



ネットワーク情報資源の組織化とメタデータ

メタデータ	言語: jpn 出版者: 公開日: 2010-08-12 キーワード (Ja): キーワード (En): 作成者: 吉田, 暁史 メールアドレス: 所属:
URL	http://hdl.handle.net/10466/10962

ネットワーク情報資源の組織化とメタデータ

吉田暁史

1. 膨大なネットワーク情報資源の出現

膨大なネットワーク情報資源が出現した現在、その中から良質な情報資源を選択し、提供する図書館サービスは必須と考えられる。良質な情報資源を提供する仕掛けとして、ネットワーク情報資源の組織化が重要である。ネットワーク情報資源組織化のためのプロセスを概観し、問題点とあるべき方向性について考える。

2. ネットワーク情報資源の特徴

ネットワーク情報資源の特徴としては、以下のようなことがあげられる。

- (1) 情報資源の量は膨大である。
- (2) 内容は玉石混淆である。
- (3) 存在が不安定である。

こういった把握しがたい情報資源の中から有益な情報を発見し、提供することが求められる。

3. 図書館におけるネットワーク情報資源

3-1. 既存の探索ツールとその問題点

既存の一般的探索ツールとしては、下記のようなものがある。

- A. 商用、汎用検索サービスの利用
- B. リンク集の利用
- C. 非インターネットのツールの利用

上記のうち最もよく用いられる商用検索サービスは、検索結果が膨大になったりするし、検索精度も粗く、良質の情報だけを効率よく検索することは難しい。BやCは、網羅性や最新性が弱いという問題がある。したがって、各図書館あるいは、図書館群の利用者にとってふさわしい、有益な情報資源に絞り込んで提供する検索システムを、図書館コミュニティが独自に構築する必要がある。

3-2. ネットワーク情報資源と所蔵資料

次に、ネットワーク情報資源と従来の所蔵資料との比較を考える。所蔵資料の場合は、選択→収集→組織化→配架→提供→保存、という流れになり、ネットワーク情報資源の場合は、これらの流れのうち、収集、配架、保存という段階がなく、選択→組織化→提供という流れになる。なおネットワーク情報資源の収集、保存も技術的には可能であり、国立図書館を中心にすでにアーカイブ実践も始まっているが、ここではこの問題に立ち入らない。両者ともに図書館利用者が等しく利用することの出来る資源であり、両者をシームレスに検索し、アクセスできるように組織化すべきだと考えられる。したがって従来の所蔵資料の目録と、ネットワーク情報資源の目録とを統合し、一体として検索・提供できる仕組みが必要だと思われる。いわば従来の所蔵目録の拡大版と捉えることができよう。

帝塚山学院大学 教授

3-3. ネットワーク情報資源を扱う上での問題点

ネットワーク情報資源を扱う上での問題点を次の7点にまとめる。

- ①情報内容が玉石混淆であり、情報内容の選択機能が必要であること。
- ②情報資源の量が膨大であり、さらに増加の一途であること。したがって資源の選択機能とともに、組織化に際して協力体制が必要となろう。
- ③所在位置の変化、情報内容の変更、削除、など種々の不安定性があり、これらの監視機能が重要であること。
- ④静止画や音声といった記録様式が多様であり、目録上でどのように扱うかという問題がある。
- ⑤情報資源があらゆる階層（粒度）に存在し、目録上の問題が生じる。
- ⑥特別なツールの必要となるケースがあり、技術的な問題が起こる。
- ⑦著作権処理の問題があり、利用形態上の対応が必要となる。

4. サブジェクト・ゲートウェイにおける情報資源の選択、目録等の過程

4-1. サブジェクト・ゲートウェイとは

前章までに述べたようなネットワーク情報資源を組織化する現実の仕組みとして、サブジェクトゲートウェイとよばれるものがある。これは必ずしも図書館の世界だけとはかぎらないが、おおむね図書館界を中心に実践されている。特に欧米では多数の実施例があり、多くの図書館が協力して組織的に行われている。種々のサブジェクトゲートウェイにおける実践活動の調査結果をもとに、サブジェクトゲートウェイにおける情報資源の現実の組織化過程を概観する。

4-2. ネットワーク情報資源の選択的収集

選択基準としては下記があげられる。

- (1) 内容、(2) 著者の信頼性、(3) 情報の鮮度、(4) 情報資源全体の構成は明確か、(5) デザインはわかり易いか、といったこと。

4-3. メタデータ・データベースの構築

ゲートウェイサービスでは、情報資源を確実に把握し、的確な検索が行えるようにするため、メタデータを付与しメタデータ・データベースの構築を行う必要がある。その構築に要するため、(1) 情報資源の書誌記述方式、(2) 主題索引の方式や典拠コントロールの方法、(3) データの収集方法、(4) データの選択基準の作成、(5) メタデータ・データベースの検索インターフェースの設計、といったことを決定する必要がある。

5. 情報資源組織化としてのメタデータ

5-1. メタデータとは何か

ここで情報資源組織化としてのメタデータについて考える。メタデータの定義はさまざまに行われているが、最も一般的には、「データに関するデータ」あるいは「データに関する構造化されたデータ」という定義となろう。しかし、メタデータの対象は事実上、デジタル情報資源、特にネットワーク情報資源に限定されているように思われる。

