

平成 27 年度 学術情報システム総合ワークショップ 調査報告書
外部データ活用のための参照 MARC 比較調査と NACSIS-CAT 新モデルの提案

【1 班】

安藤 大輝（国立国会図書館）

末田 真樹子（神戸大学）

萩原 泰子（九州大学）

原 智子（秋田大学）

目次

0. 要約	3
1. はじめに.....	4
1.1. 背景	4
1.2. 現状の課題	4
1.3. 本調査の概要.....	5
2. 新モデル案の概要.....	6
2.1 新モデル案についての背景と前提事項	6
2.1.1. 現行システムにおける書誌作成・所蔵登録	6
2.1.2. 書誌の名寄せ	7
2.1.3. VOL 崩し	7
2.1.4. 書誌作成と書誌利用の分離	7
2.2 新モデル案の構想とそのメリット	8
3. 調査について	10
3.1 調査の観点	10
3.2 調査対象.....	10
3.3 調査方法.....	11
3.4 TR の不一致項目の分類について	13
3.5 参照 MARC を流用入力する際に修正が必要となる割合.....	17
4. 調査結果	18
4.1 JPMARC の調査結果.....	18
4.1.1 JPMARC レコードと NC レコードの一致率.....	18
4.1.2 JPMARC レコードと NC レコードの不一致の分類.....	19
4.1.3 JPMARC レコードと NC レコードの要対応の不一致（ISBN 有の場合） ...	24
4.2 TRCMARC の調査結果.....	26

4.2.1	TRCMARC レコードと NC レコードの一致率	26
4.2.2	TRCMARC レコードと NC レコードの不一致の分類	27
4.2.3	TRCMARC レコードと NC レコードの要対応の不一致 (ISBN 有の場合)	32
4.3	USMARC の調査結果	34
4.3.1	USMARC レコードと NC レコードの一致率	34
4.3.2	USMARC レコードと NC レコードの不一致内容の分類	36
4.3.3	USMARC レコードと NC レコードの要対応の不一致 (ISBN 有の場合)	42
4.4	UKMARC の調査結果	43
4.4.1	UKMARC レコードと NC レコードの一致率	43
4.4.2	UKMARC レコードと NC レコードの不一致内容の分類	44
4.4.3	UKMARC レコードの要対応の不一致 (ISBN 有の場合)	49
4.5	DNMARC の調査結果	51
4.5.1	DNMARC レコードと NC レコードの一致率	51
4.5.2	DNMARC レコードと NC レコードの不一致内容の分類	52
4.5.3	DNMARC レコードの要対応の不一致 (ISBN 有の場合)	56
5.	調査を受けての新モデル案	58
5.1	外部データを活用した書誌作成のモデル	58
5.2	所蔵登録を行う対象	61
5.3	書誌と所蔵の分離	63
6.	まとめ	65
6.1	NACSIS-CAT レコードおよび参照 MARC の差異に関する調査	65
6.2	新モデル案の検討	67
6.2.1	外部データを活用した書誌作成モデル	67
6.2.2	所蔵登録について	67
	参考文献	69

別紙(a) MARC 別調査結果データ

別紙(b) 新モデルを実現するためのシステム要件など

0. 要約

平成 27 年度学術総合ワークショップのテーマは、NACSIS-CAT の持続可能なサービスの実現を目指し、「NACSIS-CAT の運用モデル再考」と設定された。これに関連して、本調査では、「外部データの活用を含めた所蔵登録の簡素化」という個別テーマのもと、参照 MARC の調査、および、NACSIS-CAT の新モデル案の検討を行った。

調査では、流用登録されたレコードを対象に、(1)参照 MARC (JP/TRC/US/UK/DN) と NACSIS-CAT レコードの一致率の算出、(2)不一致データのサンプル調査による不一致項目の分類を行った。2 つの調査により、現在の参照 MARC を利用する際に修正が必要な割合は、0.87%~6.02%であることがわかった。

この結果にもとづき、新モデル案を検討した結果、「修正レコードは新規データとして扱い、外部データと新規データを名寄せにより利用する書誌作成モデル」がもっとも実現可能性が高いという結論にいたった。

今回の調査では、外部データとして国内外の主要 MARC のみを調査したが、今後、他の MARC や MARC 以外の外部データについても同様に調査することにより、新モデル案のデータとして活用することも可能である。

1. はじめに

本稿では、目録業務の効率化の一つの方法として「外部データの活用を含めた所蔵登録の簡素化」について、新モデルの作成とその実現の可能性を検討するために行った調査を報告する。

1.1. 背景

NACSIS-CAT の運用開始から 30 年が経過しようとしている。この間、参加機関が拡大するとともにデータ登録件数も順調に増加し、ILL システムの開発や WebcatPlus のサービス開始など NACSIS-CAT は発展と進化を続けてきた。そのなかで、共同分担目録方式と参照 MARC の整備という運用方針は転換されることなく今日に至っている。

大学図書館を取り巻く環境は大きく変化するなか、目録業務については、担当人員の削減あるいは外部委託が進んでいる。

そのような状況下、米国議会図書館（以下 LC）の書誌コントロールの将来に関する米国議会図書館ワーキンググループ報告書『On the Record』^{1) 2)}（2008 年 1 月）において、また、国立情報学研究所（以下 NII）の次世代目録ワーキンググループ報告書『次世代目録所在情報サービスの在り方（最終報告）』³⁾（2009 年 3 月）においても、出版社等のメタデータを利用することによる目録作成の効率化が言及されている。

1.2. 現状の課題

現在は参照 MARC という形式で外部 MARC を活用し、目録業務における書誌新規作成の負担軽減が図られている。参照 MARC は NACSIS-CAT の仕様に合うようあらかじめ外部 MARC をフォーマット変換したものであるが、実際に流用して書誌を作成する際は、ほとんどの場合にレコード修正が必要となっている。また、NACSIS-CAT 独自の書誌階層構造のため、親書誌レコードの作成を伴う場合もあり、流用入力とは言え、書誌作成スキルが必要である。

NII が平成 23 年 3 月に実施したアンケートの報告書『NACSIS-CAT/ILL 参加館状況調査アンケート結果報告書』⁴⁾ の自由記述欄には、人手不足、書誌作成スキルの不足、そしてレコード調整の負担感を訴える意見を見ることができる。書誌作成スキル不足によりレコード調整が発生し、人手不足の現場でさらに労力を費やしてしまうという悪循環に陥っている。

参照 MARC の活用には外部 MARC のフォーマット変換というシステムティックな作業と、流用入力時のレコード修正作業という二重の作業が伴うことに、効率化の限界がある。課題の解決には、外部 MARC を含む外部データの更なる効果的な活用方法を検討し、所蔵登録を含めた目録業務全般の簡素化を目指す必要があると考えられる。

1.3. 本調査の概要

現状の課題解決のためには、外部データを無修正で利用できること、書誌と所蔵を分けて持つことの2点が必要となると考えた。これらにより、現在の目録業務で行っている、参照MARCを流用登録する際の修正作業とレコード調整作業、NC書誌のダウンロード・自館所蔵のアップロード作業を省略できる。この構想に基づき、2章ではNACSIS-CATの新モデルを提案する。3～4章では、外部データの無修正利用について検討するため、現在の修正状況を確認することとした。調査対象をNACSIS-CATにとって、もっとも身近な外部データである、参照MARCとし、どのような修正がどの程度行われているのかを明らかにした。5章では、この調査結果を受け、新モデルの修正を行っている。

2. 新モデル案の概要

1章で述べた通り、本報告書では NACSIS-CAT および参照 MARC の差異に関する調査を行い、調査結果に基づき、既存の NACSIS-CAT の枠組みを発展させ、外部データをより効率的に活用した書誌作成・所蔵登録管理の新しいモデル（以下、新モデル案）を提案する。

2.1 新モデル案についての背景と前提事項

まずは、現行システムにおける書誌作成・所蔵登録の概要について簡単に述べる。また、新モデル案では NACSIS-CAT についての各種の先行する調査や検討を前提としているため、前提事項としているもののうち、3つの重要なトピックについて簡単に説明する。

2.1.1. 現行システムにおける書誌作成・所蔵登録

NACSIS-CAT では各参加館が書誌データを分担して作成するオンライン共同分担目録を取っている。各参加館が作成した書誌は総合目録データベースに登録され、他館が作成した書誌を自館で用いることができる。また、書誌作成を効率化するため、参照ファイルとして各種 MARC を参照・利用できるようになっている。現行 NACSIS-CAT での書誌作成および所蔵登録の作業フロー概要図を図 2-1 に示す。

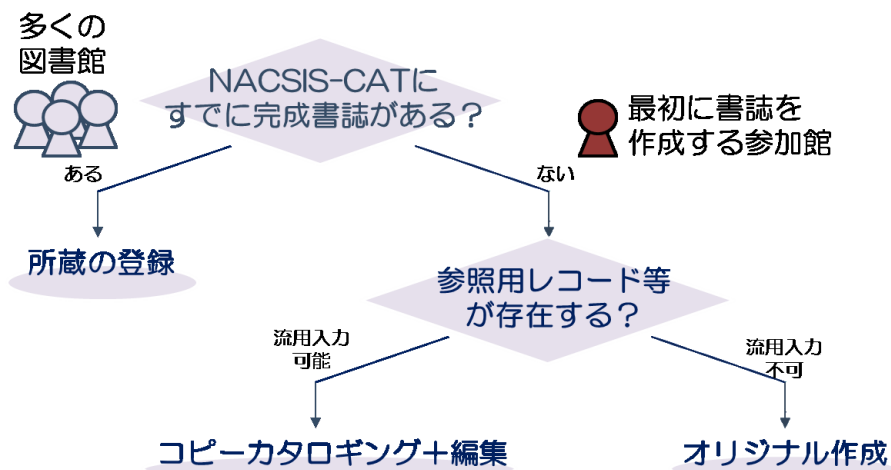


図 2-1 現行 NACSIS-CAT でのデータ登録フロー

各参加館は書誌データを登録する際、まず、登録したい書誌データが総合目録データベースに既に登録されていないかチェックする。総合目録データベースに既に登録されている場合は、自館での資料番号や排架位置などの所蔵登録処理のみを行う。総合目録データベースに登録されているデータが無い場合は続いて参照ファイルから書誌データを検索する。参照ファイルに対象となる書誌データがある場合には、参照ファイルから流用入力し、

参照ファイルにもデータが無い場合には新規でデータを作成することとなる。

ここで最初に書誌を作成（流用入力含む）した館（作成館）が、レコード調整の責任を持ち、書誌修正の責任を持つこととなっている。例えば責任表示を変更するなど、修正の内容・程度が一定以上であるものは、発見館が作成館に連絡し、協議のうえ原則として作成館が修正することとなっている。このような作成館責任の方式は、データベースの品質向上に寄与してきた一方で、レコード調整作業のコストや負担感の問題がこれまでも指摘されている³⁾。

2.1.2. 書誌の名寄せ

複数の情報源から外部のデータを効率的に利用するためには、書誌の名寄せが必須となる。NACSIS-CATと同じく国立情報学研究所が開発・運用している学術情報システムである CiNii Articles では、機械学習アルゴリズムのひとつである SVM によって書誌の同異を判別し、その判別がどれくらい明らかであるかによって確信度というパラメータを算出している。確信度が高いものについては、SVM での判別結果をそのままシステムに反映し、確信度の低いものについては人手での判断に回すことで効率的で信頼度の高い名寄せシステムを構築している⁴⁾。本報告書で提案する新モデル案では、CiNii Articles と同様の名寄せシステムを用いることを想定としている。CiNii Articles では高い精度での名寄せシステムが運用できているため、新モデル案においても機械学習と人手での判断を組み合わせることで十分な名寄せ精度が得られることを前提とする。

2.1.3. VOL 崩し

現在の NACSIS-CAT 書誌データの大きな特徴として、一つの書誌レコードが物理単位ではなく書誌単位となっている点があげられる。他の MARC などのほとんどは物理単位のレコードに対して対応する書誌情報を記述しているが、NACSIS-CAT では物理単位の情報は書誌中の項目(VOL)に含めて書誌単位で 1 レコードとしている。新モデル案では、外部データとの相互運用性を高めるために、現在の NACSIS-CAT での書誌階層構造を廃止して、既存の NACSIS-CAT 登録レコードも物理単位に変換することを前提としている。

2.1.4. 書誌作成と書誌利用の分離

これまでも NACSIS-CAT の軽量化・合理化を実現するために、書誌作成と書誌利用を分離するためのモデルが検討されてきた。「これからの学術情報システム構築検討委員会」第 10 回委員会での検討結果として、図 2-2 のようなモデルが提示されている。このモデルにおいては先に述べた 2 項(2.1.2. 書誌の名寄せ, 2.1.3. VOL 崩し)を前提として検討されている。

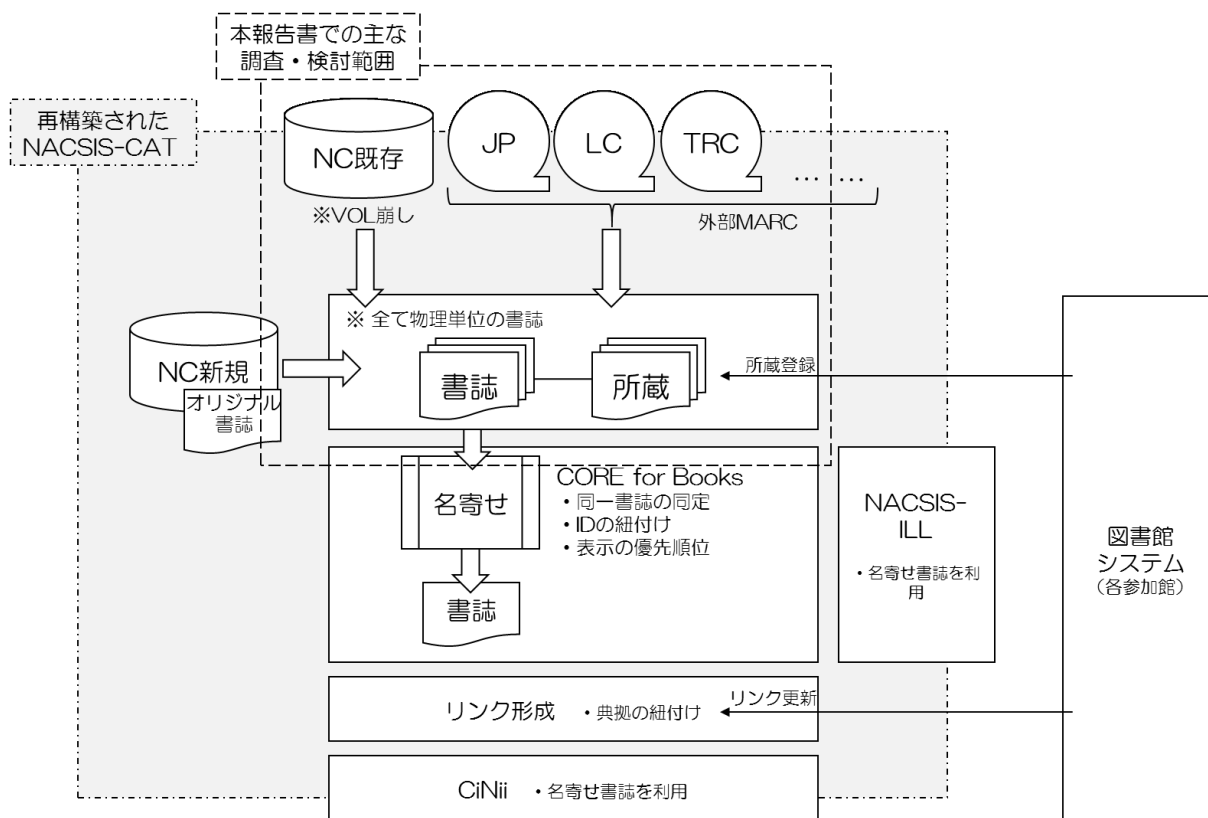


図 2-2 これまで検討されてきた書誌作成と書誌利用の分離モデル⁵⁾

2.2 新モデル案の構想とそのメリット

新モデル案では、これまでの書誌作成作業に掛かっていた負担の簡素化を図るために、外部 MARC データをなるべく人手による編集は行わずにシステムに取り込むことを想定している。新モデル案は「2.1.4. 書誌作成と書誌利用の分離」で検討されているモデルをベースとし、主に図 2-2 の長破線で示した部分(外部データからの書誌作成, 所蔵登録の方法)の実現方法を調査・検討したものになる。

書誌作成については、「2.1.4. 書誌作成と書誌利用の分離」と同様に外部 MARC を活用し、外部 MARC にデータがある場合には新規データを作成せず、当該外部データを修正せずにそのまま活用できるか調査する。外部データを修正せずに利用することができれば、書誌作成に係るコスト(レコード調整に係るコスト含め)を大幅な低減が期待できる。新モデル案で外部データ活用が実現した場合においても、特殊資料などについては外部のデータソースに書誌データがなく、引き続き参加館が作らなくてはならないものが存在する。これらの作成コストやスキル維持などの問題解消するための案としては、本調査と並行する調査において提案される。そのため本調査では対象外とし、詳細は扱わない。

また、本報告書で提案する新モデル案では、書誌データのみについては新システム側に一か所集中で保持し、排架位置などの所蔵情報は各館のローカルシステムにのみで保持す

ることの実現可能性を検討する。現行 NACSIS-CAT では排架位置などの所蔵情報も全て NACSIS-CAT に登録しているため、各館で蔵書整理などを行った時にローカルシステムと NACSIS-CAT との所蔵情報同期のために各館作業が発生していた。書誌情報は新システム側に、所蔵情報はローカル側に分離して持つことで、データ同期作業に係る労力を低減できるか検討を行う。

