

博物館情報に基づく複合的メタデータスキーマ 構築のための言語設計に関する検討

秋元 良仁 亀山 渉
早稲田大学大学院 国際情報通信研究科

概要

近年の情報技術の進展に伴い、利用できるコンテンツの量は飛躍的に増大している。また、これら膨大なコンテンツにいつでもどこでもアクセスできるユビキタスな環境も整いつつある。このような状況では、メタデータを用いて多種多様なコンテンツから利用者の要求に応じて意味のある情報を横断的に紡ぎ出す技術が求められる。そこで、本稿ではまずメタデータの現状と問題点を、特に博物館を事例として整理する。それをもとに、複合的にメタデータスキーマを構築する手法である「ファジィ・スキーマ」という概念を提案する。更に、複数の博物館の資料情報管理用メタデータスキーマを基に、ファジィ・スキーマを構築するための言語設計について検討する。

A Study on Language Design of Creating Compound Metadata Schema Method Based on Museum Information

Ryoji AKIMOTO Wataru KAMEYAMA
Global Information and Telecommunication Studies, Waseda University

Abstract

As information technology has progressed recently, the amount of contents that can be used increase. And, there is an environment in which lots of contents can be accessed at anytime and anywhere. In this situation, a technology is needed to extract useful information from various contents by using metadata in order to meet the requirements. In this paper, we summarize the current state of metadata technology and problems. Based on it, we propose the concept called Fuzzy Schema. Fuzzy Schema can flexibly construct metadata schema if necessary. And we also describe the design of Fuzzy Schema language taking an example of metadata schema of museum.

1 はじめに

近年、情報技術の活用が急速に進展し、利用者が取り扱うことのできる情報量は飛躍的に増加している。800億ページと言われるWebコンテンツを始め、放送・出版・音楽・映像・画像等の多種多様なコンテンツ、これらの複合体としてのデジタル・アーカイブ等、個人のみならず、企業や関係機関が編集を行い提供されるコンテンツが多数存在する。特に、近年博物館においては古来からの人類の営みを表現した種々の文化財のデジタルコンテンツ化が進んでいる。また、近年ではこれらのコンテンツに様々なデバイスを用いていつでもどこでもアクセスできる、ユビキタスな情報環境基盤も整いつつある。

このような環境では、利用者の要求に応じて多様なリソースからコンテンツを横断的に取得し、複合的に意味のある情報を紡ぎ出す技術が求められる。

そのためには、コンテンツに関する情報を記述した

データ（メタデータ[1]）を活用する必要がある。これまでも、コンテンツの特性を考慮した様々なメタデータが提案されており、また、幾つかの国際標準や業界標準が制定されている状況にある。

しかしながら、これらメタデータには根本的な技術的課題が内在しているため、その利活用は不十分と言わざるを得ない。

そこで、本研究ではまず2節でメタデータ、特に博物館分野におけるメタデータの技術的課題を明確にする。次に、3節においてメタデータの技術的課題解決のため、博物館情報に基づき複合的にメタデータスキーマを構築する手法である「ファジィ・スキーマ」という概念を提案する。更に、ファジィ・スキーマを構築するための言語設計について検討を加える。最後に4節でまとめと今後の課題を述べる。

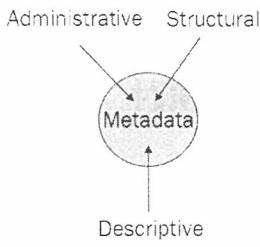


図 1: メタデータ

2 メタデータの現状と問題点

2.1 メタデータとメタデータスキーマ

コンテンツとは、デジタルデータで表現された文章や画像、映像、音楽、これらを複合的に組み合わせた情報のことを言う [2]。以下に具体的な例を示す。

- インターネット上に存在する Web コンテンツ
- TV 放送・映画に代表される映像コンテンツ
- 博物館における絵画や風景画をデジタル化した画像コンテンツ
- 電子書籍や電子コミックに代表される出版コンテンツ
- インターネットを介して配信される音楽コンテンツ
- 上記コンテンツの複合体としてのデジタル・アーカイブ

メタデータとは、「データに関する（構造化された）データ、(Structured) data about data」と定義される [1]。即ち、何らかの情報資源に関する記述は全てメタデータであると言うことができる。

メタデータの記述は、情報資源をどのように表現するかによって、Descriptive メタデータ、Structural メタデータ、Administrative メタデータの 3 種に分類することができる [3](図 1 参照)。

