

## データベース共有における データマッピングの事例的研究

原 正一郎<sup>(1)</sup>、相田 満<sup>(2)</sup>、入口 敦志<sup>(3)</sup>、江戸 英雄<sup>(2)</sup>、五島 敏芳<sup>(4)</sup>、山田 直子<sup>(3)</sup>

<sup>(1)</sup>国文学研究資料館・複合領域研究系  
<sup>(2)</sup>国文学研究資料館・文学形成研究系  
<sup>(3)</sup>国文学研究資料館・文学資源研究系  
<sup>(4)</sup>国文学研究資料館・アーカイブズ研究系

多様な人文科学情報の横断検索を目指した「資源共有化システム」の開発を行っている。資源共有化システムの特徴は、データ構造を Dublin Core メタデータにより、検索手順を Z39.50 により標準化し、データベースシステムの差異を乗り越えようとしている点にある。そのため、各データベースのレコード・フィールドと Dublin Core メタデータの要素とのマッピングが、資源共有化システムの検索精度を左右する。本研究では、国文学研究資料館の国文学論文目録データを DublinCore メタデータへマッピングする実験を通じ、Dublin Core メタデータの有効性・問題点および解決法について議論する。

### Case Study on the Data-Mapping for the Resource Sharing System

Shoichiro HARA<sup>(1)</sup>, Mitsuru AIDA<sup>(2)</sup>, Atsushi IRIGUCHI<sup>(3)</sup>, Hideo EDO<sup>(2)</sup>,  
Haruyoshi GOTHO<sup>(4)</sup>, Naoko YAMADA<sup>(3)</sup>

Department of Interdisciplinary Studies, National Institute of Japanese Literature  
Department of Literary Development Studies, National Institute of Japanese Literature  
Department of Literary Resource Studies, National Institute of Japanese Literature  
Department of Archives Studies, National Institute of Japanese Literature

The Resources Sharing System has been developed to share various humanities information among databases. The peculiarity of the Resources Sharing System is its promotion of standardization by introducing Dublin Core metadata and Z39.50 to overcome heterogeneities of record structure and retrieval procedure among databases. Thus, the mapping between original database record fields and Dublin Core metadata record fields is crucial to retrieval accuracy of Resources Sharing System. In this paper, records of the Catalog Data of Research Papers on Japanese Classical Literature were converted into Dublin Core metadata, and the validity, problems and their solutions on Dublin Core and its conversion procedure are discussed.

**Key Words:** 資源共有化システム, メタデータ, ダブリンコア, Z39.50, 国文学研究資料館, 論文目録データベース, MARC21, Resources Sharing System, metadata, Dublin Core, National Institute of Japanese Literature

#### 1. はじめに

国文学研究資料館（以下、国文研）では目録データをはじめアーカイブズ、全文、画像など多様なデジタルデータの形成を推進している。これらは個別データベースシステムであるため、関連資料を調べるために幾つものデータベースを検索しなければならないなどの問題が指摘されていた。資源共有化システムは、このような問題の解決を目指して開発中のシステムである<sup>(1)</sup>。資源共有化システムは、Dublin Core メタデータ（以下 DC メタデータ）<sup>(2)</sup>と Z39.50<sup>(3)</sup>を利用した、プロトコルレベルにおける分散データベースの統合検索システムである。つまり Z39.50 により検索手順を、DC メタデータによりレコード構造の違いを吸収する。これにより利用者は、多様なデータベースを一つのデータベースシステムとみなして検索することが可能となる。国文研の館蔵資料目録・画像・アーカイブズ・OPAC などのデータベースは、資源共有化システムの下で統合されつつある。まもなく国文研のデータベース利用者は、データベースの所在・メディア・データ構造・検索法の違いを意識することなく、国文研の全データベースを統合検索することが可能となる。もし国文学研究資料館以外の大学、研究機関、博物館、文書館などが同様の資源共有化システムを導入すれば、各機関のデータベース利用者は、これらの機関が有する多種多様なデータベースを統合的に検索することが可能となる。そこで総合研究大学院大学に参加している人文科学系大学共同利用機関を中心として、資源共有化システムによるデータベースの機関間連