新モデル案の構想については、対象を図書に限定するものではないが、今回モデルを検討するにあたって図書を対象に調査を行った。

3. 調査について

3.1 調査の観点

ここでは、4章で報告する各 MARC の調査結果について、調査対象および調査方法、不一致項目の分類、要修正の計算方法について述べる。新モデル案において、外部データに必要とされる要件（外部データ側）と運用ルールの変更内容（NACSIS-CAT 側）を明らかにすることを目標とする。

2.3 で述べたとおり、NACSIS-CAT レコードおよび参照 MARC の比較調査を行い、以下の観点から分類する。

1. システム処理で対応できる部分
2. ルールの違いによる部分
3. どこかで人が手を加えなくてはならない部分(要対応)

上記の 3 に相当する割合、すなわち「外部データに修正を加えたとしたら、何%なのか」を算出した。ここでの「修正」とは、同一資料であるにもかかわらず、**無視できない=要対応の不一致**がどれくらいあるかという意味である（e.g. データの誤り、タイトルの打ち間違い）。指している資料が違う、つまり「別資料であること」は名寄せで解決すべき問題とし、3には含まない。

3.2 調査対象

NACSIS-CAT にとって、もっとも身近な外部データである、参照 MARC を調査対象とした。NACSIS-CAT にロードされた参照 MARC と、流用登録された NACSIS-CAT レコードを比較し、不一致が生じている割合と、不一致の分類を行った。

今回の調査では、すべての参照 MARC を対象とすることが困難なため、収納件数の多い¹、以下の 5 つの参照 MARC を調査対象とする

JPMARC, TRCMARC, USMARC, UKMARC, DNARC

(1) 抽出条件

CRTDT : 20100101~20141231

データ形式 : pai 形式

(2) 出力項目

¹ NACSIS-CAT 統計情報 (2015 年 12 月 20 日現在)

<http://www.nii.ac.jp/CAT-ILL/archive/stats/cat/db.htm> (参照 2015-12-25)

調査にあたって、java プログラムを作成し、pai データを 1 レコード 1 行の tsv データに変換した。

参照 MARC :

ID, MARCID, SOUCE, ISBN, ISBN_A, CRTDT, RNWDT, TR, TRR, PLC, PBLC, PBLCDT, PHYNM, PHYSZ, NT, AL

NC 書誌 :

ID, CRTDT, RNWDT, MARCID, SOUCE, UID, VLSR, ISBN, YR1, CTR, CTGL, TXTL, TR, TRR, PLC, PBLC, PBLCDT, PHYNM, PHYSZ, NT

(3) 繰り返しのある項目について

PLC, PBLC, PBLCDT については、先頭から順に「PLC1, PLC2, PLC3, PLC4」という項目に出力した。また、NT, AL, ISBN, VLSR については、NT1~NT10, AL1~AL10 のように区切って同様に処理する。11 個以降があれば 10 番目に連結した。

存在しない項目と値が無い項目について、それぞれ、存在しない項目=空欄、値がない項目=NULL とした。

3.3 調査方法

3.2 で取得した NC レコードについて、さらに ISBN 有無によって二分した。Microsoft Access を使い、tsv データに対して、以下のクエリを実行した。

(SOUCE=<>"ORG"かつ ISBN1="" Or Is Null)

または (SOUCE=<>"ORG"かつ ISBN1=Like "**NULL")

Microsoft Excel のフィルタ機能を使用し、ISBN_2 から ISBN_10 まで、「(空白セル)」または「_NULL」のフィルタをかけた。

表 3-1 ISBN 有無別 NC レコード件数 (割合)

ISBN 有	ISBN 無
NC レコード総数 - ISBN 無	ISBN_1~10 のいずれも空白、または Null
535,335 件 (68%)	251,504 件 (32%)

(1) 参照 MARC ごとの担当分け

TRC と JP を原、US を末田、UK と DN を萩原が担当した。

(2) 一致率の算出方法

NC レコードの MARCID をキーにして、NC レコードと流用元の参照レコードとつき合わせ、TR, TRR (JP と TRC のみ), PBLC_1, PBLCDT_1~3, PHYNM, PHYSZ の一致率を参照 MARC ごとに算出した。各項目に対して、事前処理は行わず Microsoft Excel の IF 関数を用いて、完全一致か否かで一致率を算出した。

(3) 不一致項目の分類のためのサンプル抽出

TR, TRR (JP と TRC のみ), PBLC_1, PBLCDT_1, PHYNM, PHYSZ において、不一致となったレコードに対して、それぞれランダムに 500 件ずつ抽出し、NC レコードと参照レコードを目視で比較しながら、不一致を分類していった。500 件の抽出方法は以下のとおりである。

ISBN 有 : YEAR=2014, NCID=末尾が 0 (500 に満たない場合は、末尾 1,2,3... と追加)

ISBN 無 : NCID=末尾が 0 (500 に満たない場合は、末尾 1,2,3... と追加)

※ISBN 有のレコード件数が ISBN 無に比べて多いため、出版年を 2014 年に絞った。

3.4 TR の不一致項目の分類について

不一致項目の分類について説明する。各項目 (TR, TRR (JP と TRC のみ), PBLC, PBLCDT, PHYNM, PHYSZ) について、以下のように分類した。

(1) TR の不一致項目の分類

表 3-2 TR の不一致項目の分類

不一致項目の種類	対応の要不要	不一致項目の種類	対応の要不要
書名の相違	要対応	副書名の有無	
書名のヨミの相違	要対応	ルビの有無	機械処理可
責任表示の相違	要対応	PTBL による相違	
責任表示の有無		VOL 積みによる相違	機械処理可
責任表示の省略形の相違 (イニシャルなど)		スペースの有無	機械処理可
責任表示の役割表示の相違および有無		記号の相違	機械処理可
書名の順番の相違		大文字小文字などの字体の相違	
並列書名の相違	要対応	別資料であるもの	
並列書名の有無		翻字	機械処理可
副書名の相違	要対応	その他	?

書名の相違 (例)

	NC レコード	参照レコード
DN	Schriftenreihe des Lehrstuhls Baubetrieb und Bauprozessmanagement der Technischen Universität Dortmund	Schriftenreihe des Lehrstuhls Baubetrieb und Bauprozessmanagement der Universität Dortmund

責任表示の相違 (例)

	NC レコード	参照レコード
JP	越谷市 健康 福祉部高齢介護課編	越谷市福祉部高齢介護課編

綴り、言語の違い、情報源の相違の他、複数著者の場合は省略される場合もある。

責任表示の役割表示の相違および有無（例）

	NC レコード	参照レコード
UK	Barbara Vine	by Barbara Vine
US	Paula Yurkanis Bruice	Paula Yurkanis Bruice, University of California, Santa Barbara

翻訳・編集などの役割の他，所属や身分を示す語の相違・有無も含む。

副書名の相違（例）

	NC レコード	参照レコード
TRC	150 人の Shine : CHORI NEXT 創業 150 年記念特別号 / 蝶理株式会社編	150 人の Shine : 蝶理株式会社創業 150 年記念特別号 : 社内報 CHORI NEXT 特別号 / 蝶理株式会社編集

検索の際に，副書名だけを検索キーにすることは考えにくいため，相違についてのみ，検索からもれることを想定して「要対応」とした。

(2) TRR の不一致項目の分類

表 3-3 TRR の不一致項目の分類

不一致項目の種類	対応の要不要	不一致項目の種類	対応の要不要
書名(TR)の相違		スペースの有無	
ヨミの相違	要対応	記号の相違	
ヨミの有無	要対応	外国語，記号，数字等の読み方の相違（有無含む）	
並列書名のヨミの有無	要対応	大文字，小文字等の字体の相違	
副書名のヨミの有無	要対応	その他	
ワカチの相違			

ヨミの相違（例）

	NC レコード	参照レコード
JP	ムサシノ シリツ トショカン キュウゾ ウ(ショカ)モンジョ	ムサシノ シリツ トショカン キュウゾ ウ(ショカ)ブンジョ

外国語、記号、数字等の読み方の相違（有無含む）（例）

	NC レコード	参照レコード
TRC	オカシナ コヤ：イゴコチ ノ イイ ミ ニ ハウス センボウ ノ 35 ケン	オカシナ コヤ：イゴコチ ノ イイ ミ ニ ハウス センボウ ノ サンジュウゴ ケン

(3) PBLC の不一致項目の分類

表 3-4 PBLC の不一致項目の分類

不一致項目の種類	対応の要不要	不一致項目の種類	対応の要不要
出版者名の相違	要対応	記号の相違有無	
出版者名の省略形の相違		固有名詞以外の相違	
出版者名の有無	要対応	翻字	
スペースの有無		別資料	
大文字小文字などの字体の相違		その他	?
冠詞の有無			

出版社名の相違（例）

	NC レコード	参照レコード
DN	J. Cramer	Cramer in der Gebr.-Borntraeger-Verl.-Buchh

固有名詞以外の相違（例）

	NC レコード	参照レコード
UK	Emerald	Emerald Publishing

社名以外の省略形の違いや有無をここに分類した。

(4) PBLCDT の不一致項目の分類

表 3-5 PBLCDT の不一致項目の分類

不一致項目の種類	対応の要不要	不一致項目の種類	対応の要不要
出版年の相違	要対応	記号の相違	表現の相違
出版年の有無	要対応	表記の相違	機械処理可
		c マークの有無	機械処理可

出版年の相違（例）

	NC レコード	参照レコード
TRC	[2009.12 公表にあたって]	2010.2
DN	2014, c2013	2013

記号の相違（例）

	NC レコード	参照レコード
DN	[2010]	2010
JP	2013.1-	2013.1

カッコ、ハイフンなどの相違を分類する。

表記の相違（例）

	NC レコード	参照レコード
US	1770	MDCCLXX [1770]

出版年を表す表記の相違を分類した。

(5) PHYNM の不一致項目の分類

表 3-6 PHYNM の不一致項目の分類

不一致項目の種類	対応の要不要	不一致項目の種類	対応の要不要
数量の相違	要対応	単位の相違	
表記の有無	要対応	スペースの有無	
表記の相違		記号の相違	
前付, 後付ページの有無		その他	?

表記の相違（例）

	NC レコード	参照レコード
US	390 p.	390 pages
UK	240 p	240 S

ページ数などを表す表記の相違を分類した。

単位の相違（例）

	NC レコード	参照レコード
UK	viii, 120 p.	1 volume

VOL 積みなど書誌単位の相違が影響しているものを含む。

記号の相違（例）

	NC レコード	参照レコード
JP	[40]p	[40p]

かっこの順序，ピリオドの相違・有無などが含まれる。

(6) PHYSZ の不一致項目の分類

表 3-7 PHYSZ の不一致項目の分類

不一致項目の種類	対応の要不要	不一致項目の種類	対応の要不要
大きさの相違	要対応	スペースの有無	機械処理可
表記の有無		記号の相違	機械処理可

3.5 参照 MARC を流用入力する際に修正が必要となる割合

参照 MARC を流用して，NC レコード（ISBN 有）を作成するときに，修正が必要な割合を求める。書誌検索および同定の際に重要となる TR に焦点をしばり，TR が誤りであり，かつその誤りが要修正である割合を以下の計算式により算出した。

$$\begin{aligned}
 & \text{(TR が誤っているときに修正しなければならない割合)} \times \text{(TR が誤っている割合)} \\
 & = (\text{要対応の不一致} \div \text{誤りレコード数}) \times \{ \text{TR が不一致} \div (\text{TR が一致} + \text{TR が不一致}) \} \\
 & = \{ (\text{要対応の不一致} \div (500 - \text{別資料} - \text{その他})) \} \times \{ \text{TR が不一致} \div (\text{TR が一致} + \text{TR が不一致}) \}
 \end{aligned}$$

4.調査結果

4.1 JPMARC の調査結果

4.1.1 JPMARC レコードと NC レコードの一致率

(1) JPMARC 流用作成の NC レコードの抽出

3.3 のとおり，JPMARC を流用して作成された NC レコードを抽出し，ISBN の有無によって分類した。結果は表 4-1 のとおりである。

表 4-1 JPMARC 流用の NC レコード件数（割合）

NC レコード件数	ISBN 有	ISBN 無
64,836	52,973 (81.7%)	11,863 (18.3%)

(2) JPMARC レコードと NC レコードの各項目の一致率

3.3.(2)のとおり，ISBN の有無別に各項目の一致・不一致件数および一致率を調査した。結果は表 4-2，図 4-1 のとおりである。

表 4-2 JPMARC と NC レコードの一致・不一致件数および一致率

項目	TR		TRR		PBLC_1		PBLCDT_1	
	有	無	有	無	有	無	有	無
一致件数	35,724	6,054	30,211	6,058	52,062	10,473	47,552	9,698
不一致件数	17,249	5,809	22,762	5,805	911	1,390	5,421	2,165
一致率	67.4%	51.0%	57.0%	51.1%	98.3%	88.3%	89.8%	81.7%

項目	PHYNM		PHYSZ		ISBN	
	有	無	有	無	有	無
一致件数	43,458	7,843	51,858	10,150	52,402	11,608
不一致件数	9,515	4,020	1,115	1,713	571	255
一致率	82.0%	66.1%	97.9%	85.6%	98.9%	97.9%

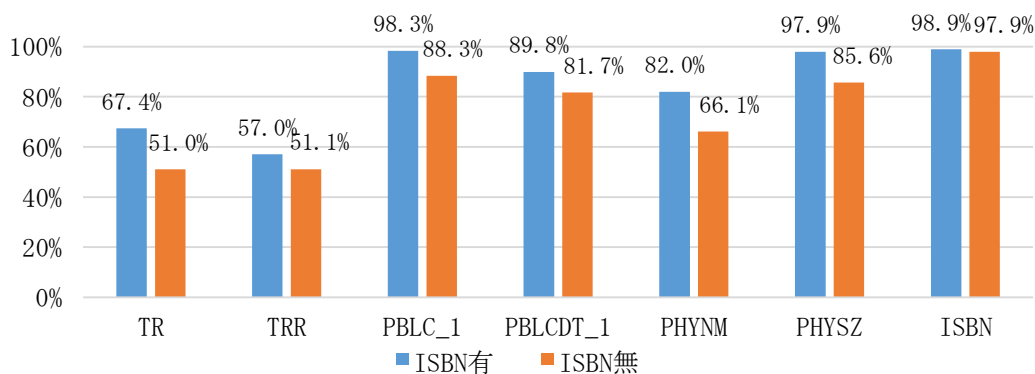


図 4-1 JPMARC と NC レコードの一致率

ISBN 有と ISBN 無では，全般に ISBN 有の方が一致率が高いが，共に TR と TRR の一

致率が低いなど各項目の一致率の傾向は同じである。ただし、PBLCDT, PHYNM, PHYSZ には、空値のための一致も含まれている。

4.1.2 JPMARC レコードと NC レコードの不一致の分類

(1) TR の不一致の割合

3.4.(1)のとおり、項目ごとにランダムに抽出した不一致レコード 500 件について、目視により不一致による分類を行い（付録 1）、その結果から不一致の割合を導き出した。TR の不一致の割合は図 4-2 のとおりである。

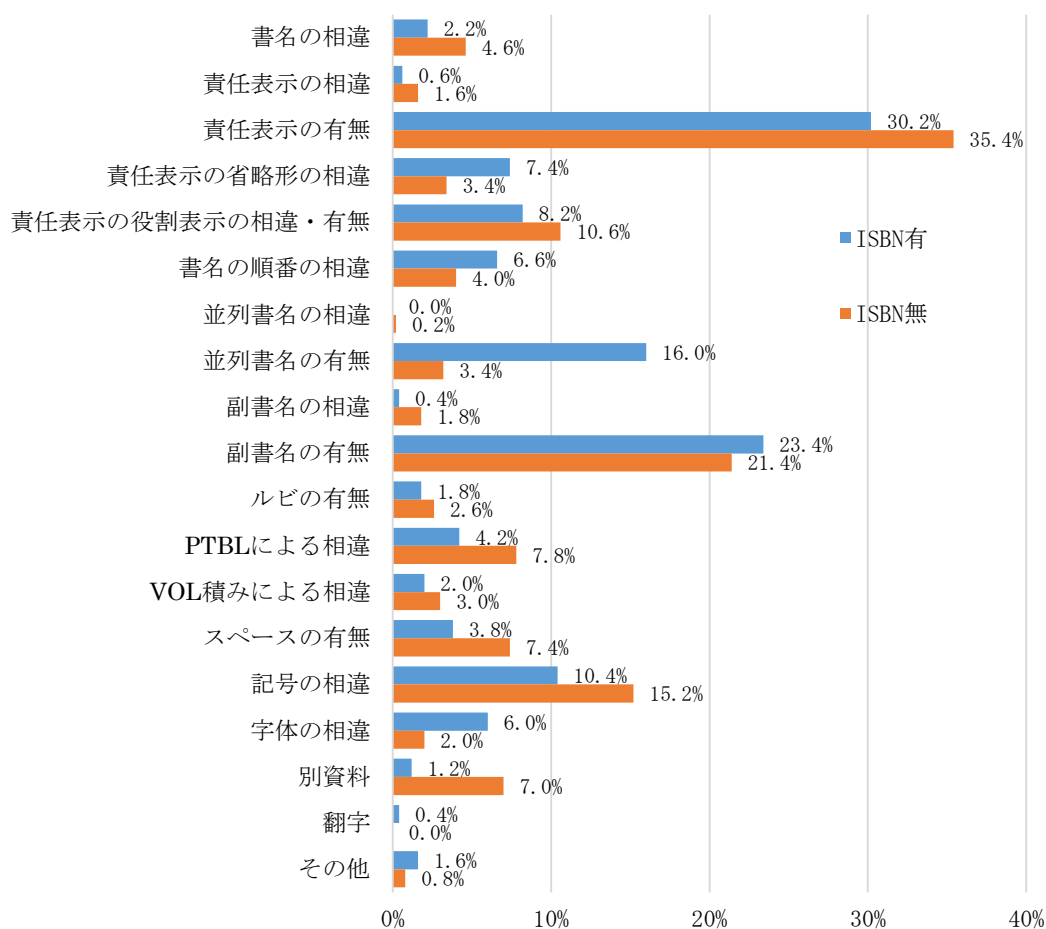


図 4-2 TR の不一致の割合

TR で不一致の割合が最も大きい内容は、ISBN の有無にかかわらず、責任表示の有無によるもので 30%を超えている。その理由の一つは、責任表示とする範囲がルールによって異なることにある。例えば、監修者について JPMARC の元となる JAPAN/MARC のルールでは責任表示の対象としているのに対し、NC のルールでは対象としていない。また、JPMARC レコードで PTBL にデータが存在している場合に責任表示が TR ではなく、PTBL にマッピングされている場合も見られる。

