Z39.50 CIMIプロファイルをもちいた民族学標本資料の情報共有にむけて

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・美術館の間で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、所蔵する民族標本資料の情報を検索・提供するシステムを試作した。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について検討する。

Towards the Cross-Agency Search for Information 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 ethnographic objects in the National Museum of Ethnology. This report describes the pilot system and discusses the feasibility 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。分野を横断した情報検索では、Dublin Core [1] の基本要素がもちいられることが多いが、博物館・美術館に分野を限定した場合、Dublin Coreだけで十分な検索ができるとはいえない。

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 [2] が作成したCIMIプロファイル [3] がある。CIMIプロファイルでは、情報検索の標準プロトコルであるZ39.50 [4] をベースに、文化財（*cultural heritage*）特有のデータ項目を定義している。各組織が持つデータベースのデータ項目をそれらに変換することにより、異なる構造のデータベース間で情報の相互アクセスを可能にしている。

今回、国立民族学博物館（以下、民博）では、所蔵する資料情報を標準の手段で提供し、また、他の博物館や異分野の情報資源と情報共有する方法のひとつとして、CIMIプロファイルに対

応したZ39.50サーバを作成し、これをもちいて情報検索システムを構成した。本稿では、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、博物館の民族標本資料におけるCIMIプロファイルの適用可能性について検討する。

2 CIMIプロファイルと民族標本資料

2.1 Z39.50 CIMIアプリケーションプロファイル

Z39.50とは、サーバ-クライアント型情報検索システムにおいて、検索要求と検索結果の取得（**返戻**）の手順を定めた国際標準規約である。クライアントは検索語に**アトリビュート**（検索項目などの属性）をつけてサーバへ質問を送り、サーバは検索結果を**タグ**（データ項目）をもちいてレコードに組み立て、クライアントに返す。

サーバとクライアントの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開発され、普及した。

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[5]、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（美術、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] [6]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の基本要素（15個）が含まれている。
- 返戻レコードを構成するタグとして、標準のtagSet-GとtagSet-M [7]、Digital Collectionsタグセット[8]に加え、CIMI独自のもの（要素数60）が定義されている。そこには、画像や音響などバイナリデータとそのメタデータを記述する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標本資料

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、世界のさまざまな地域や社会で現在使用されている、あるいは使用されていた「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とはしてない。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関する情報を整理し、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・製作の観点

[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）

からみた記述を中心とする、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、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を蓄積している。

3 CIMI プロファイルをもちいた民族標本資料データベースの試作

この節では、作成した CIMI プロファイル対応の Z39.50 サーバ（**CIMI サーバ**）による標本資料データベースについて述べる。以下、混同を避けるため、従来の民博独自のスキーマにしたがって作成されている標本資料データベース（XML 形式）を、**オリジナル DB** と呼ぶ。

3.1 CIMI サーバの実装

CIMI プロファイルに対応した Z39.50 サーバ（**CIMI サーバ**）を、サーバ部、データベース部、返戻生成部という構成で実装した（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は、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 のデータ項目を 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。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 段階でおこなう。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。

本サーバでは、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、事前に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。