Descriptive メタデータは主として情報検索等、情報資源の発見や特定するために記述されるメタデータである。Structural メタデータは、情報資源の形状や物理的な特性等、情報資源の属性に関する内容記述であり、Administrative メタデータは情報資源の保存、権利関係や利用条件等、情報資源管理に関する記述がなされる。

本稿では、これら 3 種の概念を基本とし、特にコンテンツに関する情報を記述したデータのことをメタデータと言う。

メタデータは定義が広範であり、用途目的も様々であることから、それぞれのコンテンツの特性に合わせた形でメタデータの記述項目と記述形式が提案されてきた。メタデータスキーマは XML Schema や DTD、RELAX NG に代表される、何をどのように管理すべきか、その項目名や適用範囲、データ型等を定義するタイプのものと、RDF や RDF Schema、OWL に見られる項目間の関係や意味概念までも記述するものとに分類することができる。本稿ではこの両者を併せてメタデータスキーマと呼ぶ。

メタデータやメタデータスキーマの活用は、近年その重要性が再認識され、幾つかの国際標準や業界標準が制定されている。

例えば、アーカイブズ記述の国際標準に General International Standard Archival Description(ISAD(G): 国際標準記録資料記述)[4] がある。これは様々な公式文書や資料を管理・活用するためのメタデータスキーマである。また、図書館の分野では、インターネット上の情報資源発見のためのメタデータスキーマに Dublin Core[5] がある。Dublin Core はインターネット上におけるリソースに関する情報を記述するための 15 の基本要素タイプを中心とするメタデータスキーマである(表 1 参照)。

Dublin Core をベースとした映像音声等のマルチメディア記述用のメタデータスキーマに MPEG-7[6] がある。MPEG-7 では、データ記述言語として XML のサブセットを Description Definition Language(DDL) として定義し、マルチメディアコンテンツの検索を目的に、コンテンツを表現する様々な情報をメタデータで定義している。また、TV-Anytime[7] では、Electronic Program Guide(EPG) を用いた放送番組検索のみならず、インターネットで配信されるコンテンツも含めて効率良く検索できることを目指してメタデータの標準化が行われてきた。

この他にも、電子書籍で利用される XML 形式の XMDF[8] や BBeB[9]、音楽配信用に用いられる RDF 形式の MusicBrainz[10] 等、業界で標準として利用される種々のメタデータスキーマが存在する。

また、これらのメタデータスキーマを相互に活用するためのメタデータ交換用の通信プロトコルとして、Open Archives Initiative Protocol Metadata Harvesting(OAI-PMH) [11] がある。OAI-PMH は大学や研究機関等で生産された電子知的生産物を収穫・

表 1: Dublin Core 15 element set

エレメント名	定義
title(タイトル)	情報資源に与えられた名称
creator(作成者)	情報資源の内容の作成に主たる責任を持つ実体
subject(キーワード)	情報資源の内容のトピック
description(内容記述)	情報資源の内容の説明・記述
publisher(公開者)	情報資源を公開することに対して責任を持つ実体
contributor(寄与者)	情報資源の内容に寄与、貢献した実体
date(日付)	情報資源のライフサイクルにおける事象の日付
type(資源タイプ)	情報資源の内容の性質またはジャンル
format(記録形式)	情報資源の物理的形態ないしデジタル形態での表現形式
identifier(資源識別子)	情報資源を一意に特定する識別子
source(出處)	情報資源が作り出される源の情報資源への参照
language(言語)	情報資源の内容を表す言語
relation(関係)	関連情報資源への参照
coverage(時空間範囲)	情報資源の内容が表す範囲
rights(権利管理)	情報資源に含まれる（また関わる）権利情報

保存し、原則無償でインターネット上に公開する機能を持つ。近年では文化遺産オンライン [12] に応用される等、幅広い活用が見込まれている。

2.2 メタデータが抱える技術的課題

2.2.1 メタデータスキーマ構築の問題

メタデータはメタデータスキーマに基づいて生成される。通常、メタデータスキーマはその分野に精通した専門家が構築を行う。デジュール標準として作成される場合、策定プロセスや標準内容はオープンであり、かつ専門家による高度な知識が駆使されるので、その内容は信頼性が高いものとなる。しかしその反面、策定のものには膨大な時間と手間を必要とするため、構築と利用・普及の間にタイムラグが生じやすい。また、一度構築・運用が始まると、追加・削除等の変更（柔軟な対応）が困難となる。

