みんぱくリポジトリ
国立民族学博物館学術情報リポジトリ National Museum of Ethnolo

Z39.50 CIMIプロファイルにもとづく民族学標本資料データベースの試作

メタデータ	言語: Japanese
	出版者: 情報処理学会
	公開日: 2015-11-18
	キーワード (Ja):
	キーワード (En):
	作成者: 山本, 泰則, 中川, 隆
	メールアドレス:
	所属:
URL	http://hdl.handle.net/10502/5584

Z39.50 CIMIプロファイルにもとづく民族学標本資料データベースの試作

山本 泰則
国立民族学博物館 博物館民族学研究部

中川 隆
国立民族学博物館 情報管理施設

CIMIプロファイルとは、情報検索の共通プロトコルZ39.50をもちいて博物館・美術館の間の情報共有をおこなうために提案された仕様である。国立民族学博物館では、このプロファイルに準拠して所蔵する民族資料のデータベースを試作した。本報告では、プロトタイプシステムの機能の紹介しCIMIプロファイルの適用可能性を検討する。

Prototype Database of Ethnographic Museum Objects based on Z39.50 CIMI Profile

Yasunori YAMAMOTO
Department of Museum Anthropology,
National Museum of Ethnology

Takashi NAKAGAWA
Information and Documentation Center,
National Museum of Ethnology

CIMI Profile is a set of technical specifications for using Z39.50 to search and retrieve cultural heritage information in the distributed repositories. We implemented a CIMI server to retrieve information of ethnographical objects in the museum. This report describes the system and discusses the effectiveness of the Profile.

1 はじめに

インターネットを介して文化情報資源を共有するためのさまざまな試みがなされている。博物館・美術館のもつ情報も、そういった情報資源のひとつである。博物館・美術館の間で情報を交換・共有する枠組みのひとつとして、CIMI (Computer Interchange of Museum Information) コンソーシアム [1] が作成したCIMIプロファイル [2] がある。CIMIプロファイルでは、サーバ-クライアント間での情報検索の標準プロトコルであるZ39.50 [3] をベースに、文化財（*cultural heritage*）情報特有のデータ項目を定義している。これらの項目に各組織が持つデータベースのデータ項目をマッピングすることにより、異なる構造のデータベース間で情報の相互アクセスが可能となり、組織を横断した検索が実現できる。

国立民族学博物館（以下、民博）では、CIMIプロファイルに対応したZ39.50サーバを作成し、所蔵する民族資料の情報をこのサーバをとおして検索できるようにした。本報告では、実際にシステムを実装し、約5000件のサンプルデータベースを作成した経験にもとづき、博物館の民族資料におけるCIMIプロファイルの適

用可能性について検討する。

2 Z39.50 CIMI プロファイルと民族資料

まずはじめに、前提となるとなる基本事項についてまとめておく。

2.1 Z39.50 プロトコルと CIMI アプリケーションプロファイル

Z39.50 とは、情報検索システムのサーバとクライアントの間で、検索要求と検索結果の取得（返戻）について手順を定めた規約である。クライアントは検索語にアトリビュートをつけてサーバへ質問を送り、サーバは検索結果をタグをもちいてレコードに組み立て、クライアントに返す。アトリビュートは検索語の属性（どの検索項目が対象かなど）の指定にもちい、タグは返戻レコードを構成する要素（項目）である。

サーバ-クライアント間でアトリビュートとタグを共有することにより、データベースのスキーマを抽象化してデータベース間の差異を吸収し、複数のデータベースを同じインタフェースで検索することができる。アトリビュート、タグとローカルなデータベースの項目（フィールド）間の変換はサーバの責任で、クライアントからは見えない。

Z39.50 は、はじめ書誌データベースを横断検索する目的で広まった。

Z39.50 のアトリビュートやタグ、返戻レコードの構造などを、特定の分野に応用するための指針がアプリケーションプロファイルである。現在いくつかのものが提案されているが [4]、その中に CIMI コンソーシアムが定めた **CIMI** プロファイルがある。これは、博物館や美術館間で文化財情報（芸術、建築、文化史、自然史などの分野を含む）を相互に検索・利用することを目的にしており、以下のような特徴がある。

