museum of Ethnology

† Information and Documentation Center, National Museum of Ethnology

Dublin Core メタデータは本来インターネット上の情報資源を記述し、効率よく発見することを目的に提案されたものであるため、博物館のモノ資料のような物理的な実体を記述するには必ずしも適しているわけではない。

本報告では、国立民族学博物館（以下、民博）がもつ民族学標本資料を対象に、その情報を Dublin Core で記述する方法を提案し、問題点を検討し、データベース横断検索における有効性について議論する。

2 国立民族学博物館の民族標本資料

国立民族学博物館では、民族学の研究資料として、また人々に世界の文化を紹介するため、世界のさまざまな地域や社会に関する「モノ」資料、文献・図書資料、映像・音響資料を収集している。これらのうち「モノ」資料を民博では**標本資料**と呼んでいる。

標本資料には、

- 農業・牧畜・狩猟・漁労など生産に関する用具類
- 衣・食・住をはじめ楽器、遊びなど生活に関する用具類
- 紡織製品、木製品、金属製品、陶製品など生活技術に関する用具類
- 宗教儀礼、冠婚葬祭に関連する用具類

などが含まれる。つまり、人間の活動にかかわる、ありとあらゆる種類のモノを収集している。一方、美術・骨董的価値のあるものは原則として収集の対象にはしてない。標本資料の所蔵点数は、現在約 24 万点におよぶ。

標本資料についての基本情報はデータベース化（**標本資料 DB**）されており、その内容は、資料を名称・収集・使用・製作などの観点からみた約 40 項目の記述からなる（表 1）。また、多くの標本資料について、平面・正面・側面・鳥瞰の 4 つの異なる方向から撮影したデジタル画像が蓄積されている。

カテゴリ	項目名
製作	製作者 製作年代 製作地＋ 製作状況 製作法・材料
使用	使用民族＋ 使用者 使用年代 使用地＋ 使用状況 用途・使用法
分布	流通状況 変遷・分布
コード	OWCコード OCMコード
名前	標本名＋ 現地名（ローマ字） 現地名（カタカナ） 現地名（現地語綴） 現地名（訳）＋ 検索名
収集	収集年月日 原収集者＋ （原収集者住所＋） 収集地＋ 入手状況
関連情報	関係品 文献 関連資料・情報 その他
物理特性	寸法・重量
管理	標本番号 收藏場所
展示	展示番号
記録	履歴番号 記入年月日 記入責任者

表 1 標本資料データベースの項目一覧（＋記号はフリガナ情報がある項目）

3 モノがもつ情報とメタデータ

メタデータとは「データに関する」（構造化さ

れた)データ」と、しばしば定義される [8]。しかし、モノは、はたしてデータだろうか? また、Dublin Core メタデータに関する文書を読むと「資源の内容 (the contents of the resource)」という表現によくでくわす。モノのコンテンツとは何を意味するのか? ということをあらためて考えてみると、その答えはそれほど明確ではない。その理由は、モノはその存在自体がなんらかの情報を表しており、「情報」とそれを載せている「メディア」の区別がはっきりしないからである。

CIMI コンソーシアム (Consortium for the Computer Interchange of Museum Information [4]) では、博物館資料のような物理的な実体をも Dublin Core メタデータで記述することをめざして、ガイドライン (以下 *CIMI* ガイドと呼ぶ) [5] を作成している。このガイドでは、博物館・美術館の資料やその複製品をはじめ、資料の集合としてのコレクション、資料写真、それをデジタル化した画像、データベース、インタビュー、さらに博物館・美術館そのものまでも Dublin Core で記述することを試み、多くの事例を載せている。

これらの事例を見ると、博物館のモノ資料がもつ情報というものが考えやすくなる。民博の民族標本資料についていえば、「標本資料は、それが作られ、使われ、売られていた、ある地域のある時代の文化的文脈に関する情報を担う一種の記述である」と解釈することができる*¹。

4 標本資料情報から Dublin Core へのマッピング

すでに 20 数万件の記録をもつ民博の標本資料の情報を Dublin Core メタデータで記述するためには、オリジナルデータベースの各データ項目を Dublin Core の各要素へ機械的に変換