次に不一致の割合が大きい内容は、副書名の有無である。JPMARC レコードに副書名がある場合が多く、並列書名の有無でも同じ傾向である。JPMARC レコードにある副書名や並列書名が、NC レコードでは VT 欄に標題紙以外の書名として入力されている場合が多く見られる。また、ISBN の有無による相違が最も大きいのが並列書名の有無で、ISBN 有が 16.0%、ISBN 無が 3.4% である。その差は、今回抽出した ISBN 無のレコードが元々並列書名を持たないものが多かったと推察できる。

記号の相違で最も多かったのは、区切り記号の相違で、NC レコードでは“：(コロン)”で区切って副書名としていても、JPMARC では区切らずに本タイトルとしている場合が多かった。また、カッコの用法で、NC レコードでは () や 「 」 のものが、JPMARC では 〈 〉 になっている場合が見られた。

責任表示の役割表示の相違および有無で最も多く見られたのが、“編”と“編集”，“著”と“[著]”の相違である。

(2) TRR の不一致の割合

3.4.(2) のとおり、項目ごとにランダムに抽出した不一致レコード 500 件について、目視により不一致による分類を行い(付録 1)、その結果から不一致の割合を導き出した。TRR の不一致の割合は図 4-3 のとおりである。

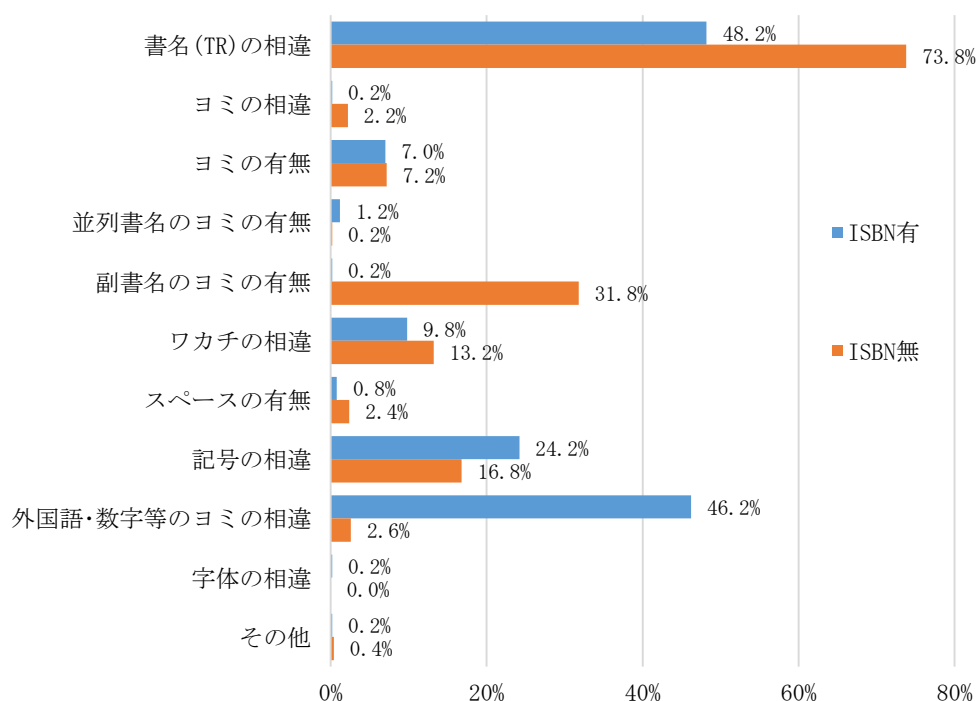


図 4-3 TRR の不一致の割合

TRR の不一致の割合が最も大きい内容は TR 中の書名自体の相違によるもので、これは

本来のヨミの相違とは異なるため、不一致データの対象外とする。

次に不一致で割合が大きいのは、外国語や数字等のヨミの相違で、ISBN 有の場合では46.2%を占める。その理由は、JAPAN/MARC のルールではラテン文字やアラビア数字にカナでヨミを付与することになっているのに対し、NC のルールでは表示されているとおりをヨミとする（発音には従わない）ためである。

副書名のヨミの有無が ISBN 無で31.8%である。多くの場合が JPMARC レコードにヨミが無い。ただし、NDL-OPAC で該当資料の書誌を確認したところ、副書名のヨミがデータとして入っているため、マッピングによるもの、またはデータ抽出時のエラーと考えられる。

(3) PBLC_1 の不一致の割合

3.4.(3)のとおり、項目ごとにランダムに抽出した不一致レコード 500 件について、目視により不一致による分類を行い（付録 1）、その結果から不一致の割合を導き出した。PBLC_1 の不一致の割合は図 4-4 のとおりである。

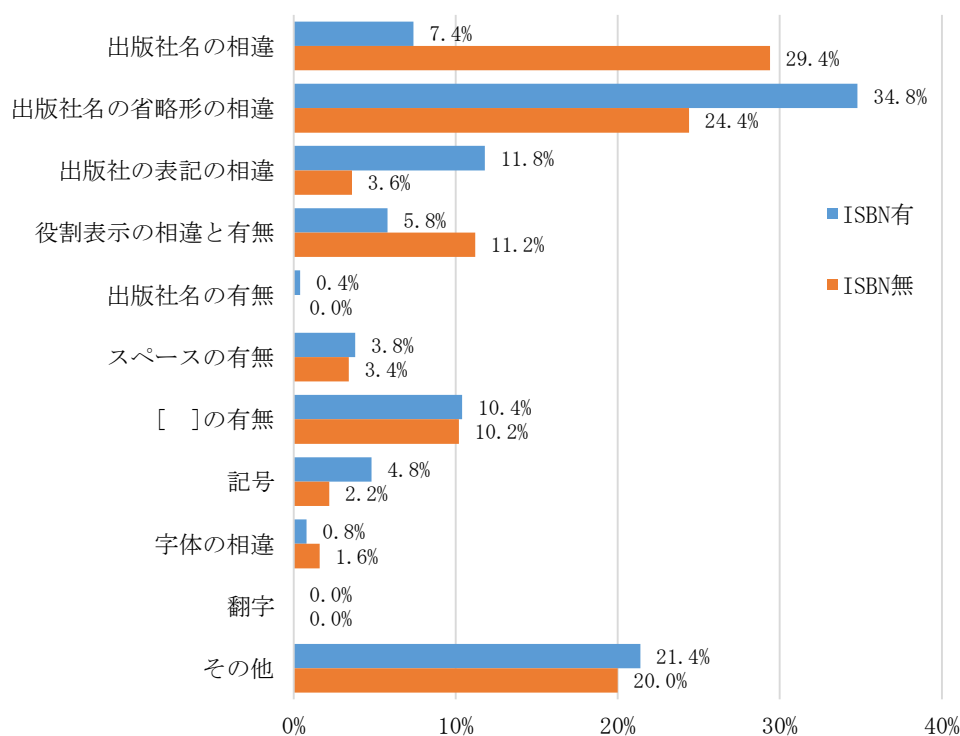


図 4-4 PBLC_1 の不一致の割合

PBLC_1 で不一致の割合が最も大きい内容は、出版社名の省略形の相違によるもので、ISBN 有で34.8%、ISBN 無で24.4%である。出版社名の省略形の相違で多く見られるパターンは、部課名等の有無と“株式会社”の省略がある。

出版社名の相違が ISBN 無で29.4%を占めているが、データを確認すると、類似資料の

JPMARC レコードを流用して作成した別資料の NC レコードである場合が多かった。

その他に分類した不一致の内容の多くは、出版社の数に相違があるものである。

(4) PBLCDT_1 の不一致の割合

3.4.(4)のとおり、項目ごとにランダムに抽出した不一致レコード 500 件について、目視により不一致による分類を行い（付録 1）、その結果から不一致の割合を導き出した。

PBLCDT_1 の不一致の割合は図 4-5 のとおりである。

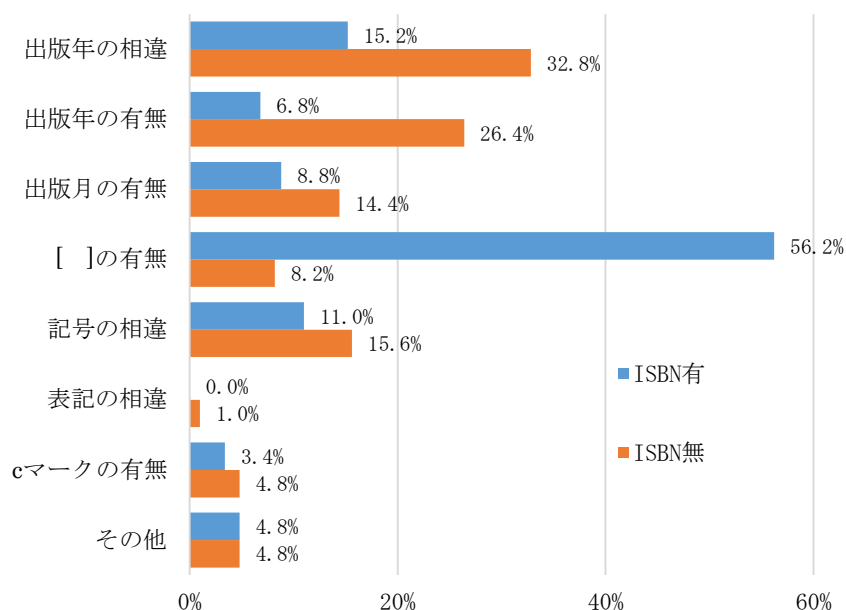


図 4-5 PBLCDT_1 の不一致の割合

PBLCDT_1 で最も不一致の割合が大きい内容は、[]の有無によるもので ISBN 有では 56.2%を占める。ほとんどの場合が NC レコードに[]がある。その理由として考えられるのが、出版年の情報源のルール相違である。JAPAN/MARC のルールでは必要に応じてカバーや箱が情報源となりうるとしているが、NC レコードは情報源の対象としていないため、カバーや箱に記載されている情報を[]を付けて補記しているものと推察できる。

次に不一致で割合が大きいのが、出版年の相違で ISBN 無で 32.8%である。ただし、データを確認すると、類似資料の JPMARC レコードを流用して作成した別資料の NC レコードである場合が多かった。

出版年の有無が ISBN 無で 26.4%である。NC レコードにデータが無い場合が 57 件、JPMARC レコードにデータが無い場合が 75 件である。NC レコードにデータが無い場合は、NC レコード側の PUB データが 2 行にわたっているため、PBLCDT_1 ではなく、PBLCDT_2 に出版年がある場合が多い。JPMARC レコード側にデータが無い場合は親書誌であることに起因しており、親書誌レコードを除くと 75 件→6 件に減少する

(5) PHYNM の不一致の割合

3.4.(5)のとおり、項目ごとにランダムに抽出した不一致レコード 500 件について、目視により不一致による分類を行い（付録 1）、その結果から不一致の割合を導き出した。PHYNM の不一致の割合は図 4-6 のとおりである。

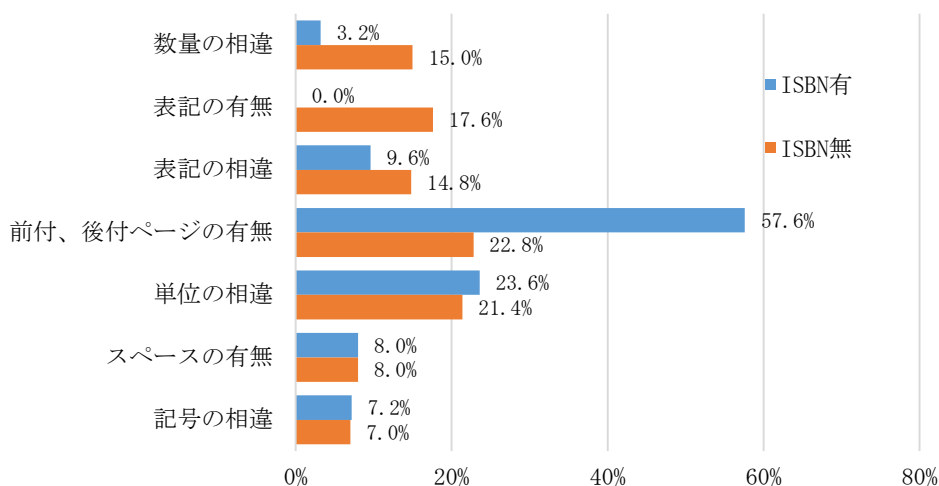


図 4-6 PHYNM の不一致の割合

PHYNM で不一致の割合が最も大きい内容は、前付・後付ページの有無で ISBN 有で 57.6%を占める。NC レコードに前付・後付ページがあり、JPMARC レコードに無い。これは JAPAN/MARC のルールでは、区別のために用いられているローマ数字は記録しない（NCR2.5.1.2 別法 2 非適用のため）ことによるものと推察できる。

単位の相違は ISBN の有無にかかわらず 20%を超えているが、NC レコードが VOL 積み書誌であることに起因している場合がある。また、ページ付けがないとき NC レコードでは“1冊”としているのに対し、JPMARC レコードでは[]を付けてページ数を補記している場合が見られる。

表記の有無が ISBN 無で 17.6%あるが、88 件中 87 件が JPMARC レコード側にデータが無く、親書誌であることに起因している。

(6) PHYSZ の不一致の割合

3.4.(6)のとおり、項目ごとにランダムに抽出した不一致レコード 500 件について、目視により不一致による分類を行い（付録 1）、その結果から不一致の割合を導き出した。PHYSZ の不一致の割合は図 4-7 のとおりである。

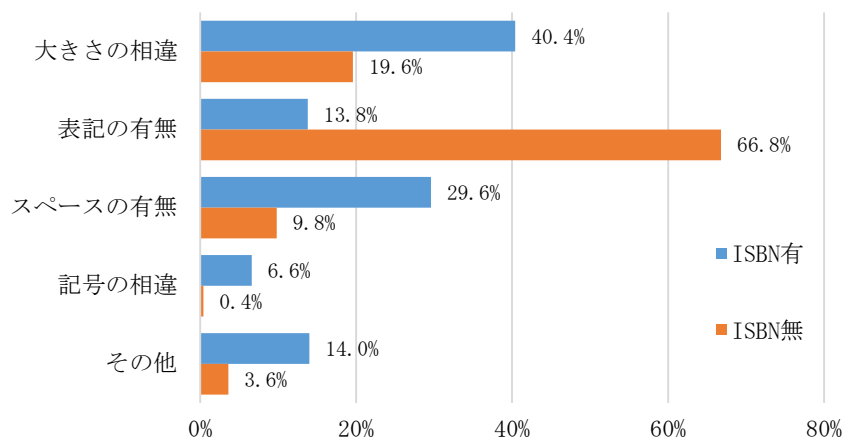


図 4-7 PHYSZ の不一致の割合

PHYSZ で不一致の割合が最も大きい内容は表記の有無である。特に ISBN 無では 66.8% を占めている。親書誌に起因している場合と、JAPAN/MARC では視聴覚資料の大きさを入力していないことによる。

大きさの相違が ISBN 有では 40.4% を占めるが、1cm 以下の相違の場合が多い。

4.1.3 JPMARC レコードと NC レコードの要対応の不一致 (ISBN 有の場合)

(1) JPMARC レコードと NC レコードの要対応の不一致の割合 (ISBN 有の場合)

4.1.2 で分類した結果のうち、ISBN 有の場合について ISBN が一致するレコードを対象に、3.4 で要対応と判断した不一致の件数を調査した。結果は表 4-3 と図 4-8 のとおりである。

表 4-3 JPMARC (ISBN 有) の要対応の不一致の件数 (割合)

項目	不一致	件数	項目	不一致	件数
TR	書名の相違	11	PBLCDT_1	出版年の相違	38
	責任表示の相違	3		出版年の有無	25
	並列書名の相違	0		計 (割合)	63(14.1%)
	副書名の相違	2	PHYNM	数量の相違	5
	本タイトル語順の相違 注)	19		表記の有無	0
	計 (割合)	35(7.2%)		計 (割合)	5(1.1%)
TRR	ヨミの相違	1	PHYSZ	大きさの相違	168
	計 (割合)	1(0.2%)		計 (割合)	168(39.1%)
PBLC_1	出版社名の相違	33			
	出版社名の有無	2			
	計 (割合)	35(8.3%)			

注) 3.4.(1)の TR の不一致項目の分類では、書名の順番の相違は無視できるとしたが、和書の場合

合は検索結果に影響するため、本タイトルの語順の相違は要対応に追加した。

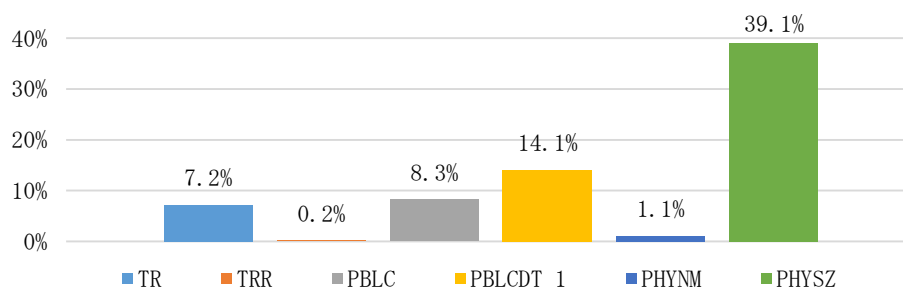


図 4-8 JPMARC (ISBN 有) の各項目の要対応の割合

要対応の割合が最も大きい項目が、PHYSZ で 39.1% である。その不一致は全て大きさの相違である。ただし、データを精査すると 1cm 以下の相違が多く、1cm 以下の相違を除くと、相違件数は 168 件 (39.1%) → 8 件 (1.8%) に激減する。採寸の誤差、もしくは刷による若干のサイズの差の場合もあるものと推察できる。