表 1. 論文目録データのレコードフィールド一覧

出現順	要素名	説明	文字種(全角)	固定	文字数	補足
1	冊子年	国文学年鑑記載年	かな,漢字,数字	固定	2 0	
2	入力者	データ入力者氏名	かな,漢字	固定	5	5文字以上の場合はカットしている
3	入力日	データ入力日付	数文字	固定	8	yyyymmdd
4	更新1	更新履歴1(専門員)	数文字	固定	8	yyyymmdd
5	更新2	更新履歴2(監修)	数文字	固定	8	yyyymmdd
6	更新3	更新履歴3(公開後)	数文字	固定	8	yyyymmdd
7	漢字	表示できない漢字等のメモ	文字	固定	2 0	原情報はここへ。関連情報は「メモ」へ
8	メモ	作業の備考	文字	固定	3 0	
9	連番	ユニークキー	数文字	固定	7	0から開始することを許す。
1 0	旧連番	以前のユニークキー	数文字	固定	1 5	
1 1	時代分類	時代による配列用番号	漢字	固定	5	選択肢
1 2	時代分類番号		数文字	固定	2	0から開始することを許す
1 3	分野1	分野1による配列用番号	漢字	固定	8	選択肢
1 4	分野1番号		数文字	固定	2	0から開始することを許す
1 5	分野2	分野2による配列用番号	文字	固定	1 0	選択肢
1 6	分野2番号	作者・作品名	数文字	固定	2	0から開始することを許す
1 7	分野3		文字	固定	1 0	氏名の場合,姓と名の間に「/」,反復はなし。
1 8	分野3よみ		かな	固定	2 0	氏名の場合,姓と名の間に「/」,反復はなし。
1 9	分野4		文字	固定	1 0	選択肢
2 0	分野4番号	分野4による配列用番号	数文字	固定	2	0から開始することを許す
2 1	分野5	作者・作品名	文字	固定	1 0	氏名の場合,姓と名の間に「/」,反復はなし。
2 2	分野5よみ		かな	固定	2 0	氏名の場合,姓と名の間に「/」,反復はなし。
2 3	分野6		文字	固定	1 0	選択肢
2 4	分野6番号	分野6による配列用番号	数文字	固定	2	0から開始することを許す
2 5	分野7	作者・作品名	文字	固定	1 0	氏名の場合,姓と名の間に「/」,反復はなし。
2 6	分野7よみ		かな	固定	2 0	氏名の場合,姓と名の間に「/」,反復はなし。
2 7	頭書		かな,カナ,漢字,記号	可変	-	
2 8	主題		かな,カナ,漢字,記号	可変	-	
2 9	副題		かな,カナ,漢字,記号	可変	-	
3 0	題名	頭書+主題+副題	かな,カナ,漢字,記号	可変	-	データの整合性チェックはしない
3 1	英文タイトル		文字	可変	-	
3 2	執筆1	執筆者表記	文字	可変	-	姓と名の間は「/」で区切る。複数著者の場合は「%」で区切る。名前以外の文字で使えるのは,「%記」と「%他」のみが最後に現れることがある。
3 3	執筆2	執筆者よみ	かな,アルファベット	可変	-	姓と名の間は「/」で区切る。複数著者の場合は「%」で区切る。「%記」と「%他」は出現しない。
3 4	英文執筆		アルファベット,記号	可変	-	姓と名の間は「」で,複数著者の場合は「%」で区切る。
3 5	種別	雑誌・単行本・新聞別	漢字	可変	-	
3 6	請求	請求記号	かな,カナ,数字文字,記号	固定	1 6	単行本の区切りには「-」を使う。例えば,“イ9-102-27A”
3 7	誌著名	雑誌・単行本・新聞のタイトル	文字	可変	-	単行本には「』」を付す。新聞の場合は朝刊・夕刊の別を示す。
3 8	英文誌	誌著名の英文タイトル	アルファベット,記号	可変	-	単行本には「』」を付す。新聞の場合は朝刊・夕刊の別を示す
3 9	巻号		数字文字,漢字,記号	固定	7	巻と号の間に「/」,新聞の場合は日付と朝夕の別。例「三夕」
4 0	通巻		数字文字,漢字	固定	4	
4 1	開始頁		数字文字	固定	4	
4 2	終了頁		数字文字	固定	4	
4 3	総頁		数字文字	固定	4	終了頁-開始頁+1
4 4	和暦年		数字文字,アルファベット	固定	3	1文字目は元号を示すアルファベット,続いて年号を示す2桁の数字。
4 5	月		数字文字,漢字	固定	3	2桁の数字。例(01)。これに春夏秋冬などの文字が続く場合がある。
4 6	日		数字文字	固定	2	2桁の数字。例(01)。
4 7	西暦年		数字文字	固定	4	
4 8	翻複1	復刻・複製表記	かな,漢字	可変	-	複数ある場合は区切りに%を用いる。
4 9	翻複2	復刻・複製表記のよみ	かな	可変	-	複数ある場合は区切りに%を用いる。
5 0	作品名		文字	可変	-	複数ある場合は区切りに%を用いる。
5 1	作者名		文字	可変	-	複数ある場合は区切りに%を用いる。
5 2	概念	主題・概念事項	文字	可変	-	複数ある場合は区切りに%を用いる。
5 3	要約		文字	可変	-	
5 4	キー	キーワード	文字	可変	-	複数ある場合は区切りに%を用いる。
5 5	作成年度	冊子体用抽出年度	数字文字,漢字	固定	3	例(H01)。
5 6	作品名よみ		かな,漢字	可変	-	複数ある場合は区切りに%を用いる。
5 7	作者名よみ		かな,漢字,文字	可変	-	複数ある場合は区切りに%を用いる。