- 検索項目を指定する Use アトリビュートには、Bib-1 アトリビュート[1] [5] から取りいれた要素（14 個）に加え、CIMI 独自の要素（55 個）と Dublin Core [6] の基本要素（15 個）が含まれている。
- 標準の tagSet-G と tagSet-M [7]、Digital Collection タグセット [8] に加え、CIMI 独自のタグセット（要素数 60）が定義されている。この中には画像や音響情報などバイナリデータとそのメタデータを記述するためのタグが含まれている。
- 返戻のデータ項目集合を規定したエレメントセットとして、標準の B と F の他に、MB (Museum Brief record) が定義されている。
- 返戻の抽象レコード構造が定義されている。
- CIMI プロファイルに対応していない Z39.50 クライアントも、最低限の情報検索ができるよう考慮されている。

2.2 国立民族学博物館の民族資料

民博では、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、世界の諸民族が現在使用している、あるいは使用していた「モノ」を収集している。これら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の対象としている。しかし、美術・骨董的価値のあるものは原則として収集していない。現在約 24 万点の標本資料を所蔵している。

約 30 年前の創設当初から、民博ではコンピュータの利用を念頭において各資料についての情報を整理し、早期にデータベース化をおこなった。その内容は、資料を収集・使用・製作

[1] 書誌情報の検索にもちいる

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

表 1: 標本資料データベースの項目一覧（＋記号はフリガナがあることを意味する）

の観点からみた記述が中心で、全部で約 40 項目の情報からなる（表 1）。現在、データベースを XML 化するための移行作業をおこなっている。

また、多くの標本資料について、4 つの異なる方向からデジタル画像を撮影し蓄積している。

3 CIMI サーバを介した標本資料データベースの試作

この節では、Z39.50 サーバと標本資料データベースの実装ついて述べる。以下、混同を避けるため、従来のスキーマにしたがって作成した民博の標本資料データベース（XML 形式）をオリジナル **DB** と呼ぶことにする。

3.1 CIMI サーバ

CIMI プロファイルに対応した Z39.50 サーバ（**CIMI** サーバ）は、サーバ部、データベース部、返戻生成部からなる（図 1）。

サーバ部　クライアントからの検索要求を解析し、データベースの検索命令に変換してデータベース部へ送る。クライアントからの返戻要求に対して、定められたレコード形式とプロトコルにしたがって検索結果をクライアントへ返す。また、セッション管理などもおこなう。

データベース部　オリジナル DB のデータを CIMI-1 の Use アトリビュートに変換したデータベースで、IBM の DB2 と全文検索エンジンテキスト情報エクステンダで管理している。データベース作成にあたっては、オリジナル DB のレコードを、XSLT によって CIMI-1 のアトリビュートに変換した。

返戻生成部　検索結果を CIMI のスキーマと Z39.50 のレコードシンタックスに変換し、返戻レコードを生成する部分である。変換は、XSLT プロセッサによって、標本資料 DB → CIMI スキーマ → 返戻レコードという 2 段階でおこわれる。XSLT はエレメントセット（B、MB、F）とレコードシンタックス（GRS-1、SUTRS）の組合せに応じて 6 種類用意されており、必要に応じて使い分けられる。

本システムでは、オリジナル DB からアトリビュートへの変換は、検索の効率を考慮して、事前おこないデータベースで管理しているのに対して、返戻レコードへの形式変換はオンデマンドでおこなっている。

検索のおおまかな流れは以下のようになる。

1. クライアントが検索要求をだすと、サーバ部はそれを解析してデータベース部へ検索命令を送る。検索結果のレコード id の集合は、サーバ部が保持している。
2. クライアントから返戻要求を受けとると、サーバ部は検索結果レコード id 集合をもとに、返戻生成部へ返戻レコードの作成を指示する。
3. 返戻生成部は、オリジナル DB のレコードを XSLT により変換して返戻レコードを作成し、それをサーバ部へ渡す。
4. サーバ部は、Z39.50 のプロトコルにしたがって返戻レコードをクライアントへ返す。