2.2.2 メタデータ作成の問題

コンテンツ利用の促進には、メタデータが大量に作成され、インターネット等のネットワーク上に流通す

る必要がある。メタデータの作成は、例えば MPEG-7 におけるビジュアル信号やオーディオ信号から自動的に取得できる特徴量（メタデータ）等、一部自動化も見られるが、コンテンツの属性記述や維持・管理用の情報等、テキスト部分のメタデータ作成は人手を介して記述を行っている現状がある。適切なメタデータの作成作業は人的・金銭的・時間的に見て非常に負荷の高い作業と言える。

また、近年のコンテンツ量の増加に伴うメタデータ量の増加も考慮しなければならない問題の一つと言える。

2.2.3 メタデータ・マッピングの問題

複数のデータベースやアーカイブ等、多様なコンテンツから意味のある情報を横断的に取得するためには、形式の異なるメタデータスキーマで定義されるメタデータ間のインターフェラビリティをどのように保証するのかという問題も存在する。

現状、オントロジ・マッピング [13] やメタデータクロスウォーク [14] と呼ばれる、専門家を介した手動による項目間のマッピングやメディエータ、あるいはラッパーや用いた相互利用の例が見受けられる。しかし、項目間の不一致やスキーマ自体の変更が頻繁に起こるため、マッピングを行ったことにより無駄な情報の増加、あるいは必要な情報の欠落が起こり、異論のない整合の取れたマッピングを実現することは困難な状況にある。

2.3 博物館分野におけるメタデータの現状と問題点

博物館は歴史・芸術・民族・産業・自然科学等、人類の営みを表現した有形・無形の文化財を幅広く取り扱う。そのため、古くから台帳あるいは目録という形で文化財に関するメタデータ（博物館メタデータ）が管理されてきた。

筆者らは、博物館メタデータの利用がどのように行われているのか、博物館で利用されている国内外約 20 の収蔵品管理用のシステムについて調査を行った [15]。その結果、表 2 に示す情報が一般的な収蔵品管理システムが管理する情報であることがわかった。

一般的な収蔵品管理システムが取り扱う博物館メタデータは、収蔵品そのものに関するメタデータと収蔵品を操作する際に生じる履歴メタデータに大別することができる。

具体的には、前者は作品/作者名、物理的なサイズ、材質や数量、テープや CD-ROM 等の記録媒体情報、收

表 2: 標準的な収蔵品管理システムの管理項目

情報名称	内容
収蔵品情報	作品名、作者名、制作年、サイズ等
人物情報	登録者、修復者、研究者等関係者情報
メディア情報	テープや CD-ROM、DVD 等の媒体管理
保険情報	評価・保険情報
図書情報	カタログ、出版物、参考文献管理情報
用地情報	不動産文化財情報
会計情報	会計システムと連携した会計情報
貸出・借入情報	展示会等による収蔵品の貸借管理情報
展示会情報	展示会行程情報
イベント情報	関連展示会情報

蔵品評価情報といった現物に関するメタデータがこれに当たる。また、後者は博物館が業務を遂行する上で基盤となり、かつ時系列に並べることができるメタデータ(例えば、ある収蔵品の受け入れから破棄までの履歴情報や、展示会情報、展示会等による収蔵品の貸借情報等)と言える。

また、これらの博物館メタデータとリンクする形である特定の分野(美術・民族・工芸・建築・考古学等)に特化したソースや辞書を有するシステムもあり、これらも含めて博物館メタデータと言うことができる。

1990年代以降、データの相互利用や業務遂行を目的とした博物館メタデータの記述方式の統一化が見られる。これまで、各博物館はそれぞれ独自の記述項目・記述形式で博物館メタデータを蓄積し、システム化してきた。しかし、より効率的な博物館の業務遂行や博物館間での情報共有の必要性から、標準的な博物館メタデータの利用が求められている。

具体的に検討されている博物館分野におけるメタデータスキーマの動向を表3に整理する。

これまで種々の博物館用標準メタデータスキーマを見てきたが、博物館分野においても、前述のメタデータが抱える技術的課題が内在する。特に、博物館間における情報共有の必要性の高まりから、メタデータ・マッピングに関する話題は議論を呼んでおり、いくつかの関連研究も見られる[21][22]。