携の実現を目指したコラボレーション・プロジェクトを開始した<sup>[4,5]</sup>。

プロジェクトの開始時点においてはシステムの接続が中心課題であり、DC メタデータの構築手順（データベースの各レコード・フィールドを DC メタデータのどのエレメントと対応させるか：以下ではマッピング）に関する検討は不十分であった。たとえば日付・年代・時代などの時間情報は Date や Coverage へマッピングされるが、どの時間情報をマッピングするかなどについての合意は得られておらず、機関やデータベースごとに ad hoc なものであった。しかしマッピングは資源共有化システムの検索精度を大きく左右する要素である。そこでプロジェクトの第二段階として、マッピングについての検討を開始した。本稿では、国文学研究資料館の国文学論文目録データ（以下、論文目録データ）<sup>[6]</sup>を DC メタデータへマッピングする実験を通じて、DC メタデータの有効性・問題点および解決法について議論する。

## 2. 実験の概要

DC メタデータへのマッピングに関する問題点を明らかにするために、国文学研究資料館の論文目録データを用いた実験を行った。論文目録データは日本文学研究論文の総合目録データベースの元となるデータであり、日本国内で発表された雑誌紀要単行本（論文集）などに収められた論文に関する書誌情報を登録している。なお論文目録データには論文と単行本の二種類の情報が登録されているが、ここでは論文の情報のみを対象とした。表 1 に論文目録データのレコード・フィールド定義の一覧を示す。

マッピング実験には、国文研において目録・アーカイブズのデータ形成に従事している教職員 6 名（目録系 4 名、アーカイブズ系 1 名、情報系 1 名）の参加を仰いだ（以下では被験者）。マッピング実験の参考資料として、論文目録データのレコードフィールド一覧表（表 1）、DC メタデータのエレメントに関する資料<sup>[7]</sup>、DC メタデータの Qualifiers に関する資料<sup>[8]</sup>、および MARC21 書誌データフォーマットに関する資料<sup>[9]</sup>を事前に配布した。与えられた参考資料に基づいて、論文目録データの各レコード・フィールド（表 1 の要素名）を可能な限り DC メタデータのエレメントへ対応付けることが、被験者に課せられた最低限の作業内容である。もし DC メタデータの Qualifiers レベルを考慮できる場合は、エレメント名に Qualifiers を含む形でマッピングを行ってもらった。また MARC の知識を持つ被験者には、論文目録データの各レコード・フィールドを可能な限り MARC21 のエレメントへ対応づける作業も行って貰った。なお論文目録データのレコード・フィールドと DC メタデータのエレメント間の対応関係は多対多とした。