次に要対応の割合が大きい項目は PBLCDT_1 で、14.1% である。不一致としては、出版年の相違と出版年の有無の 2 つの場合がある。後者については、出版社数の相違により、PBLCDT_1 と PBLCDT_2 で出版年が一致する場合があります、これを一致とみなした場合は、出版年の有無が 25 件 → 4 件となり、PBLCDT の要対応の件数が 63 件 (14.1%) → 42 件 (9.4%) に減少する。

TR については調査時に要対応の不一致として、本タイトルの語順の相違を追加した。しかし、データを検証したところ、表示の大小、記載位置などで本タイトルの語順に相違が生じていると推測され、それはルール之差ではなく、目録作成者の判断による相違によるものと見なすことができ、要対応の対象とは言えないと考えられる。その場合、TR の要対応の件数は 35 件 (7.2%) → 16 件 (3.3%) に減少する。

(2) JPMARC レコードの TR の修正が必要となる割合 (ISBN 有の場合)

(1) の結果から JPMARC の ISBN 有の場合、TR の修正が必要となる割合は、3.5 の計算式により、下記のとおりとなる。

修正が必要となる割合 = 2.14% → 47 冊に 1 冊

ただし、本タイトルの語順の相違を要対応の対象としない場合は、下記のとおりとなる。

修正が必要となる割合 = 0.87% → 115 冊に 1 冊

4.2 TRCMARC の調査結果

4.2.1 TRCMARC レコードと NC レコードの一致率

(1) TRCMARC 流用作成の NC レコードの抽出

3.3 のとおり、TRCMARC を流用して作成された NC レコードを抽出し、ISBN の有無によって分類した。結果は表 4-4 のとおりである。

表 4-4 TRCMARC 流用の NC レコード件数 (割合)

NC レコード件数	ISBN 有	ISBN 無
127,143	123,825 (97.4%)	3,318 (2.6%)

(2) TRCMARC レコードと NC レコードの各項目の一致率

3.3.(2) のとおり、ISBN の有無別に各項目の一致・不一致件数および一致率を調査した。結果は表 4-5、図 4-9 のとおりである。

表 4-5 TRCMARC と NC レコードの一致・不一致件数および一致率

項目	TR		TRR		PBLC_1		PBLCDT_1	
	有	無	有	無	有	無	有	無
一致件数	62,932	1,880	75,315	2,150	119,135	1,143	109,013	2,274
不一致件数	60,893	1,438	48,510	1,168	4,690	2,175	14,812	1,044
一致率	50.8%	56.7%	39.2%	35.2%	96.2%	34.4%	88.0%	68.5%

項目	PHYNM		PHYSZ		ISBN	
	有	無	有	無	有	無
一致件数	83,290	1,789	122,398	2,310	122,085	2,932
不一致件数	40,535	1,529	1,427	1,008	1,740	386
一致率	67.3%	53.9%	98.8%	69.6%	98.6%	88.4%

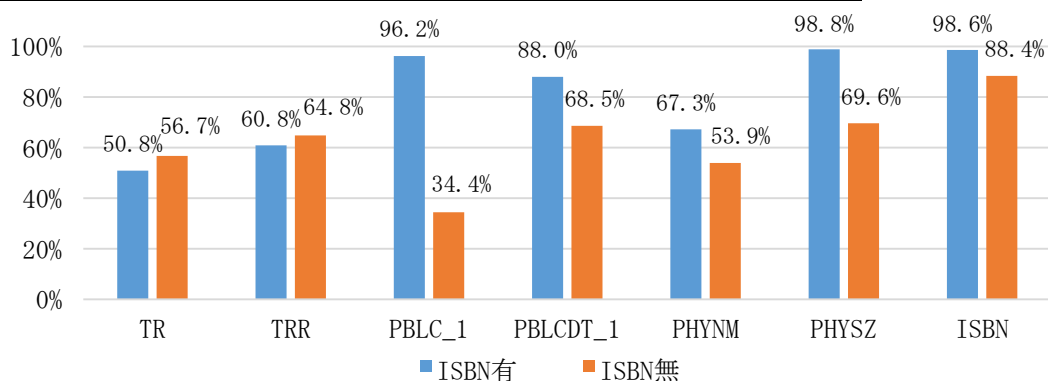


図 4-9 TRCMARC と NC レコードの一致率

ISBN 有と ISBN 無では、全般に ISBN 有の方が一致率が高いが、TR および TRR に関しては ISBN 無の方が一致率が高い。また、各項目の一致率の傾向は似ているが、PBLC_1 に関しては、ISBN 無の一致率が極端に低くなっている。ただし、PBLCDT、PHYNM、

PHYSZには、空値のための一致も含まれている。

4.2.2 TRCMARC レコードと NC レコードの不一致の分類

(1) TR の不一致の割合

3.4.(1)のとおり、のとおり、項目ごとにランダムに抽出した不一致レコード 500 件について、目視により不一致による分類を行い（付録 2）、その結果から不一致の割合を導き出した。TR の不一致の割合は図 4-10 のとおりである。

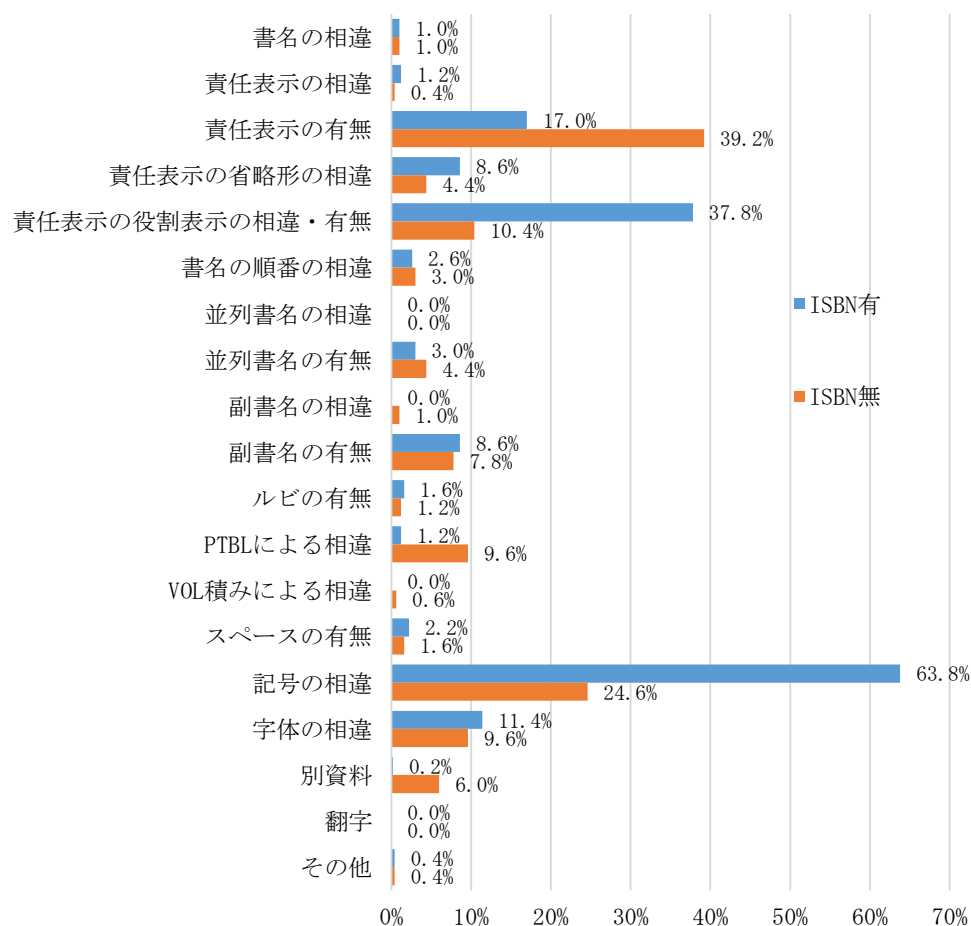


図 4-10 TR の不一致の割合

TR で不一致の割合が最も大きい内容は、ISBN 有の場合、記号の相違によるもので 63.8% を占める。その理由は、使用記号のルール相違にある。TRCMARC では、補記に使用するカッコが“ [] (亀甲カッコ) ”であり、責任表示では役割の同異に関わらず“ : (コロロン) ”を区切り記号として使用している。

次に不一致で割合が大きいのは、責任表示の有無によるもので ISBN 無では 39.2% である。責任表示のデータがないレコードを確認すると親書誌レコードの場合がほとんどである。今回の不一致項目の分類作業では、親書誌レコードであるため生じた責任表示の有無

は、PTBLによる相違ではなく、責任表示の有無に振り分けている。

責任表示の役割表示の相違および有無は ISBN 有で 37.8%である。これはルール相違によるもので、TRCMARC では 2 人以上の個人や団体が同一の役割をはたしているときも、各々に役割の種類を示す語を付与しているためである。

字体の相違が ISBN の有無とも 10%前後あるが、TRCMARC のルールでは常用漢字表に収録されている漢字は常用漢字表にある字体を使用し、旧字は新字に統一することになっているためである。また、ローマ数字は原則として使用せずアラビア数字で入力するというルールを採用している。

(2) TRR の不一致の割合

3.4.(2) のとおり、項目ごとにランダムに抽出した不一致レコード 500 件について、目視により不一致による分類を行い（付録 2）、その結果から不一致の割合を導き出した。TRR の不一致の割合は図 4-11 のとおりである。

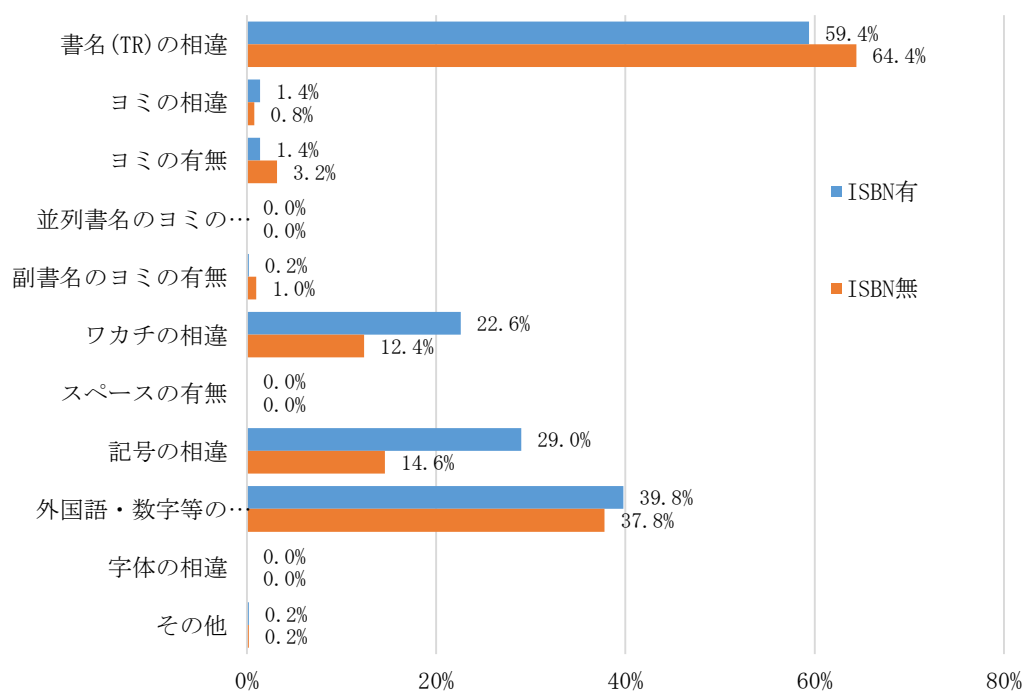


図 4-11 TRR の不一致の割合

TRR で不一致の割合が最も大きい内容は TR 中の書名自体の相違によるもので、これは本来のヨミの相違とは異なるため、不一致データの対象外とする。

次に不一致で割合が大きいのは、外国語や数字等のヨミの相違で、ISBN の有無とも 40% 近くを占める。TRCMARC のルールではラテン文字やアラビア数字にもカナでヨミを付与することになっているのに対し、NC のルールでは表示されているとおりをヨミとしている（発音には従わない）ためである。

記号の相違は ISBN 有で 29.0%である。NC レコードではヨミの場合も記号を入力しているが、TRCMARC レコードではヨミには記号を入力していないためである。

ワカチの相違は ISBN 有で 22.6%である。例えば、「～における～」という文節に対し、NC レコードでは“ニオケル”，TRC レコードでは“ニ オケル”とヨミが付与されている。

(3) PBLC_1 の不一致の割合

3.4.(3) のとおり、項目ごとにランダムに抽出した不一致レコード 500 件について、目視により不一致による分類を行い（付録 2），その結果から不一致の割合を導き出した。

PBLC_1 の不一致の割合は図 4-12 のとおりである。

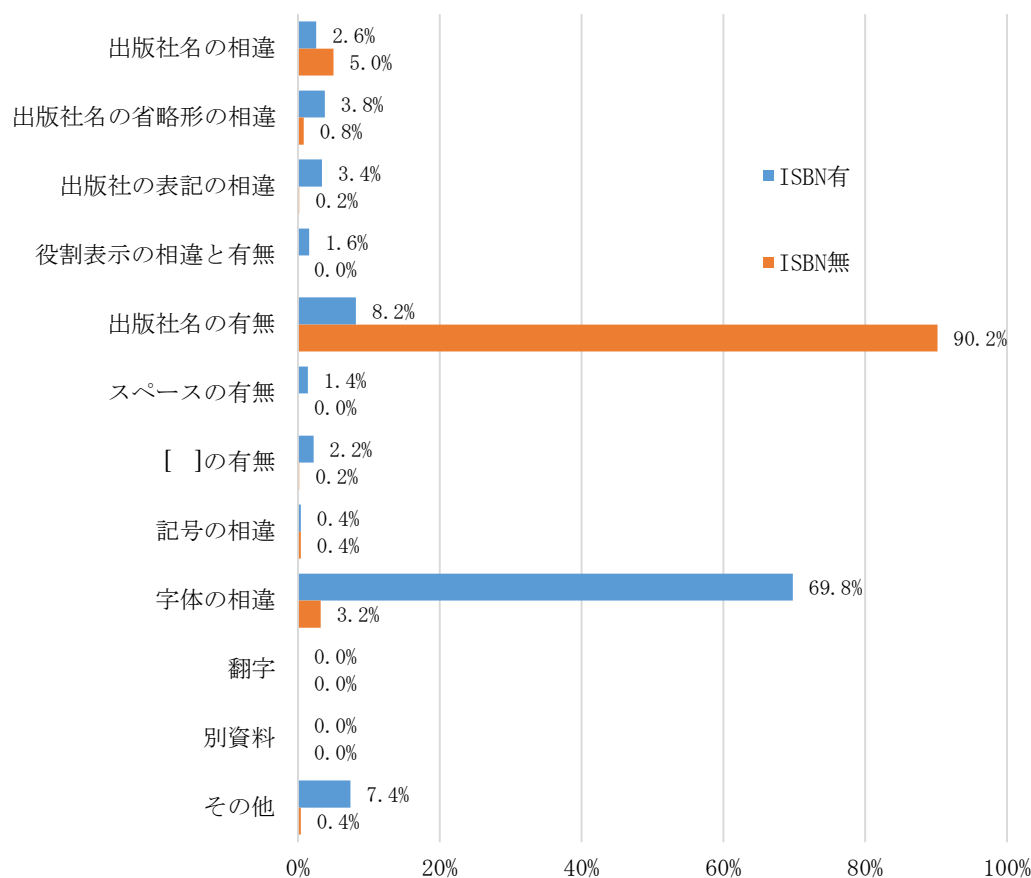


図 4-12 PBLC_1 の不一致の割合

PBLC_1 で不一致の割合が最も大きい内容は、出版社名の有無で ISBN 無では 90.2%を占める。全て TRCMARC に出版社データがないもので、親書誌であることに起因するものがほとんどである。

字体の相違が ISBN 有で 69.8%を占めるが、TR の場合と同様、TRCMARC のルールでは常用漢字表に収録されている漢字は常用漢字表にある字体を使用し、旧字は新字に統一することになっているためである。

(4) PBLCDT_1 の不一致の割合

3.4.(4)のとおり、項目ごとにランダムに抽出した不一致レコード 500 件について、目視により不一致による分類を行い（付録 2）、その結果から不一致の割合を導き出した。PBLCDT_1 の不一致の割合は図 4-13 のとおりである。