本サーバは CIMI とは別に、Bib-1 のアトリ

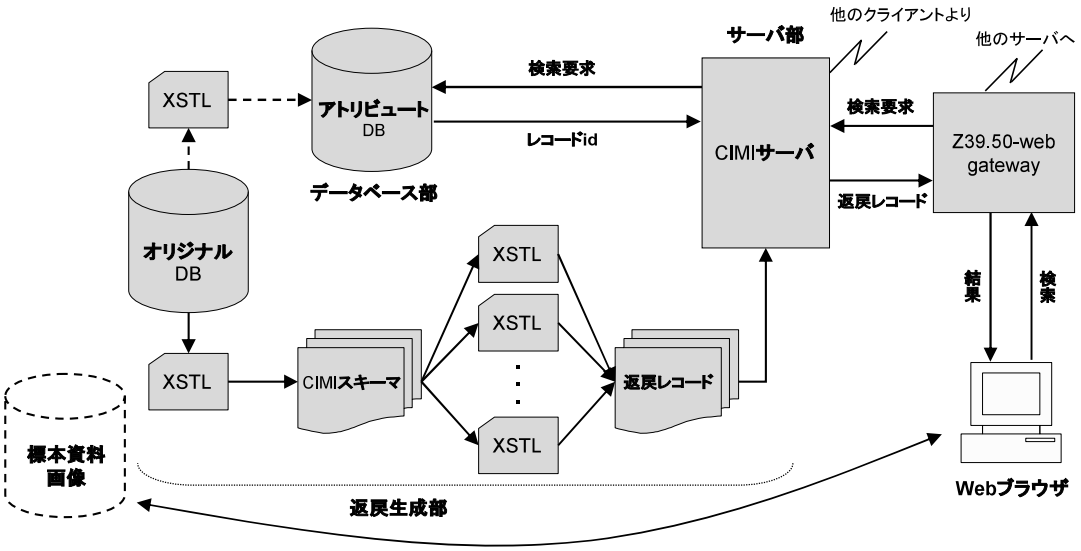


図 1: CIMI サーバの構成と Z39.50-Web gateway

ビュートに対応したデータベース部と返戻生成部ももっている。そのため、Bib クライアントに対しては Bib サーバとして機能することができる[2] 。

3.2 Z39.50-Web gateway

本サーバや他の Z39.50 サーバを Web ブラウザからも利用できるようにするため、Z39.50-Web gateway を同時に開発した。Gateway はユーザの検索要求を Z39.50 プロトコルに変換してサーバに送り、検索結果を HTML に変換してブラウザに返す。ブラウザでは 1 レコード／1 行の一覧表示と 1 レコードごとの詳細表示を選択して検索結果を見ることができる。

CIMI サーバから *mrObject/rendition* 下のタグ *resource* (5,30) として画像データの URL を受けとったとき、gateway はその画像サイズに関する記述を調べ、“thumbnail” か “wallet” の場合は画像そのものを、それ以外の場合は画像へのリンクをブラウザに表示する（図 2)。

本 gateway は、CIMI クライアントとしても Bib クライアントとしても同時に機能して、複数のサーバと同時に接続することができるので、これによってデータベース間の横断検索が可能となる[3] 。

3.3 標本資料 DB から CIMI へのマッピング

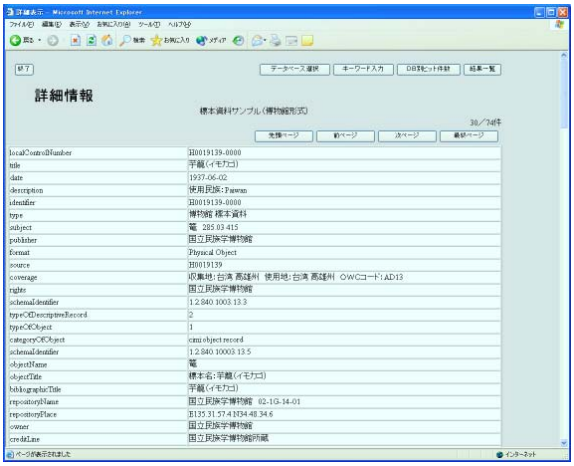
CIMI プロファイルという共通の枠組をとおして博物館間で相互に情報をアクセスできるようにするためには、民博のオリジナル DB の各項目から CIMI のアトリビュートとタグへ、いかにマッピングするかが重要である。

CIMI のアトリビュートとタグは、意味内容から次のように分類できる。CIMI 独自のもので、博物館資料の内容や管理情報を記述するもの。分野を越えて関連する情報を発見するための Dublin Core の基本要素や、*who*、*what*、*when*、*where* というアトリビュート。Bib-1 から取りこまれた書誌情報を記述するアトリビュート。画像・音声など文字でない情報を記述するタグ *mrObject*、*rendition*。