なお、CIMI サーバは、Bib にのみ対応した

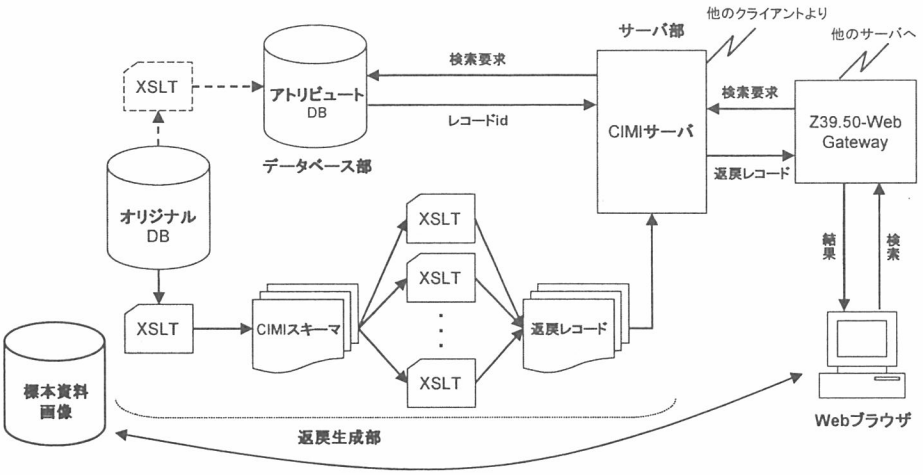

図 1: CIMI サーバの構造と Z39.50-Web Gateway

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に仕様で決められているが、本サーバは、Bib クライアントに対しては、Bib サーバとして動作する機能をもつ。

3.2 Z39.50-Web Gateway

本サーバや他の Z39.50 サーバを Web ブラウザから利用できるようにするための gateway を同時に開発した。この 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、受けとった検索結果を HTML に変換してブラウザに返す。検索結果は、1 レコード／1 行の一覧表示または 1 レコードごとの詳細表示にして、ブラウザで見ることができる。

受けとったレコードの *mrObject/rendition* 下のタグ *resource* (5,30) の情報が画像の URL であった場合、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。この機能によりデータベース間の横断検索を実現している[2]。

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

3.3 オリジナル DB から CIMI への変換

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 の基本要素やアトリビュート *who*、*what*、*when* 、*where*。Bib-1 からとり入れた書誌情報を記述するアトリビュート。画像・音声などバイナリデータとそのメタデータを記述するタグ *mrObject*、*rendition*。

マッピングは、CIMI プロファイルで定義されているアトリビュートとタグのセマンティクスにもとづいておこなったが、以下の点も考慮した（具体的なマッピングは付録の表を参照)。

- 個々のアトリビュートへは、対応する情報がありそうなオリジナル DB の複数の項目をマッピングし、検索もれがないようにする。
- タグについては、オリジナル DB の各項目が少なくともどれかのタグにマッピングされる

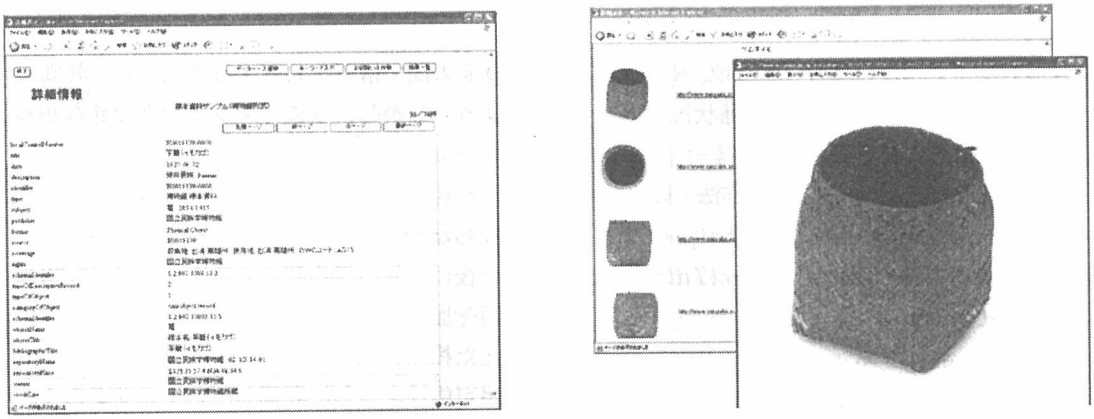

図 2: Web ブラウザで表示した検索結果（左：詳細表示、右：付随する画像）

ようにする。また、オリジナル DB のひとつの項目がいくつものタグにマッピングされて返戻レコードが冗長にならないようにする。

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ある。そのため、この項目はアトリビュートの *creatorName* にはマッピングしたが、タグの *creatorName* にはマッピングしなかった。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」をマッピングした。