3 ファジィ・スキーマの検討

3.1 ファジィ・スキーマの提案

これまで、メタデータとメタデータスキーマ、特に博物館に特化した分野における問題点と関連動向を概観してきた。

そこでは、メタデータスキーマを構築する難しさ、メタデータを作成する手間、既に作られているメタデータスキーマを相互に活用する難しさ、が問題となっている。

今後、これらの問題に加えて、コンテンツの制作者・提供者からコンテンツの消費者までシームレスなコンテンツ流通を促進するために、コンテンツの権利を明確にするとともに、メタデータの信頼性を保障する技術の確立も必要となる。

人間があるコンテンツを利用する場合、人間の思考は目的のコンテンツに対するメタデータスキーマ(想起される情報群)を状況や環境、経験に応じて動的に都度構築し、そこで定義されたメタデータを必要に応じて柔軟に取捨選択することで緩やかなチューニングを行っていると考えられる。

これに対して、現存するコンテンツ管理の手法を概観してみると、コンテンツには予め決められたメタデータスキーマに基づいたメタデータが付与されており、その内容項目の変更や改編等は通常行われない。つまり、そこには人間が思考する際に創り出すような、ダイナミックで柔軟性に富んだメタデータスキーマは存在せず、固定で柔軟性に乏しいメタデータスキーマが存在するのみである。そしてこのことが、メタデータの利用における大きな障害となっている。

あらかじめ決められたメタデータスキーマを「静的なメタデータスキーマ」と呼ぶならば、人間が想起する柔軟なメタデータスキーマはコンテクストに応じた「動的なメタデータスキーマ」ということができよう。本研究では、双方のメタデータスキーマの定性的性質を捉え、どこまでが静的に管理されるべき体系であり、どこからがコンテクストに応じて動的に創り出される体系であるのか、その最適値の算出を定量的に行うことを考える。そして、その最適値算出方法をベースとし、必要に応じて静的なメタデータスキーマと動的なメタデータスキーマを緩やかに、複合的にチューニングできる「ファジィ・スキーマ」という概念を提案する。

具体的には、まず実際に運用されている博物館用のメタデータスキーマの定性的な特性を調べ、いくつかのマッピング・パターンに分類する。更に、スキーマが定義する各項目が他の項目とどの程度類似している

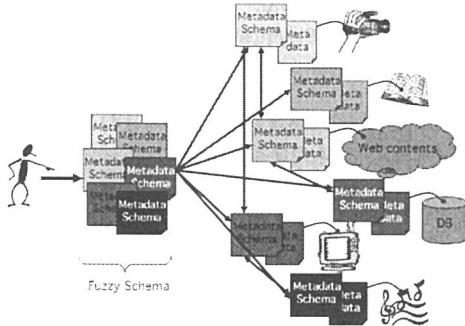


図 2: ファジィ・スキーマに基づくコンテンツ管理

のか、あるいは類似していないのかを示す尺度「あいまい度」を定義し、マッピング・パターンと組み合わせることでメタデータスキーマを複合的に作成する。

組み合わせによる項目の再利用を行うことで、メタデータスキーマ構築及びメタデータ作成の負荷軽減が図れることが期待できる。併せて、あいまい度を用いて詳細にメタデータスキーマをチューニングすることでメタデータ・マッピングの精度向上が期待できる(図2参照)。

3.2 ファジィ・スキーマの言語設計

本研究ではファジィ・スキーマを表現するために、マッピング・パターン及びあいまい度を XML 形式で記述表現する言語の設計を検討している。本稿では、マッピング・パターンに関する検討内容を示す。

マッピング・パターンとは、写像のパターンのことを指し、本稿ではメタデータスキーマで定義されるメタデータの項目対応のパターンを指す。前述の通り、これまでにも項目の対応付けは行われてきたが、数学的には全射や単射が中心であり、かつ人手を介する写像であったため、項目の不一致等が見られた。

そこで、本稿ではまず項目対応時に考えられる写像のパターンを想定し、そのパターンを記述できる言語を設計する。

マッピング・パターンを想定するにあたり、1) 博物館で利用されている収蔵品管理システムが管理するデータ項目 [23], 2) 国際・業界標準とされるデータ項目(表3)、の2点に着目した。