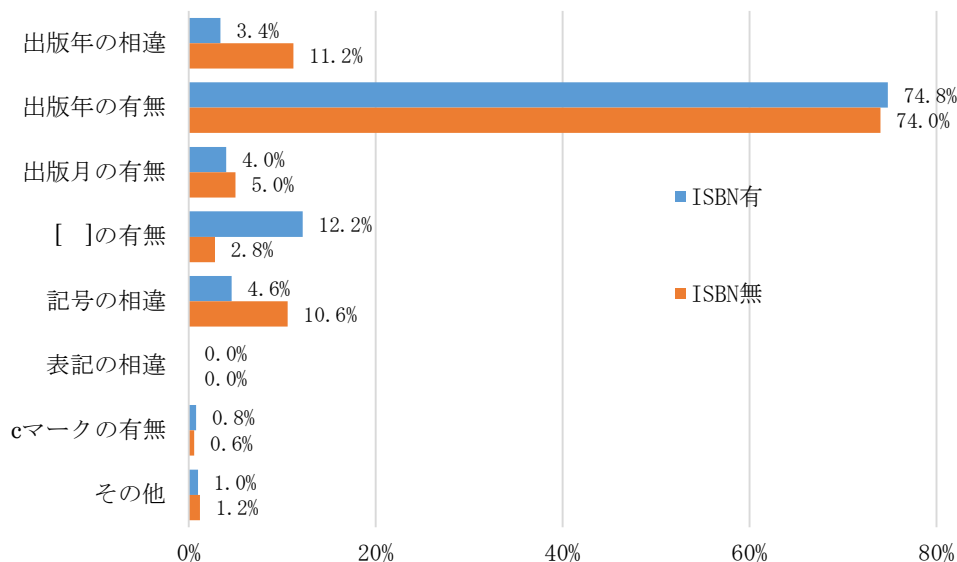


図 4-13 PBLCDT_1 の不一致の割合

PBLCDT_1 で最も不一致の割合が大きい内容は、出版年の有無によるもので ISBN の有無にかかわらず 75%近くを占める。ISBN 有の場合は、ほとんどの場合が NC レコード側に出版年が無い。これは、NC レコードに出版社のデータが複数ある場合に、PBLCDT_1 ではなく PBLCDT_2 以降に出版年のデータが入っているためであると推察できる。また、ISBN 無の場合は、TRCMARC レコード側に出版年のデータがなく、親書誌であることに起因するが多い。

(5) PHYNM の不一致の割合

3.4.(5)のとおり、項目ごとにランダムに抽出した不一致レコード 500 件について、目視により不一致による分類を行い（付録 2）、その結果から不一致の割合を導き出した。PHYNM の不一致の割合は図 4-14 のとおりである。

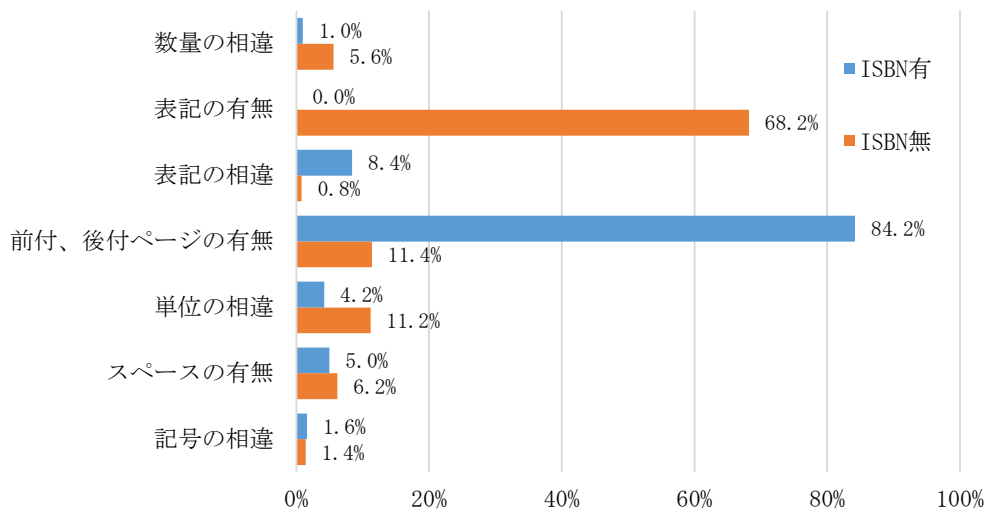


図 4-14 PHYNM の不一致の割合

PHYNM で不一致の割合が最も大きい内容は、前付・後付ページの有無で ISBN 有では 84.2% を占める。NC レコードに前付・後付ページのデータがあり、TRCMARC レコードに無い。これは TRCMARC のルールでは、前付・後付ページは入力しないことになっているためである。

次に不一致の割合が大きい内容は表記の有無で ISBN 無で 68.2% になる。多くの場合が TRCMARC にデータがなく、親書誌であることに起因するものである。

(6) PHYSZ の不一致の割合

3.4.(6) のとおり、項目ごとにランダムに抽出した不一致レコード 500 件について、目視により不一致による分類を行い(付録 2)、その結果から不一致の割合を導き出した。PHYSZ の不一致の割合は図 4-15 のとおりである。

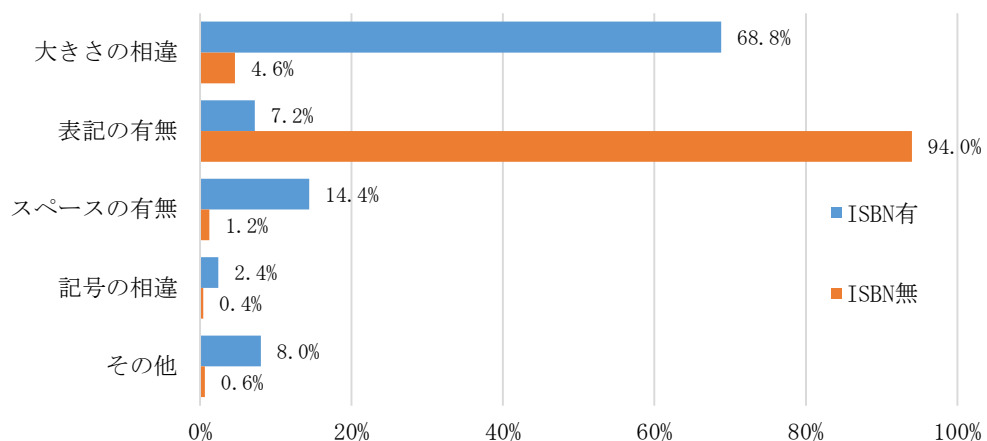


図 4-15 PHYSZ の不一致の割合

PHYSZ で不一致の割合が最も大きい内容は、表記の有無で ISBN 無で 94%を占める。多くの場合が TRCMARC にデータがなく、親書誌に起因するものである。

次に不一致の割合が大きい内容は大きさの相違で ISBN 有で 68.8%を占めるが、1cm 以下の相違が多い。

4.2.3 TRCMARC レコードと NC レコードの要対応の不一致 (ISBN 有の場合)

(1) TRCMARC レコードと NC レコードの要対応の不一致の割合 (ISBN 有の場合)

4.2.2 で分類した結果のうち、ISBN 有の場合について ISBN が一致するレコードを対象に、3.4 で要対応と判断した不一致の件数を調査した。結果は表 4-6 と図 4-16 のとおりである。

表 4-6 TRCMARC (ISBN 有) の要対応の不一致の件数 (割合)

項目	不一致	件数	項目	不一致	件数
TR	書名の相違	5	PBLCDT_1	出版年の相違	9
	責任表示の相違	6		出版年の有無	369
	並列書名の相違	0		計(割合)	378(78.6%)
	副書名の相違	0	PHYNM	数量の相違	2
	本タイトルの語順の相違 注)	11		表記の有無	0
	計(割合)	22(4.5%)		計(割合)	2(0.4%)
TRR	ヨミの相違	7	PHYSZ	大きさの相違	304
	計(割合)	7(1.4%)		計(割合)	304(70.4%)
PBLC_1	出版社名の相違	8			
	出版社名の有無	32			
	計(割合)	40(8.4%)			

注) 3.4.(1)の TR の不一致項目の分類では、書名の順番の相違は無視できるとしたが、和書の場合は検索結果に影響するため、本タイトルの語順の相違は要対応に追加した。

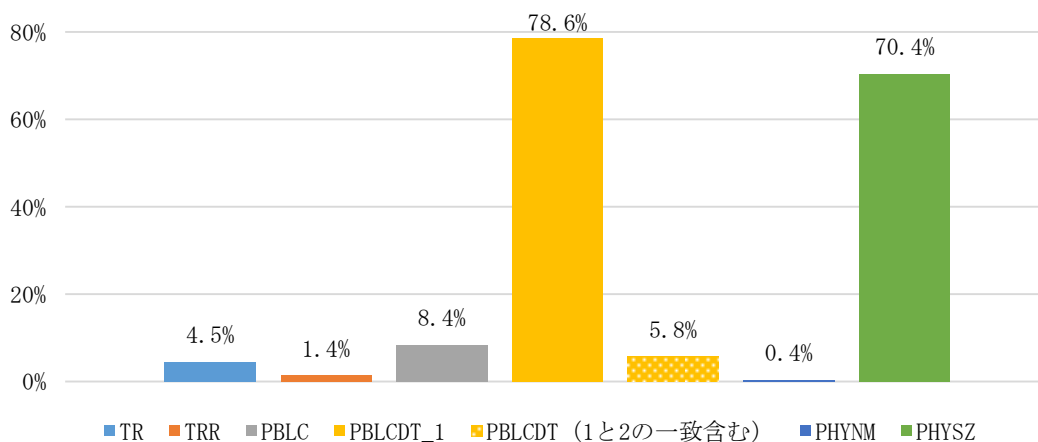


図 4-16 TRCMARC (ISBN 有) の各項目の要対応の不一致の割合

要対応の割合が最も大きい項目が、PBLCDT で 78.6%を占める。不一致としては、出版年の相違と出版年の有無の 2 つの場合があり、後者が不一致の大半を占めている。ただし、出版年の有無のうち出版社数の相違によるものは、PBLCDT_1 と PBLCDT_2 で出版年が一致する。これを一致とみなした場合は、出版年の有無が 369 件から 18 件に激減し、PBLCDT の要対応の不一致の件数が 378 件 (78.6%) →28 件 (5.8%) に減少する。

次に要対応の不一致の割合が大きい項目は PHYSZ で 70.4%である。その不一致は全て大きさの相違で 304 件に上っている。ただし、データを精査すると 1cm 以下の相違が多く、1cm 以下の相違を除くと、相違件数は 304 件 (70.4%) →12 件 (2.8%) に減少する。採寸の誤差、もしくは刷による若干のサイズの差の場合もあるものと推察できる。

TR については調査時に要対応の不一致として、本タイトルの語順の相違を追加した。しかし、データを検証したところ、表示の大小、記載位置などで本タイトルの語順に相違が生じていると推測でき、それはルールの差ではなく、目録作成者の判断による相違によるものと見なすことができ、要対応の不一致の対象とは言えないと考えられる。その場合、TR の要対応の不一致の件数は 22 件 (4.5%) から 11 件 (2.2%) に減少する。

(2) TRCMARC レコードの TR の修正が必要となる割合 (ISBN 有の場合)

(1) の結果から TRCMARC の ISBN 有の場合、TR の修正が必要となる割合は、3.5 の計算式により、下記のとおりとなる。

修正が必要となる割合 =2.18% → 46 冊に 1 冊

ただし、本タイトルの語順の相違を要対応の不一致の対象としない場合は、下記のとおりとなる。

修正が必要となる割合 =1.09% → 92 冊に 1 冊

4.3 USMARC の調査結果

4.3.1 USMARC レコードと NC レコードの一致率

(1) USMARC 流用作成の NC レコード抽出

3.3 のとおり、USMARC を流用して作成された NC レコードを抽出し、ISBN の有無によって分類した。結果は表 4-7 のとおりである。

表 4-7 USMARC 流用の NC レコード件数 (割合)

NC レコード件数	ISBN 有	ISBN 無
95,630	90,711 (94.9%)	4,919 (5.1%)

(2) MARC レコードと NC レコードの各項目の一致率

3.3.(2) のとおり、ISBN の有無別に各項目の一致・不一致件数および一致率を調査した。結果は表 4-8、図 4-17 のとおりである。

表 4-8 USMARC と NC レコードの一致・不一致件数および一致率

項目	TR_1		PBLC_1		PBLCDT_1		PHYNM		PHYSZ		ISBN ^{注)}		ISBNA	
	有	無	有	無	有	無	有	無	有	有	有	無	有	無
一致件数	63,331	2,678	64,570	3,006	60,802	3,821	55,466	3,876	68,324	37,792	37,792	4,246	80,304	4,221
不一致件数	27,380	2,241	26,141	1,913	29,909	1,098	35,245	1,043	22,387	52,919	52,919	673	10,407	698
一致率	69.8%	54.4%	71.2%	61.1%	67.0%	77.7%	61.1%	78.8%	75.3%	41.7%	86.3%	88.5%	85.8%	

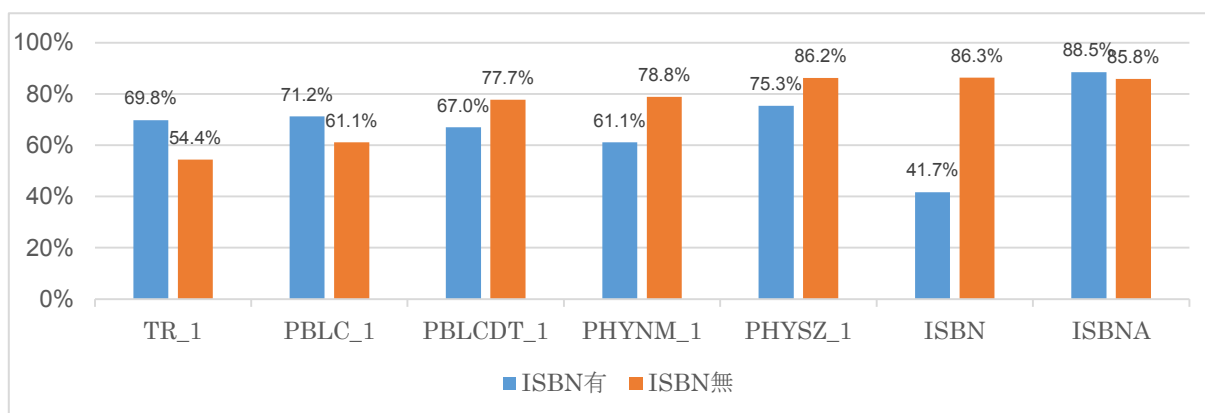


図 4-17 USMARC と NC レコードの一致率

注) ISBN なしにおける「ISBN の一致率」とは、空白または Null 値で一致していることを意味する。

ISBN 有と無の一致率を比較すると、TR と PBLC_1 については、ISBN 有が高いが、その他については ISBN 無の方が上回っており、特に ISBN 有が一致率について高い傾向にあるわけではない。ただし、PBLCDT, PHYNM, PHYSZ については空値による一致も含まれている。

ISBNA について補足する。MARC21 から CATP へのマッピング表によると、

ISBN=TAG020\$a

' (, ')' で囲まれている部分が存在し、かつ その中に以下の文字列が含まれる場合、' (, ')' で囲まれている値を、' (, ')' 含めて削除する。

' *fasc.;*'no.;

'*pt.;*'t.;

'*v.;*'vol.;

'*Abt.;*'*Bd.;*'*Folge;*'*Heft;*'*Lfg.;*'*T.;*'*V.;*'*pbk.;*'*set*'

カッコの中に例示以外の文字を含むものは、すべて ISBNA にマッピングされているため、ISBN に該当するものが ISBNA に入っている場合が多い。ISBN 有における ISBN_1 同士の一一致率は 41.7%であるが、ISBN_1(NC レコード)と ISBNA_1(参照 MARC)の一一致率は 88.5%に上昇する。

4.3.2 USMARC レコードと NC レコードの不一致内容の分類

(1) TR の不一致内容

3.4.(1)のとおり、項目ごとにランダムに抽出した不一致レコード 500 件について、目視により不一致内容による分類を行い(付録 3)、その結果から不一致内容の割合を導き出した。

TR の不一致内容の割合は図 4-18 のとおりである。

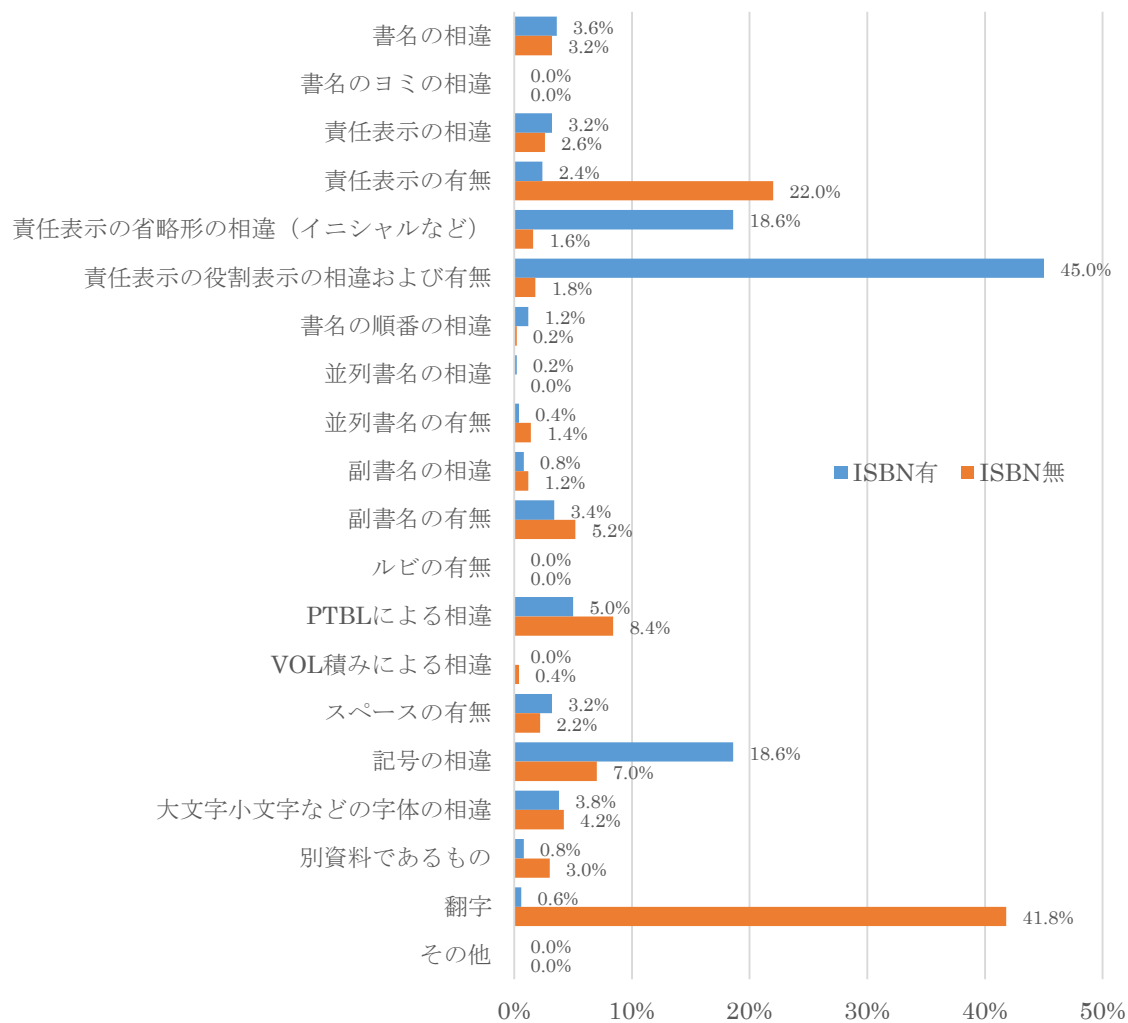


図 4-18 TR の不一致内容の割合

TR で最も不一致の割合が大きい内容は ISBN 有で「翻字」、ISBN 無では「責任表示の役割表示の相違および有無」である。

この違いは ISBN 有だけに「YR=2014」で絞り込んだことと、RDA 適用 (2013.3~) が影響していると考えられる。3.3.(3)のとおり、ISBN 有にのみ YR=2014 で絞込みをかけたため、ISBN 有には PDA 適用後のデータが、ISBN 無よりも多く含まれていると推測され