## 3. 実験の結果

表 2 はマッピング実験の結果を被験者ごとにまとめたものである。表中の行方向（冊子年～作者名よみ）は論文目録データのレコード・フィールド（表 1 の要素名と同じ）を、列方向は被験者（A1～A6）を表している。セルの内容は DC メタデータのエレメント名である。つまり被験者が論文目録データのレコード・フィールドを DC メタデータのエレメントにマッピングした結果である。例えば「冊子年」行、「A1」列のセルの内容は「date」となっているが、これは被験者 A1 が論文目録データの「冊子年」フィールドを DC メタデータの「date」エレメントにマッピングしたことを示す。マッピング関係が示されなかったセルの内容は空白となっている。論文目録データの一つのレコード・フィールドが DC メタデータの複数のエレメントへマッピングされた場合、各エレメント名はコンマ「,」で区切って記述される。例えば「分野 1」行、「A1」列のセルの内容は「subject, type」となっているが、これは被験者 A6 が論文目録データの「分野 1」フィールドを DC メタデータの「subject」エレメントと「type」エレメントの両方にマッピングしたことを示す。DC メタデータの Qualifier を指定したマッピングを行った場合、Qualifier は DC メタデータのエレメント名の後ろに [modified] のように記述される。Qualifier を指定したマッピングは被験者の裁量に任せた。ところで被験者 A3 および A6 は、データベースのレコードそのものに関する情報と、レコードのコンテンツに関する情報を区別している。データベースのレコードそのものに関する情報は、DC メタデータのエレメント名の前に「\*」を付けて区別した。このマークが無いエレメント名は、レコードのコンテンツに関する情報である。例えば「冊子年」行、「A3」列のセルの値は「\*date[issued]」となっているが、これは「冊子年」を論文の発行された date[issued]ではなく論文目録レコードが作成された date[issued]であると被験者 A1 が考えたことを示している。被験者 A1、A2、A4、A5 は、そのような区別をしていなかった。

表 2 の内容を、論文目録データのレコード・フィールドと DC メタデータのエレメントのマッピングにおけるバラツキをまとめたものが表 3 である。表中の行方向（冊子年～作者名よみ）は論文目録データのレコード・フィールドで表 2 と同じである。列方向は DC メタデータのエレメ

表 2 . マッピング実験の結果

	A1	A2	A3	A4	A5	A6
冊子年	date		*date [issued]			*date [issued]
入力者		contributor	*contributor		contributor	*contributor
入力日		date	*date [created]	date		*date Created
更新 1				date		*date [modified]
更新 2				date		*date [modified]
更新 3		date		date		*date [modified]
漢字						
メモ						
連番	identifier	identifier	*identifier		identifier	*identifier
旧連番	identifier				identifier	*identifier
時代分類	subject	subject [LCC:P]	coverage [temporal]	subject	coverage	subject, coverage [temporal]
時代分類番号	subject					
分野 1	subject	subject	type	subject	subject	subject, type
分野 1 番号	subject					
分野 2	subject		type	subject	subject	subject, type
分野 2 番号	subject					
分野 3	subject	subject		subject	subject	subject
分野 3 よみ	subject				subject	subject
分野 4	subject		type	subject	subject	subject, type
分野 4 番号	subject					
分野 5	subject	subject		subject	subject	subject, type
分野 5 よみ	subject				subject	subject, type
分野 6	subject		type	subject	subject	subject, type
分野 6 番号	subject					
分野 7	subject	subject		subject	title	subject
分野 7 よみ	subject				subject	subject
頭書	title				title	title
主題	title			title	title	title
副題	title				title	title
題名	title	title	title	title	title	title
英文タイトル	title	subject		title	title	title
執筆 1	creator	creator	creator	creator	creator	creator
執筆 2	creator	creator		creator	creator	creator
英文執筆	creator	creator		creator	creator	creator
種別	coverage			type	type	type
請求	coverage [Spatial]		identifier		identifier	identifier
誌著名	source, publisher, title		publisher	relation [Is Part of]	source	source
英文誌	source			relation [Is Part of]	source	source
巻号	source		identifier		identifier	source
通巻	source		identifier		identifier	source
開始頁	source		identifier			source
終了頁	source		identifier			source
総頁	source		format [extent]	format		source
和暦年	date		date [issued]		date	coverage [temporal]
月	date		date [issued]		date	coverage [temporal]
日	date		date [issued]		date	coverage [temporal]
西暦年	date	coverage [temporal]	date [issued]		date	coverage [temporal]
翻複 1	subject		subject	relation [Is Version of]	subject	relation [Is Versoin of]
翻複 2	title			relation [Is Version of]	subject	relation [Is Versoin of]
作品名	title		subject	subject	subject	subject
作者名	title		subject	subject	subject	subject
概念	title			subject	subject	subject
要約				Description	subject	subject, description
キー				subject	subject	subject
作成年度			*date [issued]		date	*date [issued]
作品名よみ	subject			subject	subject	subject
作者名よみ	subject			subject	subject	subject