マッピングにあたっては、CIMI プロファイルで定義されているアトリビュート、タグの意味とオリジナル DB の各項目の意味を十分吟味し、全体のバランスを考慮しながら対応づけをおこなった（具体的なマッピングの抜粋は付録

[2] CIMI サーバは、Bib クライアントからの検索要求にも、ある程度は応じられるよう、仕様できめられているが、本サーバでは、Bib クライアントに対しては、Bib サーバーとして機能する。

[3] Dublin Core の各要素に割り当てられている use attibute value は、CIMI-1 と Bib-1 では異なるのだが、この geteway はそれぞれに変換してサーバに検索要求をだす。



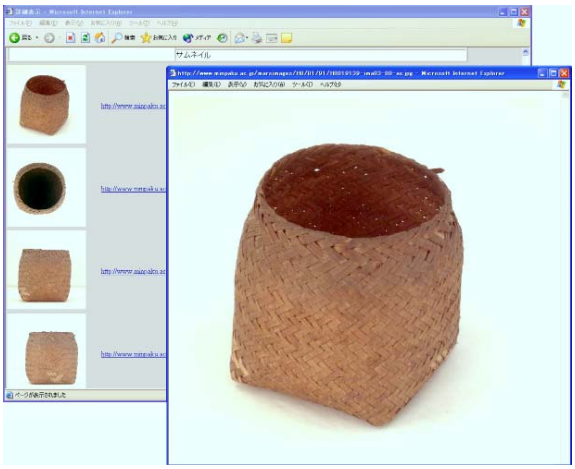


図 2: Web ブラウザで表示した検索結果と画像

の表を参照)。

アトリビュートやタグのなかで、博物館の資料の記述のための要素に注目すると、記述の詳しさはオリジナル DB の項目にほぼ等しいかより細分化されている。それらの項目へのマッピングではつぎのような場合が生じた。

- オリジナル DB にほぼ対応する項目があるもの（標本番号→ *objectID*、寸法・重量→ *demensions*、収蔵場所→ *ropositoryName*、原収集者→ *fieldCollector*、製作地→ *placeOfOrigin*、関係品→ *relatedObjects*）、画像情報→ *mrObject*）
- マッピングする項目がオリジナル DB にないもの。これは 3 つの場合に分かれ、固定値を補える場合（*owner*、*repositoryPlace*、*address*、*administrator*）；資料から情報が得られるはずだがオリジナル DB には採録されていないか、資料の性格上内容がそぐわない（*inscriptionMark*、*physicalDescription*、*creatorDateOfBirth*、*agePeriod*、*typeSpecimen*、*contextArchaelogical*；オリジナル DB の項目の方が記述の範囲が広く、必ず該当する情報があるとはかぎらない（製作者、製作法・材料、変遷・分布、*contentGeneral*、*contextHistorical*）である。
- オリジナル DB の項目にはあるが、マッピングすべき CIMI のアトリビュートやタグがないもの（収集地、使用民族、使用年代、使用地、用途・使用法、製作状況、流通状況）。

最後の 2 つ場合も、オリジナル DB の項目がなんらかの検索の対象となるように、また返戻レコードの中に情報があらわれるように、*associationGeneral* や *DC-description*、*DC-coverage*、また *who*、*what*、*when*、*where* など、包括的なと情報を記述するアトリビュートやタグへマッピングした。その結果、情報の焦点がぼやけてしまうのは否めない。

アトリビュートは検索時に情報の属性を指定するものであるのに対し、タグは返戻レコードを記述するためのものという両者の機能のちがいを考慮して、対応するアトリビュートとタグでマッピングのしかたを変えた場合がある。つまり、アトリビュートの方は、関連した情報がありそうなオリジナル DB の項目にできだけ広く網をはり検索もれがないようにした。一方タグの方は、オリジナル DB の同じ項目の内容がいくつものタグにマッピングされ返戻レコードが冗長になることができるだけないように、また、オリジナル DB の項目が少なくとも 1 か所には出現するように、マッピングを決定した。