また、オリジナル DB の「用途・使用法」のように、対応する項目が CIMI にない場合も、この情報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*、*who*、*what*、*when*、*where* など、包括的な情報を記述するアトリビュートやタグへマッピングをおこなった。

以上のように、多くのマッピングはオリジナル DB のデータ項目と CIMI のアトリビュートやタグの意味解釈にもとづいて決めるため、ad hoc な要因が入らざるをえない。

ところで、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している 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

オリジナル DB から CIMI のアトリビュートとタグへのマッピングについて、CIMI の特徴である博物館資料記述のための項目を中心に分析すると、つぎのように分類できる。

1. ほぼ 1 対 1 に対応する場合（標本番号→ *objectID*、寸法・重量→ *demensions*、原収集者→ *fieldCollector*、製作地→ *placeOfOrigin*、関係品→ *relatedObjects*）、画像情報→ *mrObject*、*rendition* など）
2. マッピング元がない場合。これには 2 つの場合があり、固定値を補えるもの（*owner*、*repositoryPlace*、*address*、*administrator* など）；原資料に情報はあるがオリジナル DB には採録されていない、または資料の性格上情報のないもの（*inscriptionMark*、*physicalDescription*、*creatorDateOfBirth*、*agePeriod*、*typeSpecimen*、*contentGeneral*、*contextArchaeological* など）

3. マッピング先がない場合（収集地、使用民族、使用年代、用途・使用法、流通状況など）。
4. 1対多の対応になる場合（製作者→ {*creatorGeneral*, *creatorName*}、「製作法・材料」→ {*materialMedium*, *processTechnique*} など）
5. 多対1の対応になる場合（*objectTitle*、*associationGeneral* など）

上記4.と5.は、オリジナルDBの項目が表す概念と、ほぼそれに近いCIMIの項目が表す概念の範囲が違うために生じる。

上記3.の場合、今回はやや強引に包括的な情報を表す項目へ対応づけた。その結果、情報の性格がぼやけてしまったのは否めない。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た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、この場合、検索条件をどう表現するかという問題は残る。

このようなマッピングの困難さは、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て興味深い。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ある。研究者は、それをどんな人が作り、どのように流通し、どう使われているかに関心をもっている。

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*）という項目がある。情報を絞りこむときに、こういった側面に注目する方法は、多種多様な情報の中から求めるものを発見するために、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。

以上のようにマッピング上の課題はあるが、CIMIプロファイルは、博物館のモノ資料に関する、画像も含めた基本情報を、標準規格の枠組みの中だけで扱うことができる。その点で、博物館の間で情報を共有する手段として評価できよう。しかし、うまくマッピングできなかったローカルなデータベース固有の情報をいかに扱うべきかについては、引き続き検討しなければならない。

今後は、他の博物館のCIMIサーバとの相互検索や歴史系データベース[9]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。

謝辞 システムの作成にご協力いただいた日本アイ・ビー・エム（株）に感謝いたします。

参考文献、URL

[1] Dublin Core Metadata Initiative. <http://dublincore.org/>
[2] Consortium for the Computer Interchange of Museum Information. <http://www.cimi.org>
[3] 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 >
[4] Z39.50 Maintenance Agency. <http://lcweb.loc.gov/z3950/agency>
[5] Z39.50 Profiles <http://lcweb.loc.gov/z3950/agency/profiles/profiles.html>
[6] "Attribute Set Bib-1 (Z39.50-1995): Semantics" (1995.9) <ftp://ftp.loc.gov/pub/z3950/defs/bib1.txt>
[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)
[10] Finnigan,S. and Bird,L., "Harnessing cultural resources: the Web, Z39.50 and the CIMI profile", Proceedings of the CIDOC '98 Conference (1998.10) <http://www.willpowerinfo.myby.co.uk/cidoc/cidoc98/finnigan.doc>

A　付録　オリジナルDBからCIMIのAttribute、Tagへのマッピング（抜粋）[3]

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）

[3] 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	—	—
contextArchaeological	—	—
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”