表3. マッピング実験の要約

	Title	Creator	Subject	Description	Publisher	Contributor	Date	Type	Format	Identifier	Source	Language	Relation	Coverage	Right	SUM
冊子年							3									3
入力者						4										4
入力日							4									4
更新1							2									2
更新2							2									2
更新3							3									3
漢字																0
メモ																0
連番										5						5
旧連番										3						3
時代分類			4											3		7
時代分類番号			1													1
分野1			5					2								7
分野1番号			1													1
分野2			4					2								6
分野2番号			1													1
分野3			5													5
分野3よみ			3													3
分野4			4					2								6
分野4番号			1													1
分野5			5					1								6
分野5よみ			3					1								4
分野6			4					2								6
分野6番号			1													1
分野7	1		4													5
分野7よみ			3													3
頭書	3															3
主題	4															4
副題	3															3
題名	6															6
英文タイトル	4		1													5
執筆1		6														6
執筆2		5														5
英文執筆		5														5
種別								3						1		4
請求										3				1		4
誌著名	1				2						3		1			7
英文誌											3		1			4
巻号										2	2					4
通巻										2	2					4
開始頁										1	2					3
終了頁										1	2					3
総頁									2		2					4
和暦年							3							1		4
月							3							1		4
日							3							1		4
西暦年							3							2		5
翻複1			3										2			5
翻複2	1		1										2			4
作品名	1		4													5
作者名	1		4													5
概念	1		3													4
要約			2	2												4
キー			3													3
作成年度							3									3
作品名よみ			4													4
作者名よみ			4													4
SUM	26	16	78	2	2	4	29	13	2	17	16	0	6	10	0	221

ント (Title ~ Right) を表す。各セルの内容は論文目録データのレコード・エレメントが DC メタデータのエレメントにマッピングされた回数である。もし 6 人の被験者全員が、論文目録データのあるレコード・フィールドを DC メタデータの一つエレメントにマッピングした場合、それに該当するセルの内容は 6 となる。例えば論文目録データの「題名」フィールドは全員が DC メタデータの「title」エレメントにマッピングしているの、「題名」行、「title」列のセルの内容が 6 となっている。一方「時代分類」行を見ると、「subject」に 4 回、「coverage」に 3 回マッピングされている。これは被験者 A1、A2、A4 が subject に、A3 と A5 が coverage に、A6 が subject と coverage の両方にマッピングした結果を示している (表 2 参照)。

#### 4. 考察

マッピングで最初に問題となるのは、「何についての情報をメタデータとして記述するか」である。論文目録データの場合、(1)論文目録データを収録した冊子体の国文学年鑑、(2)レコードそのもの、(3)論文の書誌の 3 つの情報がメタデータの対象となりうる。表 1 において、「冊子年」は国文学年鑑に関連する情報である。「入力者」から「旧連番」および「作品年度」はレコードそのものに関連する情報である。それ以外が論文の内容と書誌に関する情報である。