たとえば、オリジナル DB の「製作者」という項目には、製作者の名前だけでなく、製作者の職業や地域社会や家庭内での役割などが記述されていることも多い。そのため、この項目はアトリビュートの *creatorName* にはマップしたが、タグの *creatorName* にはマップしなかっ

た。それに対して、*creatorGeneral* についてはアトリビュートへもタグへも「製作者」をマップした。

構造をもつタグ *mrObject* には、標本資料の画像のメタデータを記述した。ひとつの標本資料について撮影方向の異なる画像ごとに *mrObject* を、サムネイル画像と精細画像という異なる解像度ごとに *rendition* を記述した。*rendition* の下位の要素である *resource* には画像の URL を割りあてた。この情報を受けとった CIMI クライアントは標本の画像を蓄積する Web サーバにアクセスして、画像を表示することができる。

4 考察

前節で述べたマッピングのずれは、民族資料と美術系、歴史系、自然史系の博物館資料との関心のちがいを反映している。民族資料では、資料の作者の個人名が問題になることは少ないし、鉱物や化石のような体系的な分類はできないし、考古学的な時代とは無縁である。一方、民族資料の多くは、使われている社会にとってはごくありふれた日常のモノで、それをだれが作り、どのように流通し、どう使われるかに関心がある。

CIMI に適当なマップ先がなかったオリジナル DB の項目については、今回は *associationGeneral* や Dublin Core の *coverage* や *description* というような包括的な情報をあらわす項目へマッピングをおこなった。しかし、Z39.50 には「サーバが local に定義したタグ」というタグタイプ（3）があるので、それを使ってオリジナル DB の項目をそのまま生かして返戻レコードを構成した方より正確な情報になると思われる。ただし、local なアトリビュートは定義できないので、検索項目へのマッピングの問題は残る。

ところで、CIMI のアトリビュートには、人についての情報（*who*）、ことがらについての情報（*what*）、日付や時代についての情報（*when*）、場所についての情報（*where*）という項目があった。こういった側面から情報をしぼりこむ方法は、多種多様な情報の中から求める情報を発見するために、Dublin Core のメタデータを補う有効な手段になる可能性がある。

さて、歴史系のいくつかの研究機関では、国文学研究資料館が中心になって、Dublin Core と Z39.50 プロトコルをベースにデータベースを相互運用する計画が進んでいる [9]。CIMI プロファイルにも Dublin Core に対応する項目があり、検索プロトコルも同じなので、これらのデータベースとの横断検索も技術的にはむずかしいことではない。今後は、他機関の CIMI サーバとの相互検索や歴史系データベースのような分野を越えた横断検索のテストをおこない、CIMI の適用可能性をさらに評価していく必要がある。

参考文献、URL

[1] Consortium for the Computer Interchange of Museum Information. <http://www.cimi.org>

[2] Consortium for the Computer Interchance of Museum Information, Z39.50 Working Group, “The CIMI Profile: Z39.50 Application Profile for Cultural Heritage Information, Release 1.0H” (1998.9) < http://www.cimi.org/products/cimi_products.html#THREE >

[3] Z39.50 Maintenance Agency. <http://lcweb.loc.gov/z3950/agency>

[4] Z39.50 Profiles <http://lcweb.loc.gov/z3950/agency/profiles/profiles.html>

[5] “Attribute Set Bib-1 (Z39.50-1995): Semantics” (1995.9) <ftp://ftp.loc.gov/pub/z3950/defs/bib1.txt>

[6] Dublin Core Metadata Initiative. <http://dublincore.org/>

[7] Z39.50 Maintenance Agency, “TagSet -G and -M Elements” <http://lcweb.loc.gov/z3950/agency/defns/tag-gm.html>

[8] Library of Congress, “Z39.50 Profile for Access to Digital Collections” (1996) <http://lcweb.loc.gov/z3950/agency/profiles/collections.html>

[9] 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』、vol.43, no.9 (2002.9)

A　付録　標本資料 DB から Attribute、Tag へのマッピング（抜粋）[4]