5-2. メタデータの種類

メタデータは非常に広い分野で用いられるものであり下記のような種類がある。

- (1) 利用目的によって分ける方法
 - A. 管理的なもの (administrative)
 - B. 記述的なもの (descriptive)

- C. 保存的なもの(preservation)
- D. 技術的なもの (technical)
- E. 利用的なもの(use)

利用目的によって分けると、Aは管理的な観点から必要なデータを扱う。権利関係、アクセスへの法的な問題、等を中心にしたもの。Bは情報資源を特定するためのいわゆる書誌記述が中心であり、目録レコード、関連資源へのリンクなどを含む。Cは情報資源を保存するために必要なデータを扱い、資源の物理的状态などを中心としたもの。Dはコンピュータシステムが扱うために必要なデータを中心としたもの。Eは実際に利用するための使用条件などを規定するメタデータである。

現実のメタデータシステムは、例えば記述的なものを中心とし、他のものも従属的に含むようなものということになるが、図書館界におけるメタデータはやはりCの記述的なものが基本だと思われる。

(2) データの複雑さ、構造化の程度によって分ける方法

	第1レベル	第2レベル	第3レベル
レコード	単純フォーマット	構造化フォーマット	高度なフォーマット
特性	独自 フルテキスト索引	新興標準 フィールド構造	国際標準 精緻なタグ付け
使用例	Lycos AltaVista Yahoo	Dublin Core IAFA template RFC 1807 SOIF	CIMI EAD TEI Header MARC

メタデータを、データの複雑さ、構造化の程度によって分ける方法がある。精粗3つのレベルで区分する。それぞれ、単純なものから第1レベル、第2レベル、第3レベルと名付ける。第1レベルの特徴はレコード形式が単純フォーマットであるということであり、特性として、独自形式で、フルテキストから索引を作成することがあげられる。使用例はLycosなどの検索エンジンであるが、事実上構造化されておらず、最もプリミティブなメタデータでしかない。第2レベルの特徴は、レコード形式は構造化フォーマットが用いられ、フィールド構造を備えているもの。第3レベルでは、レコード形式として高度なフォーマットが用いられ、精緻なタグ付け等が行われるもので、実例としてはMARCやTEIヘッダがある。

図書館界で、ネットワーク情報資源の目録をとるときに使われることの多いダブリンコアであるが、第2レベルという位置付けになると思われる。また、われわれが日常用いている目録はMARC形式であるが、これはダブリンコアなどよりはるかに詳しく、かつ精密に構造化されており、美術館におけるCIMI、文書館におけるEADなどとともに、第3レベルに属するとみなされる。

5-3. 図書館界に相応しいメタデータ

(1) 図書館で扱うメタデータとは

ダブリンコアが事実上の標準仕様になりつつあるが、図書館界において対象とする主たる情報資源と、それにふさわしいメタデータは何かを考える。まず扱う対象は基本的にドキュメント類似のオブジェクト (DLO)、つまり文書的情報資源が中心であると考えられる。もっともテキスト情報以外にも、画像情報、音声情報も当然扱われるが、それらの記述内容は文書情報の延長上にあると考えられる。

(2) 従来 of 目録規則では何が困るのか。

従来 of 目録規則は次のような問題点を持っている。

- A. ファイル形式などデジタル情報特有 of 情報を適切に記述できない
 - B. 著作権処理などの管理データを表す適切な記述要素がない
 - C. データ要素 of 不安定さへの対応が不十分
 - D. タイトルや著者などが把握しにくいケースが多い
 - E. 物的な側面を表す必要がない、あるいは表現できないことが多い
- といったことである。

こういった欠点を持つがゆえに、電子資料とりわけネットワーク情報資源 of 記述が困難であった。このような問題点を取りあえず回避するために、サブジェクト・ゲートウェイ等ではダブリンコアが採用されたということであろう。ダブリンコアが情報資源 of 構造分析をしっかり行っており、論理的に的確に記述しようという理由ではなく、それがまるでおおざっぱなものであるがゆえに、苦勞せずすむという消極的な理由からであったと思われる。

(3) 今後どうあるべきか

以上 of 結果、次のようなことが必要であろう。

- (A) データ要素 of 不安定さを適切に処理する仕組み。
- (B) 物としての側面と、内容としての側面を切り分けることにより、両者を分離して記述する枠組みを作る。
- (C) 作業性を考慮して、必須データ要素、オプション要素、といったランク付けを行うこと。

こういった方法で従来 of MARC をより論理的な構造に変更し、また実務的側面を補うことにより、MARC 形式でも十分に記述が可能になると考えられる。図書館界においては、所蔵資料とネットワーク情報資源とをシームレスに検索・アクセスできなければならないとしたが、そのためにも両者の目録基盤は共通でなければならない。改良された MARC 形式を基本的に双方で使用するべきだと考える。