1) の収蔵品管理システムは国内外約 20 のシステムを調査したが、標準的なシステムはいずれも国際・業

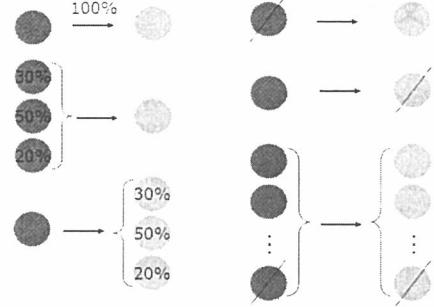


図 3: マッピング・パターン例

界標準に準拠している場合が多いため、本研究では特に 2)を中心検討を行った。

図3に想定されるマッピング・パターンの例を示す。データベースやアーカイブは異なる設計方針に基づき設計されているため、様々な異種性(メタデータスキーマの違いやデータそのものの違い)が生じる。個々のデータベースにおける記述能力を明らかにする狙いとして、関係モデルにおけるデータ異種性の分類[24]が知られている。ここでは、この分類をベースにマッピング・パターンの分類を試みた。パターンは大きく3種に分類される。つまり、1対1でマッピングできるもの、複数個の項目がある割合で1つの項目にマッピングできるもの、片方には存在せず(空集合)もう片方に存在するもの、である。特に、複数個の項目がある割合で1つの項目にマッピングされる例に関して、割合の配分は項目の類似度合を定量的に表現するあいまい度を適用することで解決を図る。

4 まとめと今後の課題

本稿では、博物館を中心とするメタデータの技術的課題を整理した。そこでは、メタデータスキーマ構築の問題、メタデータ作成の問題、メタデータ・マッピングの問題が存在することがわかった。そこで、これらの課題を解決するため、複合的にメタデータスキーマを構築する手法である「ファジィ・スキーマ」という概念を提案した。更に、ファジィ・スキーマを構築するための言語設計の検討を行った。

今後は、マッピング・パターンの具体的な記述方法及びあいまい度の定量化アルゴリズムの検討を行っていく予定である。

参考文献

- [1] Dempsey *et al.*: “*Metadata: A Current View of Practice and Issues*”, J. of Documentation, Vol. 54, No. 2, pp. 145-172(Mar. 1998).
- [2] 経済産業省商務情報政策局監修: “デジタルコンテンツ白書 2005”, デジタルコンテンツ協会(2005)
- [3] 日本図書館情報学会研究委員会編: “図書館目録とメタデータ”, 勉誠出版(2004)
- [4] ISAD(G):General International Standard Archival Description(参照 2006-11-15)
URL:http://www.ica.org/biblio/isad_g_2e.pdf
- [5] Dublin Core Metadata Initiative(参照 2006-11-15)
URL:<http://dublincore.org/>
- [6] MPEG-7(参照 2006-11-15)
URL:<http://www.itscj.ipsj.or.jp/mpeg7/>
- [7] TV-Anytime(参照 2006-11-15)
URL:<http://tv-anytime.org/>
- [8] 北村他: “電子出版とXMDF技術”, シャープ技報, 第84号(2002-12)
- [9] Broad Band e-Book(参照 2006-11-15)
URL:<http://ps.canon-sol.jp/bc/kikaku.htm>
- [10] MusicBrainz(参照 2006-11-15)
URL:<http://musicbrainz.org>
- [11] The Open Archives Initiative Protocol for Metadata Harvesting(参照 2006-11-15)
URL:<http://www.openarchives.org/>
- [12] 文化遺産オンライン(参照 2006-11-15)
URL:<http://bunka.nii.ac.jp/>
- [13] Cheng Zhu *et al.*: “*Ontology Mapping For Interaction in Agent Society*”, Proc. of IEEE Int'l Conf. on SCC'04, pp. 619-622(15-18 Sept. 2004).
- [14] Diane I. Hillmann, Elaine L. Westbrooks: “*Metadata in Practice*”, American Library Association(2004)
- [15] 秋元他: “博物館情報の相互利用を目的とした文化財情報システムの提案”, 情処シンポジウム「じんもんこん 2004」,(2004-12)
- [16] ミュージアム資料情報構造化モデル (参照 2006/11/15)
URL:<http://webarchives.tnm.jp/docs//informatics/smmoi/>
- [17] ICOM-CIDOC IC(参照 2006-11-15)
URL:<http://www.willpowerinfo.myby.co.uk//cidoc/>
- [18] The CIDOC Conceptual Reference Model(参照 2006-11-15)
URL:<http://cidoc.ics.forth.gr/>
- [19] MDA SPECTRUM(参照 2006-11-15)
URL:<http://www.mda.org.uk/spectrum.htm>
- [20] Categories for the Description of Works of Art (CDWA) (参照 2006-11-15)
URL:http://www.getty.edu/research//conducting_research/standards/cdwa/
- [21] 原ほか: “データベースの共有におけるデータマッピングの事例的研究”, 情処研報, Vol. 2005, No. 76, pp. 31-38(2005-07).
- [22] 山田ほか: “博物館情報横断検索における分散オンライン検索”, 画像電子学会研究会 (2006-03).
- [23] 秋元: “博物館の収蔵品管理におけるメタデータの利用と問題点”, 情処研報, Vol. 2004, No. 36, pp. 55-62(2004-03).
- [24] Kim, W *et al.*: “*Classifying Schematic and Data Heterogeneity in Multidatabase Systems*”, Computer, Vol. 24, No. 12, pp. 12-18(1993).