*¹ 同様の考え方は、化石や動植物標本など自然史で扱う標本資料にも適用することができる。

(マッピング) する方法をとらざるをえない。この節では、民博の民族学標本資料の情報をどのような方針でマッピングをおこなったかについて述べる。

4.1 マッピングの原則

Dublin Core メタデータはいろいろな種類の情報資源を記述することを前提にしているため、各要素の意味する内容はかなり抽象的である。そのため、解釈を広げると、ひとつの要素にさまざまな情報を対応づけることが可能になる。ここでは記述をできるだけ標本資料と直接関係することに限定して、以下のような原則をたてた。

1. 記述の対象は、モノそのものとする。

実際は標本資料 DB からマッピングをおこなうが、記述の対象は標本資料 DB のレコードや標本の画像ではなく、あくまでもモノ自体である。

2. マッピングにあたっては、標本資料 DB がもつ情報が、できるだけ Dublin Core メタデータに反映されるように配慮する。

3. CIMI ガイドをマッピングの参考にする。

4. Qualifier のない simple Dublin Core を用いる*²。

4.2 マッピングの実際

普通、何かを記述するときには「5 つの W と 1 つの H」、つまり「誰がいつ、どこで、何を、なぜ、どんなふうにしたか」を漏らさずに記述しなければならないといわれる。標本資料 DB の約 40 のデータ項目をこの軸で整理すると Dublin Core の要素へ対応づける時の見通しがつけやすくなる (表 2)。

以下、Dublin Core の要素ごとに、マッピングの根拠と考慮した点を述べる。

*² 共同研究でのとり決めによる。このプロジェクトでは、システムの検索プロトコルとして Z39.50 を用いているが、Dublin Core の qualifier を適切に表現する Z39.50 の既存のアトリビュートセットとタグセットがないことも理由のひとつである。

	who	what	when	where	how	misc
製作	製作者 Creator	Description 製作法・材料	製作年代	製作地	製作状況 製作法・材料	
使用	使用民族 使用者	用途・使用法	使用年代	使用地	使用状況 用途・使用法	
分布			流通状況 衰退・分布			
コード	OWCコード		OWCコード			OCMコード Subject
名前	Title Description	標本名 現地名(ローマ字) 現地名(カタカナ) 現地名(現地語) 現地名(訳) 検索名	Date 収集年月日	収集地	入手状況	
収集	原収集者 (原収集者住所)	Subject				
関連情報	Relation	関係品	Coverage		Relation	文献 関連資料・情報 その他
	Format					
物理特性	寸法・重量					Description
管理		標本番号	Identifier	収蔵場所		
展示		展示番号	Relation			
記録		履歴番号				
	記入責任者		記入年月日			

表2 標本資料データベースの項目分類と Dublin Core の要素

- Title 標本資料 DB の標本名と 4 種の現地名をこの要素に割りあてた。
- Creator 標本資料 DB の製作者を割りあてた。製作者には、個人名、団体名、民族名などが含まれている。
- Subject OCM コードと検索名を割りあてた。OCM コードは文化人類学の視点から人間の活動を文化要素に分類したコード [12] である。また、検索名は検索の便を考え、標本名をある程度統一した用語に置きかえたものである。
- Description この要素は「資源の内容についての説明」と定義されている。本来なら標本資料についての短い説明文が入るべきだが、これに

該当する標本資料 DB のデータ項目はない。ここでは、この標本資料の説明を構成するのに必要な情報を含んでいると思われる標本資料 DB のいくつかの項目を割りあてた。

同時に、他の Dublin Core の要素にマッピングされない項目をこの項目に含められないかを検討した。

Dublin Core のユーザガイド [9] には、“Description may include but is not limited to: ... reference to a graphical representation of content” と書かれている。今回はおこなわなかったが、標本資料の画像への URL をこの要素に含める可能性もある。