る。RDA 適用後は、転記を本則している²ため、ISBN 有には翻字による違いがほとんど確認されなかった。

責任表示の役割表示の相違および有無について、RDA 適用後は転記が原則となったため、著者の所属や身分などが責任表示に記述されるようになり、流用登録の際にそれらを削除したために ISBN 有に多くの不一致が確認された。責任表示の省略形の相違（イニシャルなど）も同様に、ISBN 有により多くの不一致が確認されたと考えられる。

続いて「責任表示の有無」について、ISBN 無のレコードを確認すると、NC レコード側が親書誌であることが多かった。今回の不一致項目の分類作業では、親書誌レコードであるため生じた責任表示の有無による違いは、「PTBL による相違」ではなく、「責任表示による有無」に振り分けている。

² 和中幹雄「RDA と国際化・個人名の典拠系アクセスポイントの形式をめぐって」『資料組織化研究-e』 No.66(2015.03)

(2) PBLC_1 の不一致内容

3.4.(3)のとおり、項目ごとにランダムに抽出した不一致レコード 500 件について、目視により不一致内容による分類を行い(付録 3), その結果から不一致内容の割合を導き出した。PBLC_1 の不一致内容の割合は図 4-19 のとおりである。

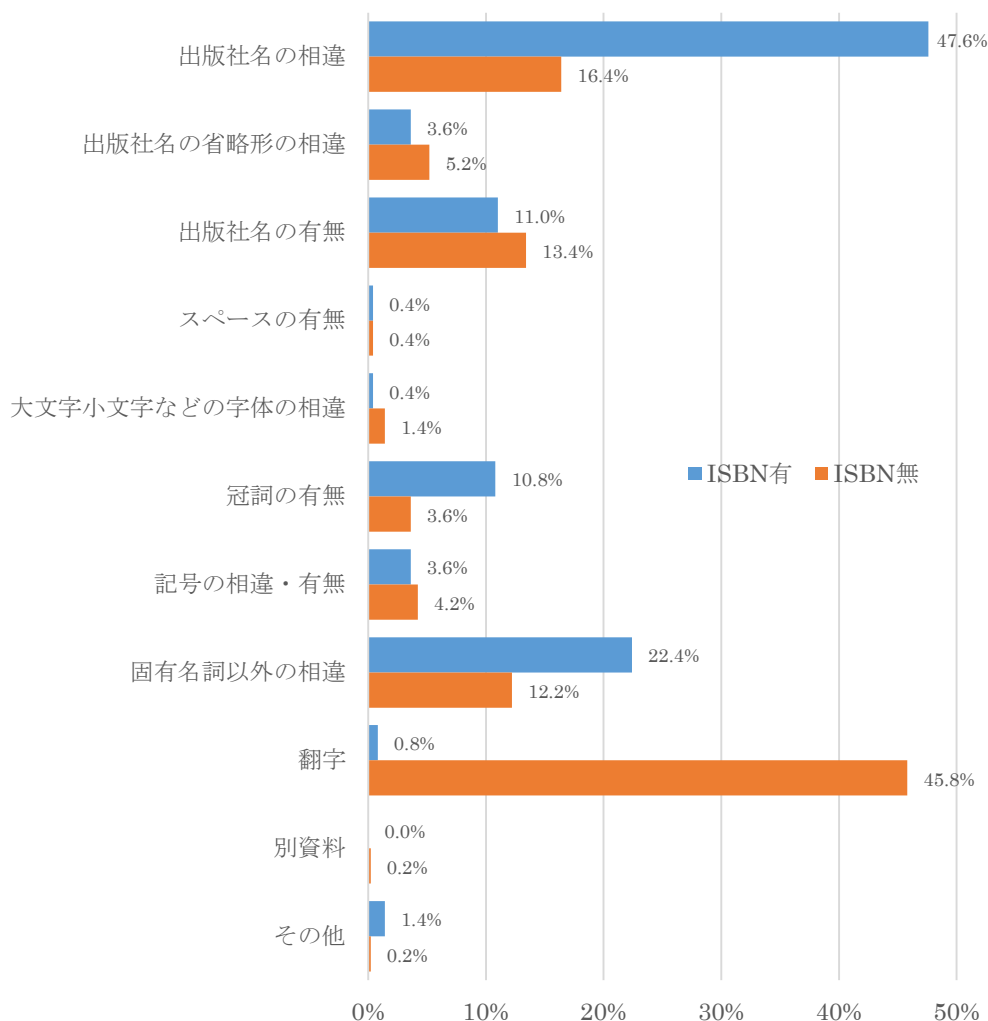


図 4-19 PBLC_1 の不一致内容の割合

PBLC_1 でもっとも不一致の割合が大きい内容は、ISBN 有で出版社の相違、ISBN 無で翻字であった。ISBN 無で翻字の割合が大きいのは、TR と同様に、ISBN 有のみに「YR=2014」のフィルタをかけたため、RDA 適用 (2013.3～) の影響を受けたためと考えられる。

ISBN 有について出版社の相違が 47.6%を占めている。そこから親書誌または PTBL リンクを持つ子書誌である数を除くと、ISBN 有における出版社の相違が占める割合は、238 件→123 件に減少する。ISBN 無においても、82 件→29 件に減少する。

ISBN 無において出版社名の有無は 13.4%を占めているが、そのほとんどが、NC レコー

ド側が親書誌であることに起因している。親書誌または PTBL リンクを持つ子書誌であるレコードを除くと、ISBN 無における出版社の有無が占める割合は、67 件→22 件に減少する。ISBN 有においては、55 件→31 件に減少する。

(3) PBLCDT_1 の不一致内容

3.4.(4)のとおり、項目ごとにランダムに抽出した不一致レコード 500 件について、目視により不一致内容による分類を行い(付録 3)、その結果から不一致内容の割合を導き出した。PBLCDT_1 の不一致内容の割合は図 4-20 のとおりである。

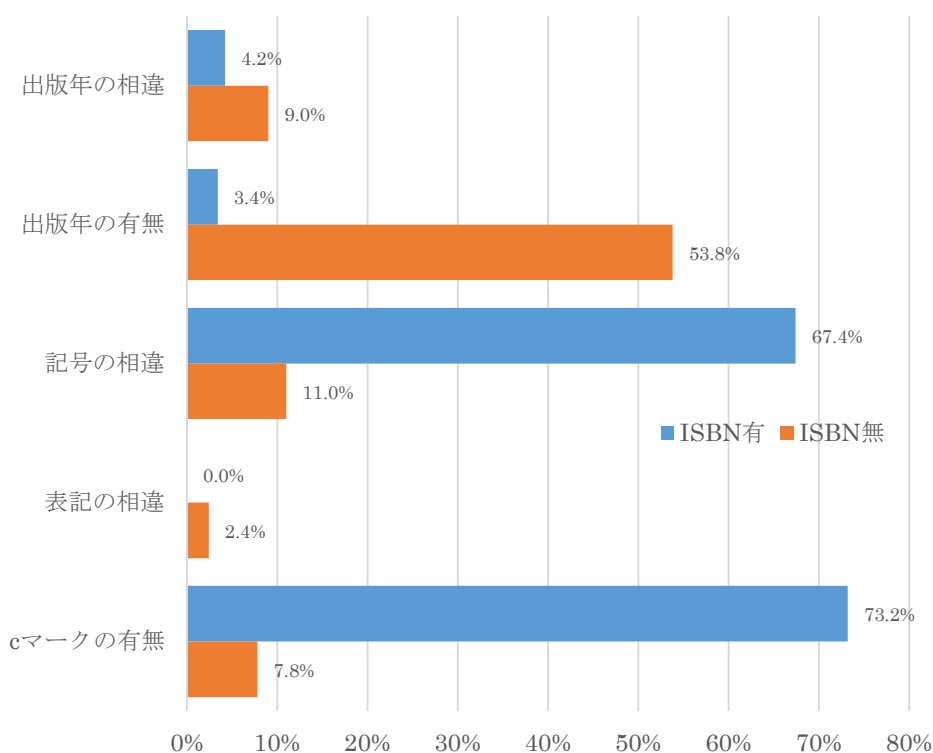


図 4-20 PBLCDT_1 の不一致内容の割合

PBLC_1 でもっとも不一致の割合が大きい内容は、ISBN 有で c マークの有無と記号の相違、ISBN 無で出版年の有無である。ISBN 有のレコードを確認すると、記号の相違 337 件のうち、334 件は NC レコード側の補記を示す “[]” の有無による。

ISBN 無において出版年の有無は 53.8%を占めているが、そのほとんどが、NC レコード側が親書誌であることに起因している。親書誌または PTBL リンクを持つ子書誌であるレコードを除くと、ISBN 無における表記の有無が占める割合は、269 件→63 件に減少する。

(4) PHYNM の不一致内容

3.4.(5)のとおり、項目ごとにランダムに抽出した不一致レコード 500 件について、目視により不一致内容による分類を行い(付録 3), その結果から不一致内容の割合を導き出した。PHYNM の不一致内容の割合は図 4-21 のとおりである。

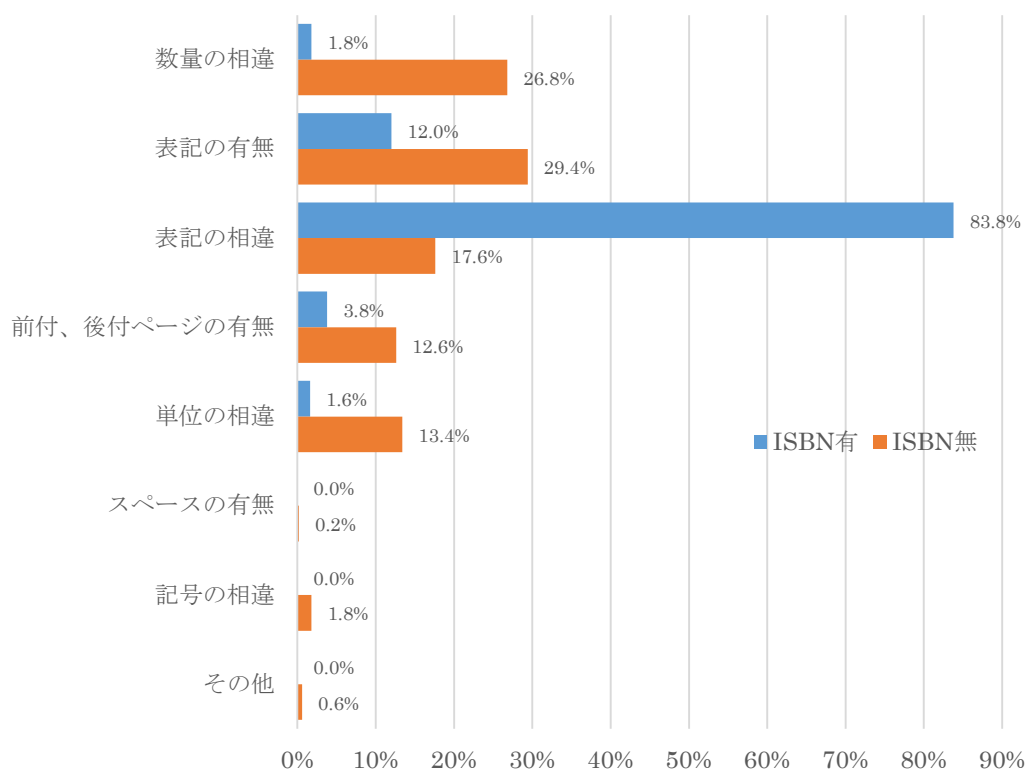


図 4-21 PHYNM の不一致内容の割合

PHYNM でもっとも不一致の割合が大きい内容は、ISBN 有で表記の相違、ISBN 無で数量の相違である。

表記の相違は主に「p.」と「page」によるもので、各レコードが準拠しているルールの違いによる。

ISBN 無において表記の有無は 29.4%を占めているが、そのほとんどが、NC レコード側が親書誌であることに起因している。親書誌または PTBL リンクを持つ子書誌であるレコードを除くと、ISBN 無における表記の有無が占める割合は、147 件→6 件に減少する。

(5) PHYSZ の不一致内容

3.4.(6)のとおり，項目ごとにランダムに抽出した不一致レコード 500 件について，目視により不一致内容による分類を行い(付録 3)，その結果から不一致内容の割合を導き出した。PHYSZ の不一致内容の割合は図 4-22 のとおりである。

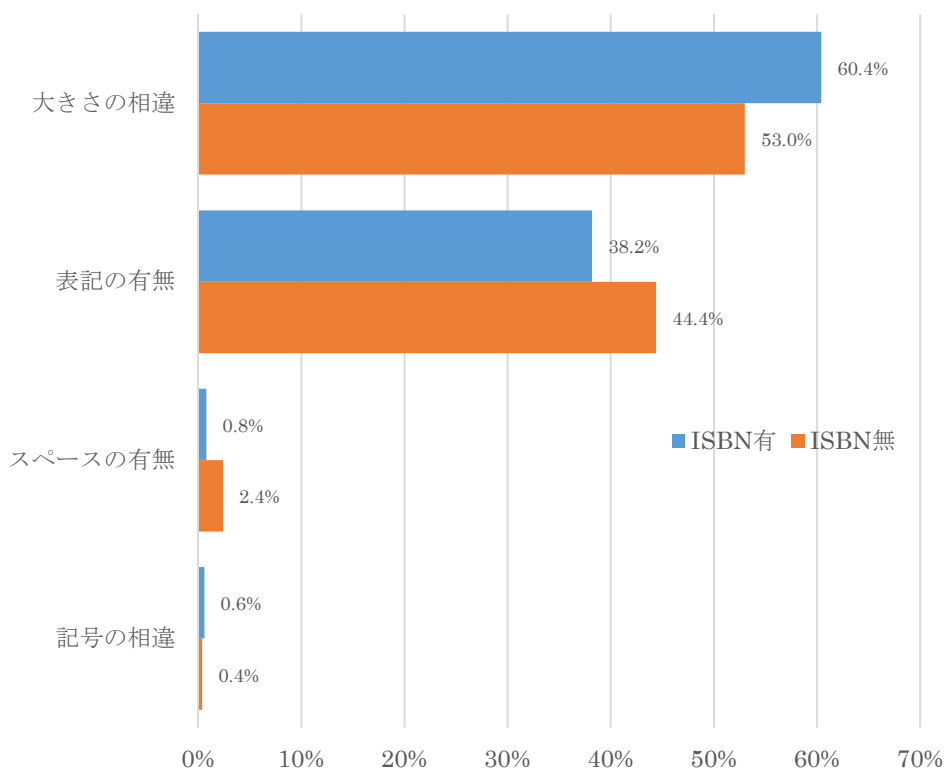


図 4-22 PHYSZ の不一致内容の割合

PHYSZ でもっとも不一致の割合が大きい内容は，ISBN 有無によらず，大きさの相違である。AACR2 も NC のルールも小数点以下は切り上げであり，ルールの違いによる差異ではないと考えられる。ただし，ISBN 有において，大きさの相違の 302 件のうち，1cm の差であるものは 271 件である。ISBN 無において，大きさの相違の 265 件のうち，1cm の差が占めるのは 171 件である。

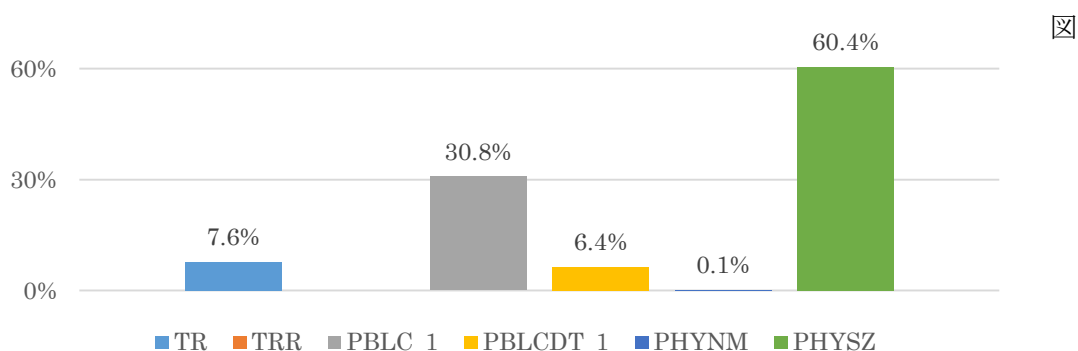
4.3.3 USMARC レコードと NC レコードの要対応の不一致 (ISBN 有の場合)