表 3: 博物館分野におけるメタデータスキーマ

名称	内容
ミュージアム資料情報モデル [16]	東京国立博物館が国内向けに提案。人文科学系博物館の移動可能な有形文化財が対象。博物館収蔵品に関する業務支援と博物館同士の情報共有の基盤提供が目的。4特性、34属性、108詳細要素からなる。また、RDF Schemaによるボキャブラリが策定されている。国際標準との互換あり。2005年発行。
CIDOC/IC[17]	博物館資料情報のための国際ガイドライン。International Council of Museums(ICOM:国際博物館会議)のInternational Committee for Documentation(CIDOC:国際ドキュメンテーション委員会)が策定。メタデータを記述するためのガイドラインとして、資料の受入から管理、廃棄に至るまでの管理プロセスを22の情報グループと74の情報カテゴリに分類して示している。1995年発行。
CIDOC/CRM	「CIDOC Conceptual Reference Model(CRM:概念参照モデル)」[18]。CIDOC ICの情報共有を目的としたモデル。各博物館が持つメタデータの交換・統合を行うためのフレームワークであり、RDF/XML形式で記述ができる。ISO21127:2006。
SPECTRUM[19]	Museum Documentation Association(MDA:博物館ドキュメンテーション協会)が英国内の博物館メタデータの標準化推進のため導入した手続き型モデル。資料の受入から廃棄に至るまでの具体的な資料管理手順20項目が示されている。メタデータの語彙やシソーラス等も含めたMDAデータ標準も提案されている。博物館登録にSPECTRUMを用いることを条件としたため、英国内では約75%の博物館が採用している[3]。2006年現在、最新版はVersion3.0。
CDWA	Categories for Description of Works of Art(CDWA)[20]。米国のGetty Research Instituteが中心となり策定された美術分野におけるメタデータスキーマ。Art Information Task Force(Getty Research Instituteの美術史情報プログラムとCollege Art Associationの共同プロジェクト)が1996年に発行。CDWAをXML Schemaで表現するためのCDWA Lite[?]も開発されている。
Museum Handbook	米国国立公園管理局は1990年、博物館資料の管理、保存、記録化、アクセス等全般的な取り扱いに関するハンドブックを発行している。
AFRICOM Handbook of Standards	盜難、略奪、密輸防止のため、アフリカの約120の博物館が共同で作成したメタデータ記述の標準マニュアル。ICOM/CIDOCが指導協力。1996年発行。
RLG Cultural Materials	米国の約40の参加機関(大学や博物館等)の総計25万点に上るメタデータをEncoded Archival Description(EAD)やMetadata Object Description Schema(MODS)のXML形式で記述することで横断検索が可能なシステムを開発している。
DC.Culture	資料の電子化・識別管理のため、欧州のMINERVAプロジェクトで開発。Dublin Coreをベースとしている。
博物館遺物管理電算化のための遺物分類標準	Dublin Coreをベースとした基本16項目、オプション116項目を要したメタデータスキーマであり、韓国内の博物館や大学等約90の参加機関で共通に利用されている。1995年発行。
Object ID	博物館資料の追跡調査(盗難防止のため)を行えるよう、メタデータとして資料(作品、盗難品)を同定する最低限の情報を記述する。1997年発行。
CIMI profile	Computer Interchange of Museum Information ProfileはZ39.50プロトコルをベースに博物館メタデータを定義している。1995年発行。
VRA	Virtual Resources Association Core Categories for Visual Resources、美術作品記述のためのカテゴリ、2000年発行
AMICO Data Directory	Art Museum Image Consortiumによる美術館画像のためのデータ仕様、2001年発行。