ここで時間情報を取り上げて、この問題を考える。時間情報に関連する DC メタデータのエレメントは Date エレメントと Coverage エレメントの Temporal Qualifier である。Date にはリソースが作成あるいは有効になった時間が記述される。ここでリソースを何にするかが問題となるが、一般には電子化されたリソースのようであり、この場合はレコードの作成された時間となる。一方 Coverage にはコンテンツに関する時間 (Temporal Qualifier) や場所 (Spatial Qualifier) に関する情報を記述できる。そこで「和暦年」など論文の内容に関わる時間情報を Temporal Qualifier に、「入力日」などのレコードに関する時間情報を Date にマッピングすれば、両者の情報を区別できる。被験者 A2 と A6 は、このような考え方に基づいたマッピングを行ったと考えられる(「入力日」などを「Date」へ、「和暦年」～「西暦年」を「Coverage」にマッピングしている)。しかし国文学年鑑に関連する時間情報を区別することはできない。もし Qualifier が使えない場合、Date エレメントのみが利用可能となる (Coverage にはコンテンツに関する時間や場所に関する情報を記述できるものの、Qualifier を使わないと情報の種類を区別できないので、Coverage エレメントを使うのは好ましくない)。被験者 A3 と A6 はこの問題を意識していたと考えられる。ところで論文目録データには場所に関する情報がない、つまり Qualifier で時間と場所を区別する必要がないので、論文の内容に関する「和暦年」から「西暦年」までの時間情報を「Coverage」にマッピングすることができる。また検索を主目的とすれば、「入力日」から「更新 3」までを「Date」にマッピングする必要性は低い。その代わりに該当する情報が登録されている国文学年鑑に関する「冊子年」を「Date」にマッピングしたほうが、利用者にとっては有益であろう。

以下マッピングの結果に差の見られたレコード・フィールドについて眺めてみる。「時代分類」は論文が対象としている時代であり、「Subject」と「Coverage[Temporal]」に判断が分かれている。論文目録データの場合、「時代分類」は統制されているので「Subject」とした方が適切と考えられる。

「分野 1」から「分野 6」は「Subject」と「Type」に判断が分かれている。DC メタデータの Type エレメントはコンテンツのジャンルに関連した情報を記述する。「分野 1」は国語・和歌・漢文学など所謂ジャンル別の分類であり「Type」とするのが適切であるように思われる。一方「分野 3」は論文が対象としている作家名などであり、ジャンルよりも主題に近いと考えられる。これらについては再検討する必要がある。

「刊号」から「終了頁」についても「Source」と「Identifier」に判断が分かれている。DC メタデータの「Identifier」エレメントはリソースを一意に同定するための情報で、ISBN や URL などが該当する。ここでもリソースが何かにより状況が変わる。もしリソースをレコードそのものとするならば「連番」や「旧連番」が「Identifier」にマッピングされるべきで、その場合「刊号」から「終了頁」は「Source」にマッピングするのが適切であると考えられる。一方リソースをレコードの内容とすると、「刊号」から「終了頁」を「Identifier」にマッピングしてもおかしくない。しかし全ての被験者が「連番」を「Identifier」にマッピングしているので (表 3)、リソースを区別する意味でも、「刊号」から「終了頁」は「Source」へマッピングするのは適切であると考えられる。

「翻複 1」と「翻複 2」についても「Subject」と「Relation」に判断が分かれている。「翻複」は復刻・複製に関する情報であり、「Relation」へのマッピングの方が適切であると考えられる。ところで論文目録データには主題 (subject) が無い (レコード・フィールド「主題」は主たる題名である)。一方で、「時代分類」, 「分野 1」, 「分野 2」, 「分野 3」, 「分野 3 のみ」, 「分野 4」,