(1) USMARC レコードと NC レコードの要対応の不一致の割合 (ISBN 有の場合)

4.3.2 で分類した結果のうち、ISBN 有の場合について、ISBN が一致するレコードを対象に、3.4 で要対応と判断した不一致内容の件数を調査した。ここでは、PBLC、PBLCDT_1、PHYNM については、階層構造を持ち、親書誌または PTBL リンクを持つ子書誌のいずれかであるものは除外した。結果は表 4-9 と図 4-23 のとおりである。

表 4-9 USMARC (ISBN 有) の要対応の不一致の件数

項目	不一致内容	件数	項目	不一致内容	件数
TR	書名の相違	18	PBLCDT_1	出版年の相違	21
	責任表示の相違	16		出版年の有無	17
	並列書名の相違	1		小計	38(7.6%)
	副書名の相違	4	PHYNM	数量の相違	9
	小計 ³	38(7.6%)		表記の有無	23
PBLC_1	出版社名の相違	123	小計	32(6.4%)	
	出版社名の有無	31	PHYSZ	大きさの相違	302
	小計	154(30.8%)		小計	302(60.4%)



4-23 USMARC (ISBN 有) の各項目の要対応の不一致の割合

(2) USMARC レコードの TR の修正が必要となる割合 (ISBN 有の場合)

(1) の結果から USMARC の ISBN 有の場合、TR の修正が必要となる割合は、3.5 の計算式により、下記のとおりとなる。

$$\text{修正が必要となる割合} = 6.02\% \rightarrow 17 \text{ 冊に } 1 \text{ 冊}$$

³ 1 レコードに対して複数の分類項目を付与しているため、不一致内容の総数と異なる

4.4 UKMARC の調査結果

4.4.1 UKMARC レコードと NC レコードの一致率

(1) UKMARC 流用作成の NC レコードの抽出

3.3 のとおり、UKMARC を流用して作成された NC レコードを抽出し、ISBN の有無によって分類した。結果は表 4-10 のとおりである。

表 4-10 UKMARC 流用の NC レコード件数 (割合)

NC レコード件数	ISBN 有	ISBN 無
40,736	39,756 (97.6%)	980 (2.4%)

(2) UKMARC レコードと NC レコードの各項目の一致率

3.3.(2) のとおり、ISBN の有無別に各項目の一致・不一致件数および一致率を調査した。結果は表 4-11、図 4-24 のとおりである。

表 4-11 UKMARC と NC レコードの一致・不一致件数および一致率

項目	TR		PBLC_1		PBLCDT_1	
	有	無	有	無	有	無
一致件数	22,736	495	28,966	751	21,051	733
不一致件数	17,020	485	10,790	229	18,705	247
一致率	57.2%	50.5%	72.9%	76.6%	53.0%	74.8%

項目	PHYNM		PHYSZ		ISBN	
	有	無	有	無	有	無
一致件数	15,896	627	27,599	768	36,727	685
不一致件数	23,860	353	12,157	212	3,029	295
一致率	40.0%	64.0%	69.4%	78.4%	92.4%	69.9%

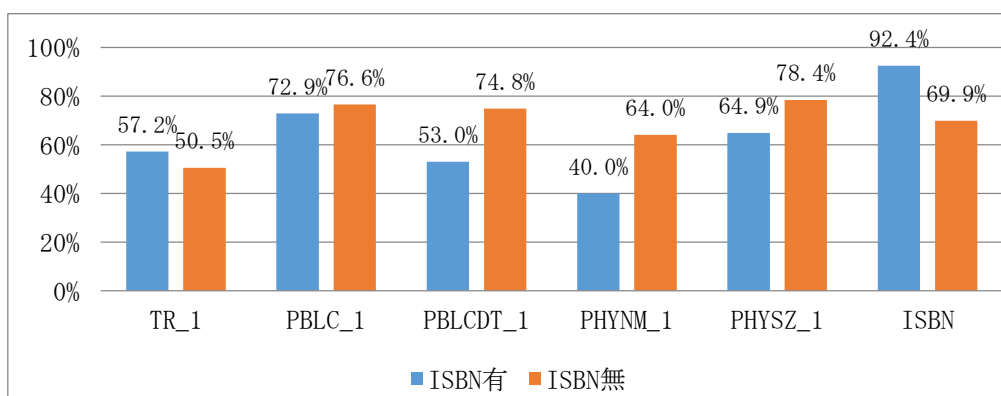


図 4-24 UKMARC と NC レコードの一致率

TR と ISBN 以外では、ISBN 無の一致率が高いが、PBLCDT、PHYNM、PHYSZ には、空値のための一致も含まれている。TR については、わずかに ISBN 有の一致率が高くなった。

ているものの、ほとんど変わらない。ISBNの有無が各項目の一致率に影響を与えているわけではないようであった。

4.4.2 UKMARC レコードと NC レコードの不一致内容の分類

(1) TR の不一致内容

3.4.(1)のとおり、項目ごとにランダムに抽出した不一致レコード 500 件（ただし ISBN 無は 486 件）について、目視により不一致内容による分類を行い（付録 4）、その結果から不一致内容の割合を導き出した。TR の不一致内容の割合は図 4-25 のとおりである。

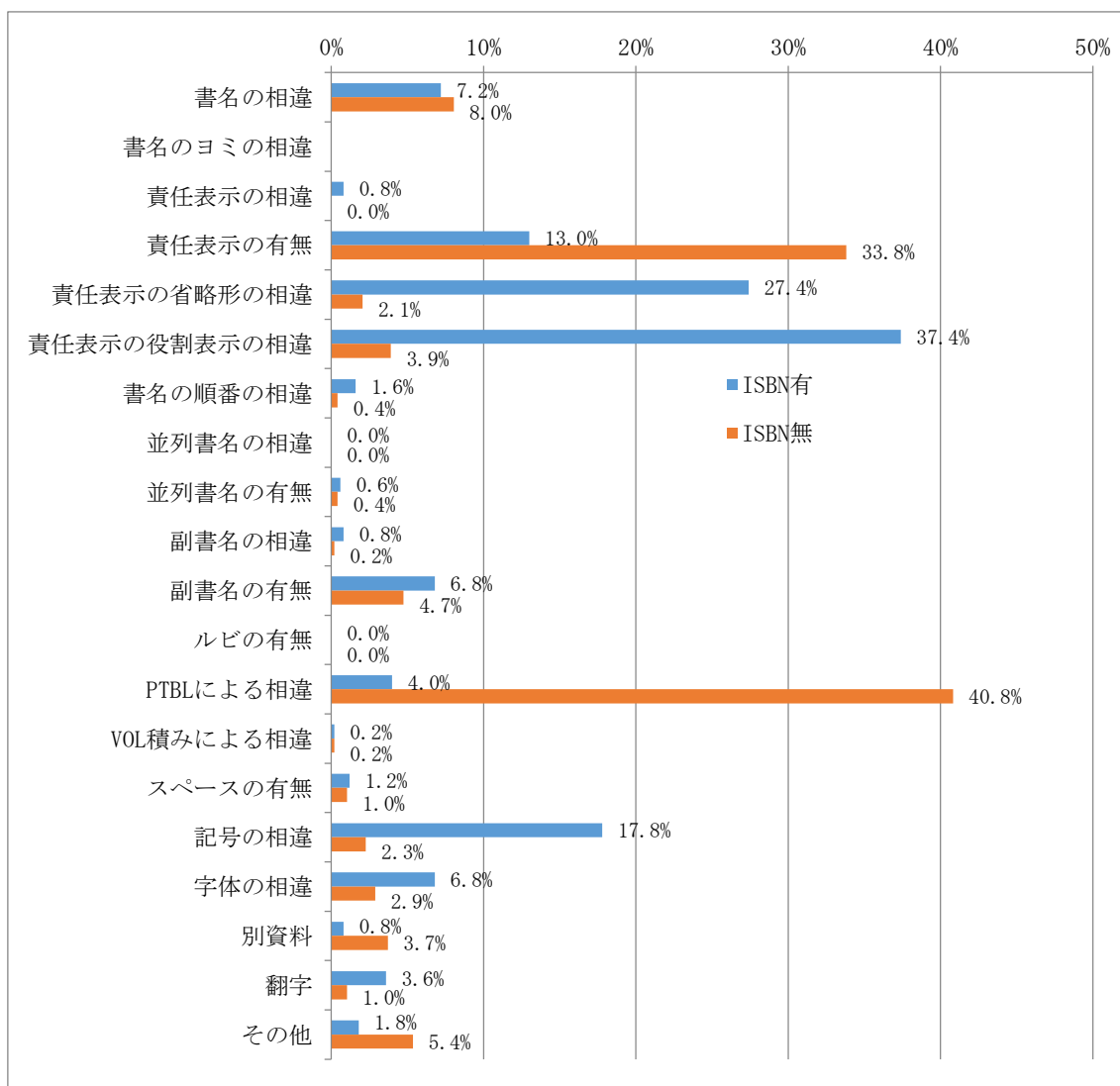


図 4-25 TR の不一致内容の割合

TR で最も不一致の割合が大きい内容は ISBN 有で「責任表示の役割の相違」および「責任表示の省略形の相違」、ISBN 無では「PTBL による相違」および「責任表示の有無」である。

ISBN 有については、USMARC と同様、ISBN 有だけを「YR=2014」で絞り込んだことと、RDA 適用 (2013.3～) が影響していると考えられる (4.3.2 (1) 参照)。すなわち、RDA 適用後は転記が原則となったが、NC では著者の所属などの表記は除かれるなどのため、ISBN 有の書誌で責任表示の役割と責任表示の省略形に多くの相違が確認された。

ISBN 無の責任表示の有無については、USMARC と同様に NC レコード側に親書誌が多いことが原因である。今回の作業では、書名が違うことなどから NC レコードが親子関係にある書誌だと確認できた場合は、PTBL による相違に分類したが、親書誌レコードであるために生じた責任表示の有無は、PTBL による相違ではなく、責任表示による有無に振り分けた。そのため、ISBN 無の相違については、親子関係を持つ書誌の作成に依拠するところが大きいと言える。

(2) PBLC_1 の不一致内容

3.4.(3)のとおり、項目ごとにランダムに抽出した不一致レコード 500 件について、目視により不一致内容による分類を行い(付録 4)、その結果から不一致内容の割合を導き出した。PBLC_1 の不一致内容の割合は図 4-26 のとおりである。

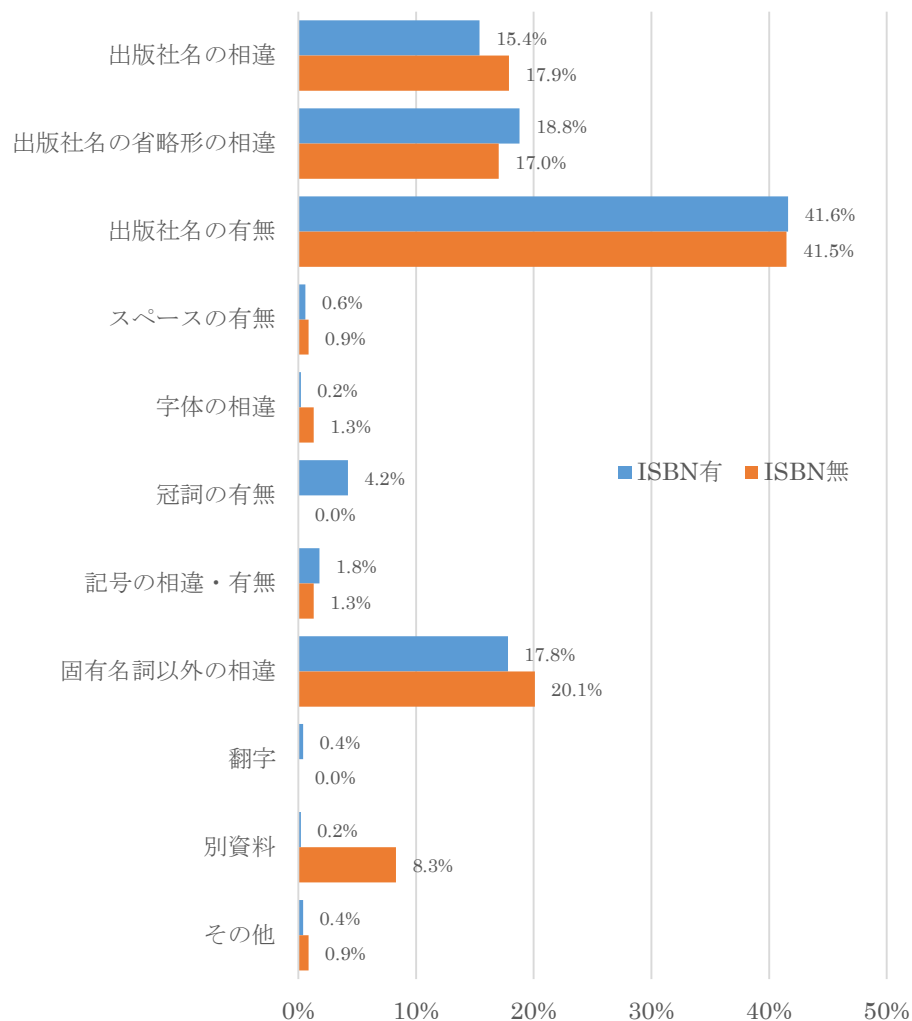


図 4-26 PBLC_1 の不一致内容の割合

PBLC_1でもっとも不一致の割合が大きい内容は、ISBN有・無ともに「出版社名の有無」である。ただし、ISBN無では、親書誌またはPTBLリンクを持つ子書誌である数を除くと、出版社名の有無による不一致の件数は95件→5件に減少する。このため、TRと同様に、親子関係を持つ書誌作成により不一致が生じていることが分かる。なお、親書誌・子書誌の判断については、次の2つと同じ書誌IDを持つデータとした。(1)TRの不一致確認でPTBLによる相違と分類した書誌。または、(2)ISBN無でNCレコードのPHYNMがv.となっている書誌（NCの親書誌作成のルールに基づく）。

一方ISBN有では、親書誌・子書誌を除いても208件→207件と、1件しか減少しないが、これは、上記(2)によるPTBLの確認ができないためもあると考えられる。試みに、出版社名の有無による不一致データをISBN有でランダムに10件抜き出して確認したところ、4件がPTBLリンクを持つ子書誌であった。そのため、ISBN有についても親子関係を持つ

書誌作成により不一致が生じている可能性が高いと言える。

(3) PBLCDT_1 の不一致内容

3.4.(4)のとおり、項目ごとにランダムに抽出した不一致レコード 500 件について、目視により不一致内容による分類を行い(付録 4)、その結果から不一致内容の割合を導き出した。PBLCDT_1 の不一致内容の割合は図 4-27 のとおりである。

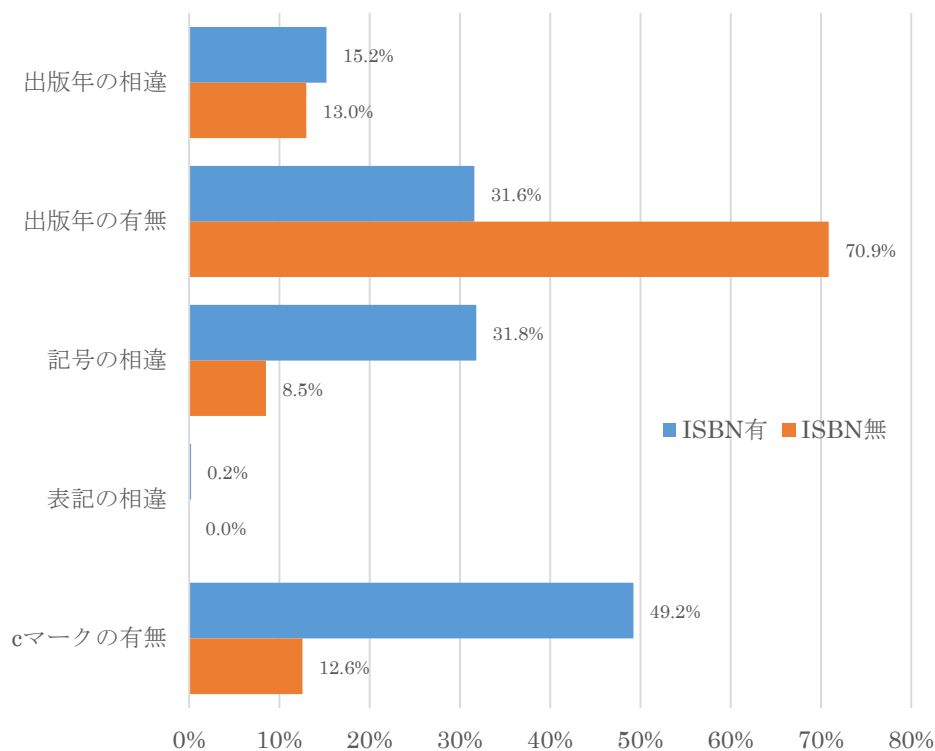


図 4-27 PBLCDT_1 の不一致内容の割合

PBLCDT_1 については、ISBN 無の「出版年の有無」による不一致の割合が最も大きい。ただし、これも親書誌または PTBL リンクを持つ子書誌の数を除くと、不一致件数は 175 件→16 件に減少する。その他、ISBN 有の「c マークの有無」による不一致も多く見られた。

(4) PHYNM の不一致内容

3.4.(5)のとおり、項目ごとにランダムに抽出した不一致レコード 500 件について、目視により不一致内容による分類を行い(付録 4)、その結果から不一致内容の割合を導き出した。PHYNM の不一致内容の割合は図 4-28 のとおりである。