DC エlement	標本資料データベースのフィールド名
Title	標本名、フリガナ、現地名（ローマ字）、現地名（カナ）、 現地名（現地語綴り）、現地名（訳）
Creator	製作者
Subject	検索名、OCM コード
Description	用途・使用法、使用民族、製作法・材料、原収集者、入手状況、その他
Publisher	“国立民族学博物館” [固定値]
Contributor	[なし]
Date	収集年月日
Type	<i>physical object</i> [固定値] <i>cultural</i> [固定値]
Format	寸法・重量
Identifier	標本番号
Source	[なし]
Language	[なし]
Relation	関係品、文献、関連資料・情報、展示番号、 原 DB のレコードを表示する URL
Coverage	収集地、使用地、製作地、流通状況、使用年代、使用状況、 製作年代、製作状況、変遷・分布、OWC コード
Rights	“国立民族学博物館” [固定値]

表3 標本資料データベースの各フィールドから Dublin Core のエレメントへのマッピング

■Publisher Dublin Core の定義ではこの要素は「資源を利用可能にしている実体」である。標本資料を展示、貸出している組織という意味で「国立民族学博物館」という固定値をこの要素に与えた。

■Contributor 標本資料 DB に、この要素に対応するデータ項目はない。

■Date Dublin Core の定義では「資源の生成や利用可能となった日と関連する日」である。標本資料では、資料が製作された日、民博に受入れた日、公開した日などが該当すると考えられる。しかし、民博の標本資料には、アートの作品のように制作年月日がはっきりしているものは少ない。CIMI のガイドでは、自然史資料の場合、資料が観測や収集された日をこの要素に入れるこ

とを提案しているの、それに習って標本資料 DB の収集年月日を割りあてた。

■Type 資源の性質やジャンルを表すこの要素には、“DCMI Type Vocabulary” [11] のように統制された語を使用することが推奨されている。今回はこの語彙の中から *Physical Object* を割りあてた。また CIMI ガイドでは、この要素のとり値として、さらに *party*、*place*、*original or surrogate*、*natural or cultural* を追加することを提案している。その中から *cultural* を固定値として採用した。