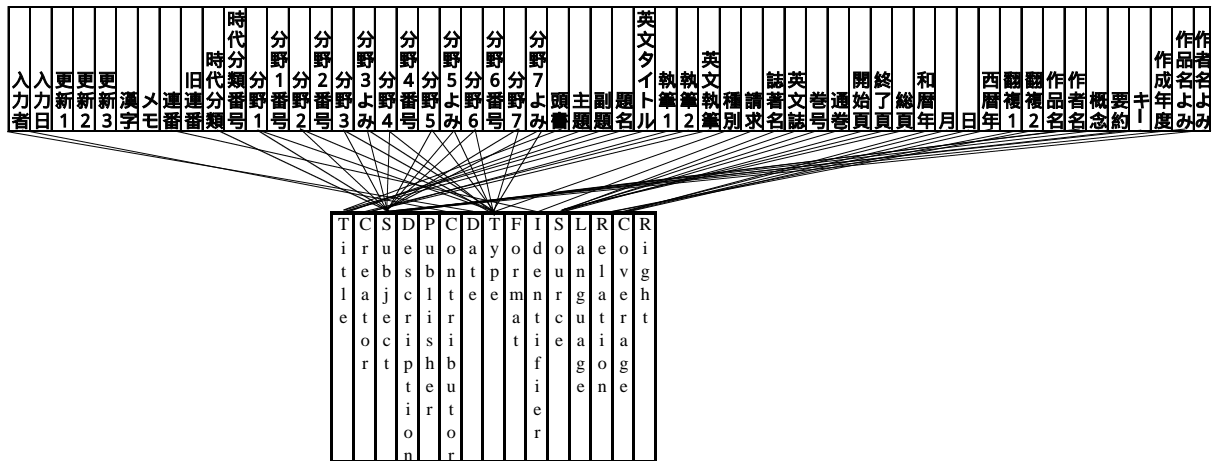


表4 . MARC21 へのマッピング実験

	M 1	M 2
冊子年		
入力者		
入力日	005	005
更新1	005	
更新2	005	
更新3	005	005
漢字		
ヌモ		
連番	001	001
旧連番		
時代分類	648#4\$a	648 #4\$a
時代分類番号		
分野1	655#4\$a	
分野1番号		
分野2	648#4\$a	
分野2番号		
分野3	600\$a	600\$a
分野3よみ	600\$a	
分野4		
分野4番号		
分野5		600\$a
分野5よみ		
分野6		
分野6番号		
分野7		600\$a
分野7よみ		
頭書		
主題	245\$a	245\$a
副題	245\$b	245\$b
題名	240 \$a or 130\$a	
英文タイトル	242\$a	242\$a
執筆1	100\$a	100\$a
執筆2	100\$a	100\$c
英文執筆	100\$a	100\$a or 242\$c
種別		
請求	084 or 024\$a	084\$a
誌者名	440\$a	440\$a
英文誌		440\$a
巻号	440\$v	440\$v
通巻	440\$v	440\$v
開始頁		
終了頁		
総頁	300\$a	300\$a
和暦年		
月		
日		
西暦年	260\$c	260\$c
翻複1	534\$e	533\$n
翻複2	534\$e	533\$n
作品名	630\$a	
作者名	600\$a	
概念	650\$a	650\$a
要約	520\$a	520\$a
主一	650\$a	650\$v
作成年度		
作品名よみ	630\$a	
作者名よみ	600\$a	

図1 . 論文目録データから DC メタデータへの直接マッピング

「分野5」、「分野5よみ」、「分野6」、「分野7」、「分野7よみ」、「題名」、「英文タイトル」、「作品名」、「作品名よみ」、「作者名」、「作者名よみ」、「概念」、「要約」、「キー」を組み合わせたものを「キーワード」と称して検索に利用している。これ「Subject」として利用することも考えられる。表2にも、その傾向が見られる。以上の議論をまとめると、論文目録データと DC メタデータ間の推奨される直接マッピングは図1のようになる。

これまでの議論は、論文目録データのレコード・フィールドを DC メタデータのエレメントへ直接マッピングする場合であった。一方、論文目録データベースと DC メタデータの間領域特異メタデータを介在させる方法が考えられる。領域特異メタデータとは、その領域で広く使われている、あるいは使うことを想定して規定された比較的情報量の多いメタデータである。論文目録データは目録データの種類であるので、MARC が領域特異メタデータの候補となる。論文目録と MARC は構造が似ているため、マッピング関係は1対1に近くなる、つまりマッピングの揺れが小さくなるものと期待される。一方、領域特異メタデータと DC メタデータ間のマッピングは専門家による提案がなされており、一般に crosswalk と呼ばれ公開されている。つまり MARC と crosswalk を組み合わせることにより、論文目録データから DC メタデータへの間接的なマッピングを行うことができる。

そこで MARC として MARC21<sup>[9]</sup>、また MARC21 から DC メタデータへのマッピングとして米国議会図書館の Crosswalk<sup>[9]</sup> を利用した間接マッピング実験を試みた。

表4にマッピング実験の結果を示す。表中の行方向は論文目録データのレコード・フィールドを、列方向は各被験者(M1とM2)を表している。各セルの内容はMARC21のフィールド名であり、被験者が論文目録データのレコード・フィールドをMARC21のフィールド名にマッピングした結果である。表2と比べると(1)論文目録データのレコード・フィールドからMARC21へフィールドのマッピングはほぼ1対1である、(2)マッピング結果の被験者によるバラツキが少ない、という結果が読み取れる。表4のマッピングにCrosswalkを適用して作成された、論文目録データとDCメタデータの間接マッピングは図2のようになる。