5-4. 他 of コミュニティとの連携を行うためのメタデータ of 条件

(1) 各コミュニティにおけるメタデータの相違

コミュニティ	主として用いられるメタデータ
図書館	MARC
文書館	EAD
博物館・美術館	CIMI

他 of コミュニティとの連携という観点を考慮して、コミュニティ内、コミュニティ間という2つの局面でのメタデータ of 条件について考える。コミュニティ間連携 of 例として、図書館、文書館、博物館・美術館をあげる。コミュニティ内部で主として用いられるメタデータは、図書館では MARC である。文書館では EAD が用いられている。文書館でも文書情報が中心であるが、オリジナルな資料そのものを扱い、しかも資料群としての来歴情報などを重視するという違いがあ

りこういった諸点に対応するものとなっている。次に、博物館・美術館では CIMI が用いられている。ここでは文書情報以外に 3 次元資料や画像情報など多様な形式を扱い、かつオリジナル資料を扱うという特色がある。

(2) コミュニティ内部および相互流通のためのメタデータ

所蔵資料に対しては MARC を用い、ネットワーク情報資源に対してはダブリンコアを用いるという方向ではなく、コミュニティ内部ではそれぞれの領域特有の高度な記述内容が求められるために、詳細な第 3 レベルを用い、コミュニティ間における情報流通のためには、大雑把な自由記述形式であるダブリンコアのような第 2 レベルを用い、第 2 レベルにおいて各フォーマット間の差異を吸収していくという方向性が望ましいと考えられる。

6. メタデータにおけるセマンティックスとシンタックス

目録規則、そしてより広義にはメタデータの仕組みを考える場合、近年ではセマンティックスとシンタックスという 2 つの軸に分けて考えるようになってきている。

・セマンティックス：データ要素の定義と意味内容。

図書館目録では、ISBD にほぼ相当する部分。

・シンタックス：データ要素を記載する文法。

図書館目録では、MARC フォーマットにほぼ相当する部分。

これら両者を切り分けて理解することが必要である。以下では典型的なセマンティックス規則、およびシンタックス規則のいくつかを紹介する。今後は、セマンティックスを定める規則とシンタックスを定める規則とを使い分けて採用し、それぞれ標準規約にもとづいて目録を作成していくことになる。

6-1. ダブリンコア

・Dublin Core Metadata Element Set の略である。

・15 の記述要素を定める。それぞれの記載方法は自由記述方式である。

・セマンティックスのみを定め、シンタックスに関しては他の基準を採用する。

・DCQ(Dublin Core Qualified)と DCS(Dublin Core Simplified)がある。

qualifier を使う場合を DCQ といい、ダム・ダウン原則 (Dumb-Down Principle) を持つ。

qualifier を使わない場合を DCS という。

・以下のような 15 の基本エレメントセットから成る。

Title, Creator, Contributor, Type, Publisher

Date, Format, Language, Description, Subject

Relation, Identifier, Source, Coverage, Rights

6-2. MODS (Metadata Object Description Schema)

・LC が策定する図書館系の各種メタデータ形式を統合するような構造記述。簡単なセマンティックスを定めると同時に、シンタックス面を詳細に定めている。書式表現は XML による。全体としてはシンタックス的な枠組みと考えられる。

・MARC のサブセットとなっており、DC のような簡略形式から、MARC21 のような詳細形式までに対応する。MARC との親和性が高いが、MARC ほど複雑ではないがゆえに、記述を作りやすい。

・単独でも使えないことはないが、基本的には MARC をはじめとして、図書館系のさまざまなメタデータ形式を吸収し、一つに統合する受け皿的な役割を果たす。

・MODS の構造

主要エレメントとして下記を持つ。

titleInfo [mandatory], name, typeOfResource
Genre, originInfo, physicalDescription
Language, abstract, tableOfContents, note
Subject, classification, relatedItem, identifier
targetAudience, accessConditions
extension, recordInfo

そのほか、各々はサブエレメントを持つ。主要エレメントは ISBD でいうエリアに、サブエレメントは要素にほぼ相当する。さらに主要エレメントおよびサブエレメントは属性を持ち、より詳細なデータを記述しうる。

6-3. METS (Metadata Encoding and Transmission Standard)

- ・ LC が策定する、メタデータを管理するための技術手法。シンタックスのみを定める。
- ・ デジタル対象物をとらえる、さまざまなメタデータ群を、一つの様式で表現する。
つまりひとつの記述対象が、異なる記述的メタデータと管理的メタデータを持つことがあるが、これらを統一的な記述手法を用いてひとまとめにすることによって、管理を確実にするための技術的な手法である。

6-4. RDF (Resource Description Framework)

- ・ W3C が制定した、メタデータを記述するための枠組みを提供する、純粋にシンタックス的な規約。記述的メタデータのみならず、産業界を含めてあらゆる方面に適用できる。
- ・ 名前空間機能を使うことによって、複数のスキーマを使用してメタデータを表現することができる。
- ・ 記載のためには XML を用いる。

7. 主な略語一覧

ダブリンコア	Dublin Core Metadata Element Set
TEI	Text Encoding Initiative
EAD	Encoded Archival Description
CIMI	Consortium for the Computer Interchange of Museum Information (現在は CIMI が正式名称)