標本によっては *original or surrogate* の情報を付加できる可能性があるが、機械的なマッピングはできない。

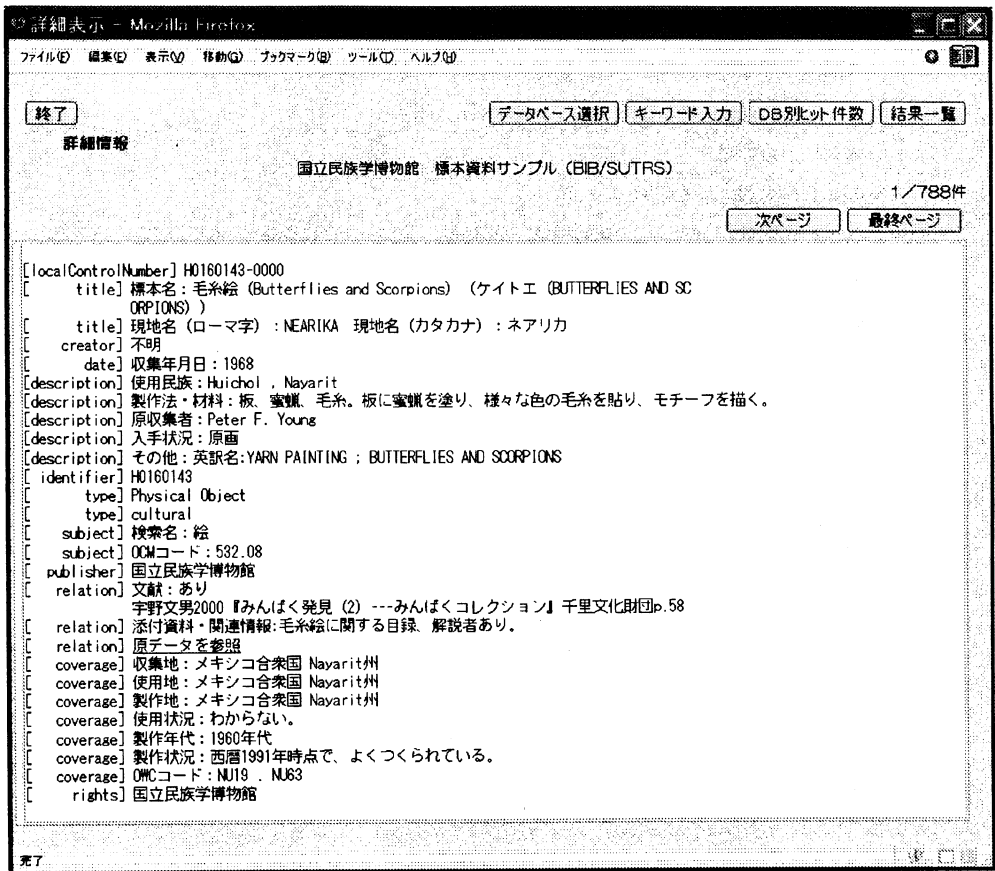


図1 Dublin Core へマッピングした標本資料 DB レコード

■Format この要素には標本資料 DB の寸法・重量を割りあてた。

材質の情報もこの要素に含まれると考える^{*3}。しかし、標本資料 DB のデータ項目製作法・材料から材料に関する情報だけを機械的に抽出することはできないこと、また、CIMI ガイドでは、材質や製作技術の情報は Description に含めることを推奨しているので、製作法・材料は Description の方に割りあてた。

■Identifier 電子的なデータではなく物理的なモノそのものを Dublin Core で記述するという

原則にしたがって、民博で標本資料を識別するために与えている標本番号を割りあてた。

■Source この要素はいくらでも拡大解釈できる可能性がある。CIMI ガイドも、意味のあいまいな要素であることを指摘している。今回は、この要素にマッピングするものはないと考えた。

機械的なマッピングはできないが、たとえば展示用に作成したレプリカ資料の場合、その元になった資料をさし示す情報がここに入る可能性はあろう。

■Language 標本資料 DB に該当する項目がないためマッピングはしない。経典や石碑のように文字を多く含む資料については、個別に情報を補える可能性がある。

^{*3} 実際、qualifier 付きの Dublin Core では、Format 要素に Medium という qualifier が提案されている [10]。

■Relation 「その標本資料と組みあわせてセットを構成する他の標本資料(たとえば、コーヒークップに対するソーサーなど)」を意味する関係品や文献、関連資料・情報をこの要素に含めた。また、元の標本資料 DB のレコードを参照する URL もこの要素に含めた。

■Coverage 標本資料は、そのモノが作られ使われていた、ある社会や地域、時代についての情報を記述していると考え、そのような時間・空間に関する情報を含んでいそうな標本資料 DB の項目は、すべてこの要素にマッピングした。この中には、世界の人々を、社会・民族、時代で区分した OWC コード [13] も含まれる。ただし、OWC コードを除いて、これらの項目のほとんどは記述的に書かれているので、Dublin Core で定義しているように時代や地域の表記の統制はむずかしい。

■Rights この要素には「国立民族学博物館」という固定を与えた。

以上のようにマッピングをおこなったが、標本資料 DB のデータ項目の中で、資料管理のための情報である収蔵場所とレコード自体の管理情報である記入責任者、記入年月日、履歴番号はマッピングしなかった。

このマッピングにしたがって標本資料 DB を変換し民博の Z39.50 サーバ [14] にインプリメントしたときのデータベースのレコードの一例を図 1 に示す。このインプリメントでは、アトリビュートセット Bib-1 [15] と TagSet-G [16] 中の Dublin Core に対応する部分を用いている。Dublin Core の各要素の値は、「使用地:メキシコ合衆国」のように元の標本資料 DB のデータ項目名を補って表示している。補った「使用地:」は検索結果レコード生成の最終段でオンデマンドで生成しているため、検索の対象にはならない。

5 まとめ

本稿では、人文科学系データベースを横断検索するために Dublin Core メタデータを用いたとき、博物館のモノ資料、特に民族学標本資料の情報を Dublin Core でどう記述するかについて、国立民族学博物館の標本資料を例に示した。議論の余地はあるにせよ、Dublin Core ヘマッピングの方法に一定の方針を与えられたと考える。

Dublin Core ヘマッピングすることにより、オリジナルデータベースのもつ情報の意味がぼやけてしまうのは否めない。しかし、これは Dublin Core に限ったことではなく、共通のメタデータに変換することにより構造の異なるデータベースの横断検索を実現した場合、程度の差こそあれ必ず生じる問題である。今回のように、オリジナルデータベースを適宜参照できる機能を付加することを前提にするなら、むしろ、メタデータの共通化により横断検索が実現できることの方を評価すべきであろう。つまり、Dublin Core の *resource description* としての機能よりも *resource discovery* の機能の方を重視すべきではないだろうか。