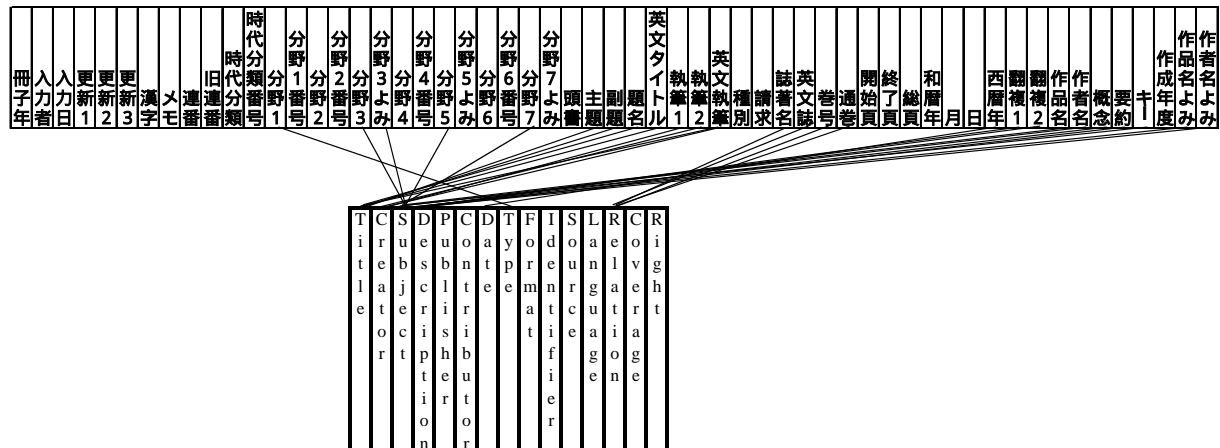


図2. MARC21 を経由した論文目録データから DC メタデータへの間接マッピング

図1と2を比較すると、間接マッピングではレコードの内容に関する情報のみがマッピングされ、レコードそのものや国文学年鑑などの情報はDCメタデータに反映されていない。また内容に関しても「分野」関係の情報は「Type」ではなく「Subject」へマッピングされている。「Source」へのマッピングが無くなっているが、これは「刊号」などが「Relation」へマッピングされた結果である。一方「Relation」へマッピングされるはずの「翻複」などは「Description」へマッピングされている。

## 5. まとめ

直接マッピングでは、被験者のリソースのとらえ方によりマッピングの結果に大きなブレが生ずる可能性のあること示された。間接的マッピングの場合、論文目録からMARC21へのマッピングでは期待通りにブレが少なかったが、最終的な結果は直接マッピングとかなり異なっていた。今後のより詳細な検討が必要である。

## 参考文献

- [1]原正一郎,安永尚志: 国文学電子資料館システム, 国文学研究資料館紀要, No.26, pp.25-52, 2000.
- [2]Dublin Core Metadata Initiative: The Dublin Core Element Set Version 1.1, 1999-07-02.
- [3]ANSI/NISO: ANSI/NISO Z39.50-1995 Information Retrieval (Z39.50) Application Service Definition and Protocol Specification, 1995.
- [4]原正一郎,柴山守,安永尚志:メタデータによるデータベースの機関間連携の実現 - 人文科学データ共有のための標準化 -, 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.
- [5]山本泰則,原正一郎,柴山守,安達文夫,合庭惇,安永尚志: 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(\*\*), pp.\*\*-\*\*, 2004.
- [6]論文目録データベース: 国文学研究資料館, [http://base1.nijl.ac.jp/~ronbun/cgi-bin/r\\_s\\_srch.cgi](http://base1.nijl.ac.jp/~ronbun/cgi-bin/r_s_srch.cgi).
- [7]Dublin Core Metadata Initiative: Dublin Core Metadata Element Set Version 1.1 Reference Description, <http://dublincore.org/documents/dces/>, 2004.
- [8]Dublin Core Metadata Initiative: Dublin Core Qualifiers, <http://dublincore.org/documents/2000/07/11/dcmes-qualifiers/>, 2000.
- [9]Library of Congress: MARC 21 Format for Bibliographic Data National Level Record---Bibliographic Full Level & Minimal Level, <http://www.loc.gov/marc/bibliographic/nlr/nlr.html#intro>, 2004.