ラベル	Attribute へのマップ	Tag へのマップ
objectName	検索名	検索名
objectTitle	標本名+、現地名（ローマ字）、現地名（カタカナ）、現地名（現地語綴り）、現地名-訳+	標本名+、現地名（ローマ字）、現地名（カタカナ）、現地名（現地語綴り）、現地名-訳
creatorGeneral	製作者	製作者
creatorName	製作者	—
creatorNationality CultureRace	—	—
creatorDateOfBirth	—	—
creatorDateOfDeath	—	—
creatorRole	—	—
fieldCollector	原収集者+	原収集者
repositoryName	”国立民族学博物館”、収蔵場所	”国立民族学博物館”、収蔵場所
repositoryPlace	”E135.31.57.4 N34.48.34.6”	”E135.31.57.4 N34.48.34.6”
owner	”国立民族学博物館”	”国立民族学博物館”
creditLine	”国立民族学博物館所蔵”	”国立民族学博物館所蔵”
subjectContent	検索名、OCM コード	検索名、OCM コード
objectID	標本番号	標本番号
materialMedium	製作法・材料	—
processTechnique	製作法・材料	製作法・材料
dimensions	寸法・重量	寸法・重量
placeOfOrigin	製作地+	製作地
dateOfOrigin	製作年代	製作年代
dateCollected	収集年月日	収集年月日
agePeriod	—	—
typeSpecimen	—	—
stylePeriod	—	—
periodName	—	—
provenance	入手状況	入手状況
quantity	—	—
award	—	—
collection	—	—
inscriptionMark	—	—
objectLanguage	—	—
condition	—	—
physicalDescription	—	—
protectionStatus	—	—
protectionDate	—	—
spatial ReferencingSystem	”緯度経度”	”緯度経度”
x-coordinate InReferencingSystem	”E135.31.57.4”	”E135.31.57.4”
y-coordinate InReferencingSystem	”N34.48.34.6”	”N34.48.34.6”

（次ページへつづく）

[4] 表中で、イタリックのラベルはアトリビュートの項目名を、それ以外はタグの要素名をあらわす。 また、+のついた項目名はフリガナ情報もマップしたことを、— はオリジナル DB にマップ元がないことを意味する。

（前ページより）

ラベル	**Attribute へのマップ**	**Tag へのマップ**
address	”565-8511 大阪府吹田市千里万博公園 10-1”	”565-8511 大阪府吹田市千里万博公園 10-1”
relatedObjects	関係品	関係品
related TextualReferences	文献	文献
associationGeneral	展示番号、使用者、製作状況、流通状況、関連資料・情報、その他	展示番号、使用者、製作状況、流通状況、関連資料・情報、その他
contentGeneral	—	—
contextHistorical	—	—
contextArchaelogical	—	—
copyrightRestriction	—	—
wallTextLabel	—	—
administrative EventGeneral	—	—
administrator	”国立民族学博物館”	”国立民族学博物館”
resource	画像ファイルの URL	画像ファイルの URL（と画像サイズの情報）
image	true	*[no associated element (s)]*
who	記入責任者、原収集者+、使用民族+、使用者、製作者、流通状況、入手状況	*[no associated element (s)]*
what	標本名+、現地名（ローマ字）、現地名（カタカナ）、現地名（現地語綴り）、現地名-訳+、検索名、用途・使用法、製作法・材料、関係品	*[no associated element (s)]*
when	記入年月日、収集年月日、使用年代、使用状況、製作年代、製作状況、変遷・分布	*[no associated element (s)]*
where	ＯＷＣコード、収集地+、使用地+、収蔵場所、製作地+、流通状況、変遷・分布	*[no associated element (s)]*
DC-title	標本名+	標本名’（’フリガナ’）’
DC-creator	製作者	製作者
DC-contributors	—	—
DC-date	収集年月日	収集年月日
DC-description	用途・使用法、使用民族+、使用者、使用年代、使用状況、製作状況、製作法・材料	用途・使用法、使用民族、使用年代、使用状況
DC-identifier	標本番号’-’履歴番号	標本番号’-’履歴番号
DC-type	”博物館 標本資料”	”博物館 標本資料”
DC-language	—	—
DC-subject	検索名、OCM コード	検索名、OCM コード
DC-publisher	”国立民族学博物館”	”国立民族学博物館”
DC-format	”Physical Object”	”Physical Object”
DC-source	標本番号	標本番号
DC-relation	関係品	関係品
DC-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	収集地、使用地、製作地、使用年代、製作年代、変遷・分布、OWC コード
DC-rights	”国立民族学博物館”	”国立民族学博物館”