今回は Dublin Core に変換したレコードの中で、たとえば「[coverage] 使用地:メキシコ合衆国」のように、データの中に元のデータベースの項目名を埋めこむ方法をとった。これは、ある意味では私製の *qualifier* (検索条件には指定できないが) を導入したのに等しい。この方法は、人が見るためのレコードとしては、意味の理解を助けるが、機械処理のため入力データとしては、余分な情報を付加したことになる。

Dublin Core へのマッピングの方法は一意に決められるものではなく、マッピングを定義する者によって、また博物館によってゆれが生じる可能性がある。博物館のコミュニティでマッピングの方針について合意が形成されることが望まれる。ただその場合も、博物館のモノ資料

だけがうまく記述できるだけでは不十分である。他の性質の情報、他の分野の情報が Dublin Core でどう記述されるかを考慮に入れたうえで最適なマッピングを求めるべきであろう。

謝辞 本研究は、総合研究大学院大学共同研究プロジェクト「文化科学研究分野における情報資源共有化のためのコラボレーション研究」および、国文学研究資料館共同研究「文化情報資源の共有化システムに関する研究」の研究成果の一部である。日頃、有益なご意見とご討論をいただく同研究会の方々に感謝の意を表する。

参考文献、URL

- [1] 原正一郎：Z39.50 とメタデータによる研究機関連携，情報処理，43(9)，pp.968-974，2002.9.
- [2] 原，柴山，安永：メタデータによるデータベースの機関間連携の実現，人文科学とコンピュータシンポジウム論文集，情報処理学会シンポジウムシリーズ 2003(21)，pp.17-22，2003.12.
- [3] 山本，原，柴山，他：Dublin Core メタデータと Z39.50 プロトコルにもとづく人文科学系データベースの統合検索に関する実証実験，人文科学とコンピュータシンポジウム論文集，情報処理学会シンポジウムシリーズ 2004(17)，pp.199-205，2004.12.
- [4] Consortium for the Computer Interchange of Museum Information. <<http://www.cimi.org>>
- [5] CIMI Consortium, *Guide to Best Practice: Dublin Core Version 1.1*, 2000.4. <http://www.cimi.org/public_docs/meta_bestprac_v1_1_210400.pdf>
- [6] Dublin Core Metadata Initiative. <<http://dublincore.org/>>
- [7] Dublin Core Metadata Initiative: *Dublin Core Metadata Element Set, Version 1.1: Reference Description*, 2004.12. <<http://dublincore.org/documents/dces/>>
- [8] Dublin Core Metadata Initiative: *DCMI Frequently Asked Questions (FAQ)*. <<http://dublincore.org/resources/faq/>>
- [9] Dublin Core Metadata Initiative: *Using Dublin Core*, 2005.5. <<http://dublincore.org/documents/usageguide/>>
- [10] Dublin Core Metadata Initiative: *Using Dublin Core - Dublin Core Qualifiers*, 2005.5. <<http://www.dublincore.org/documents/2005/05/26/usageguide/qualifiers.shtml>>
- [11] Dublin Core Metadata Initiative: *DCMI Type Vocabulary*, 2004.6. <<http://dublincore.org/documents/dcmi-type-vocabulary/>>
- [12] Murdock, G.P. et al.: *Outline of Cultural Materials*, HUMAN RELATIONS AREA FILES, INC., 1982.
- [13] Murdock, G.P.: *Outline of World Cultures*, HUMAN RELATIONS AREA FILES, INC., 1983.
- [14] 山本，中川：博物館資料情報共有の試み — Z39.50 CIMI プロファイルの民族学標本資料データベースへの応用，画像電子学会第 32 回年次大会予稿集，pp.101-108，2003.6.
- [15] Bib-1 Attribute Set, 2003.12. <<http://www.loc.gov/z3950/agency/defns/bib1.html>>
- [16] Z39.50 Maintenance Agency: TagSet -G and -M Elements. <<http://www.loc.gov/z3950/agency/defns/tag-gm.html>>