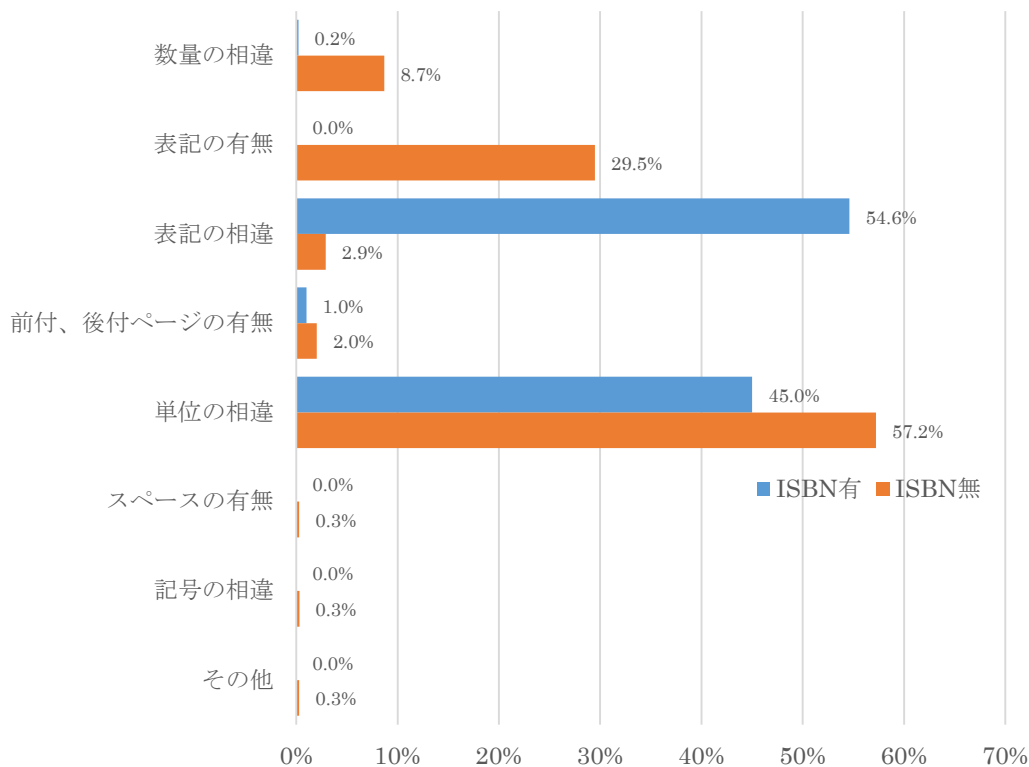


図 4-28 PHYNM の不一致内容の割合

PHYNM の ISBN 有でもっとも不一致の割合が大きい内容は、「表記の相違」である。これは、主に「p.」と「page」によるもので、各レコードが準拠しているルールの違いによる。ISBN 無でこの不一致が少ないのは、PTBL リンクを持つ書誌が多いため、v.で表される書誌が多いことによる。

全体の中では、「単位の相違」による不一致の割合が大きい。これには、UKMARC では volume で表記されているが、NC レコードでは p.で表記されているデータなどが含まれる。ISBN 有・無ともに他 2 つの海外 MARC と比較して、この単位の相違が多くみられることから、UKMARC の不一致の特徴であると考えられる。

その他、「数量の相違」について不一致の割合が低いことも、UKMARC の特徴である。

(5) PHYSZ の不一致内容

3.4.(6)のとおり、項目ごとにランダムに抽出した不一致レコード 500 件について、目視により不一致内容による分類を行い(付録 4)、その結果から不一致内容の割合を導き出した。PHYSZ の不一致内容の割合は図 4-29 のとおりである。

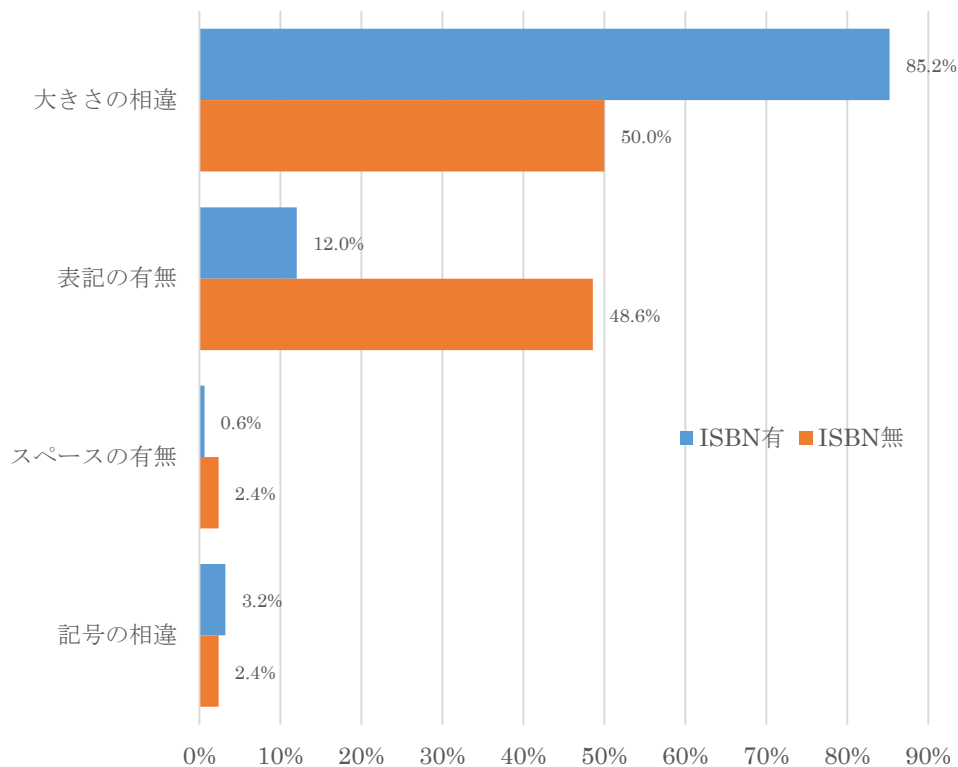


図 4-29 PHYSZ の不一致内容の割合

PHYSZ でもっとも不一致の割合が大きい内容は、大きさの相違である。AACR2 も NC のルールも小数点以下は切り上げであり、ルールの違いによる差異ではない。

ISBN 無の表記の有無に関する不一致の割合も大きいですが、ここでも親書誌または PTBL リンクを持つ子書誌の数を除くと、不一致件数は 103 件→8 件に減少する。

4.4.3 UKMARC レコードの要対応の不一致 (ISBN 有の場合)

(1) 4.4.2 で分類した結果のうち、ISBN 有の場合について、ISBN が一致するレコードを対象に、3.4 で要対応と判断した不一致の件数を調査した。この調査では、PBLC_1, PBLCDT_1, PHYNM については、階層構造を持ち、親書誌または PTBL リンクを持つ子書誌のいずれかであるものは除外した。結果は表 4-12 と図 4-30 のとおりである。

表 4-12 UKMARC (ISBN 有) の要対応の不一致の件数

項目	不一致内容	件数	項目	不一致内容	件数
TR	書名の相違	36	PBLCDT_1	出版年の相違	74
	責任表示の相違	4		出版年の有無	155
	並列書名の相違	0		小計	229(45.8%)
	副書名の相違	4	PHYNM	数量の相違	1
	小計 ⁴	43(8.6%)		表記の有無	0
PBLC_1	出版社名の相違	77	小計	1(0.2%)	
	出版社名の有無	207	PHYSZ	大きさの相違	426
	小計	284(56.8%)		小計	426(85.2%)

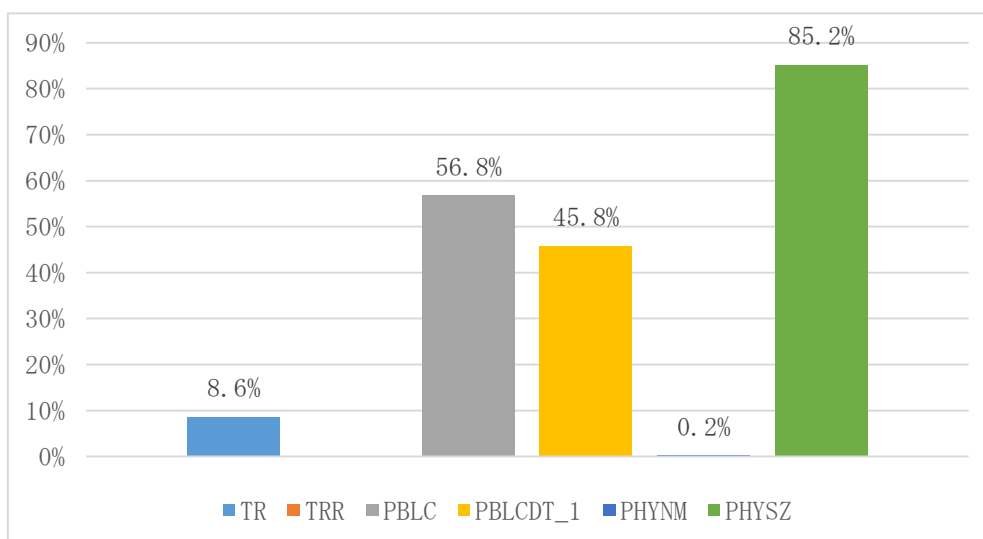


図 4-30 UKMARC (ISBN 有) の要対応の不一致の割合

(2) UKMARC レコードの TR の修正が必要となる割合 (ISBN 有の場合)

(1) の結果から UKMARC の ISBN 有の場合、TR の修正が必要となる割合は、3.5 の計算式により、下記のとおりとなる。

修正が必要となる割合 = 3.97% → 25 冊に 1 冊

⁴ 1 レコードに対して複数の分類項目を付与しているため、不一致内容の総数と異なる

4.5 DNIMARC の調査結果

4.5.1 DNIMARC レコードと NC レコードの一致率

(1) DNIMARC 流用作成の NC レコードの抽出

3.3 のとおり、DNIMARC を流用して作成された NC レコードを抽出し、ISBN の有無によって分類した。結果は表 4-13 のとおりである。

表 4-13 DNIMARC 流用の NC レコード件数 (割合)

NC レコード件数	ISBN 有	ISBN 無
22,337	21,398 (95.8%)	939 (4.2%)

(2) DNIMARC レコードと NC レコードの各項目の一致率

3.3.(2) のとおり、ISBN の有無別に各項目の一致・不一致件数および一致率を調査した。結果は表 4-14、図 4-31 のとおりである。

表 4-14 DNIMARC と NC レコードの一致・不一致件数および一致率

項目	TR		PBLC_1		PBLCDT_1	
	有	無	有	無	有	無
一致件数	8,124	507	11,751	485	8,861	793
不一致件数	13,274	432	9,647	454	12,537	146
一致率	38.0%	54.0%	54.9%	51.7%	41.4%	84.5%

項目	PHYNM		PHYSZ		ISBN	
	有	無	有	無	有	無
一致件数	47	705	12,705	779	20,817	826
不一致件数	21,351	234	8,693	160	581	113
一致率	0.2%	75.1%	59.4%	83.0%	97.3%	88.0%

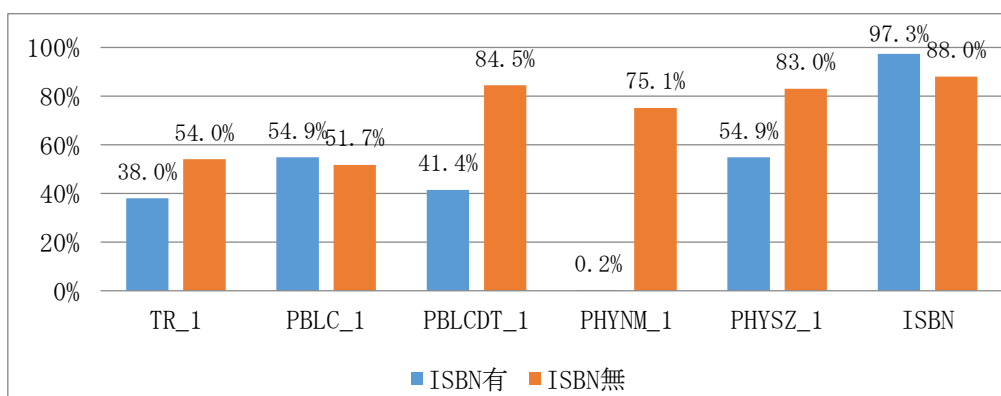


図 4-31 DNIMARC と NC レコードの一致率

TR, PBLCDT, PHYNM, PHYSZ で、ISBN 無の一致率が高くなっている。ただし、PBLCDT, PHYNM, PHYSZ には空値のための一致も含まれている。PBLC については、

わずかに ISBN 有の一致率が高くなっているものの、ほとんど変わらない。ISBN の有無が各項目の一致率に影響を与えているわけではないようであった。

4.5.2 DN/MARC レコードと NC レコードの不一致内容の分類

(1) TR の不一致内容

3.4.(1)のとおり、項目ごとにランダムに抽出した不一致レコード 500 件について、目視により不一致内容による分類を行い(付録 5)、その結果から不一致内容の割合を導き出した。

TR の不一致内容の割合は図 4-32 のとおりである。

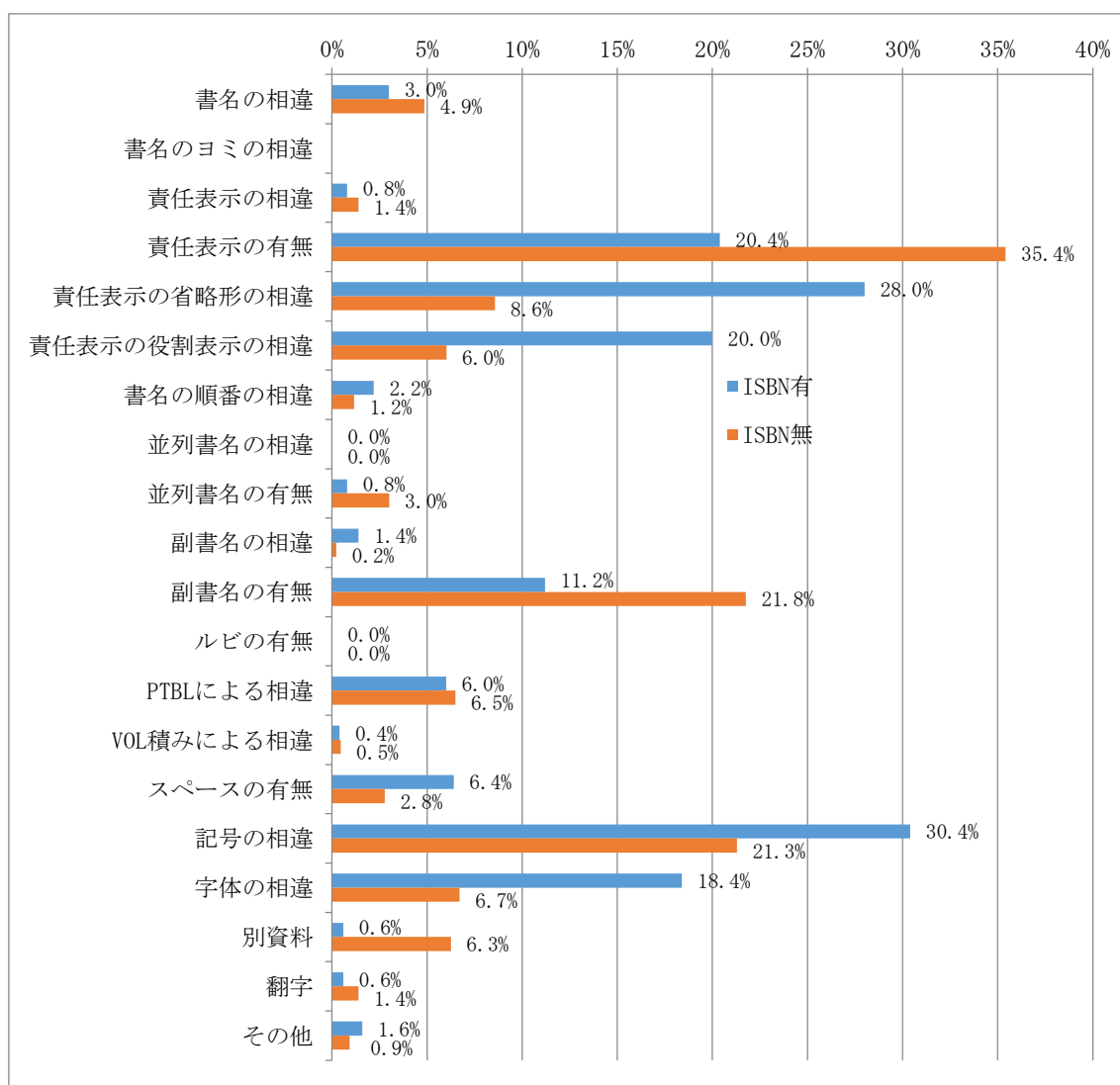


図 4-32 TR の不一致内容の割合

TR で最も不一致の割合が大きい内容は ISBN 有で「記号の相違」および「責任表示の省略形の相違」、ISBN 無では「責任表示の有無」である。

ISBN 有の不一致については、NC と DN のルールの違いが影響していると考えられる。

例えば、記号の相違については、同一の役割の責任表示が複数ある場合、NCでは“著者名 1,△著者名 2”のようにカンマで区切るが、DNでは、“著者名 1△;△著者名 2”のようにセミコロンが使われている書誌が多く確認された。また、責任表示の省略形については、NCで“herausgegeben”がDNで“Hrsg.”、また、NCで“edited”がDNで“ed.”と表記されているような相違が多く確認された。

ISBN 無の責任表示の有無については、USMARC、UKMARCと同様にNCレコード側に親書誌が多いことが原因であると考えられる。ただし、DNではRDAの適用は2015年からのため、本調査で対象としたDNMARCへのRDAによる影響はない。

(2) PBLC_1の不一致内容

3.4.(3)のとおり、項目ごとにランダムに抽出した不一致レコード500件について、目視により不一致内容による分類を行い(付録5)、その結果から不一致内容の割合を導き出した。

PBLC_1の不一致内容の割合は図4-33のとおりである。

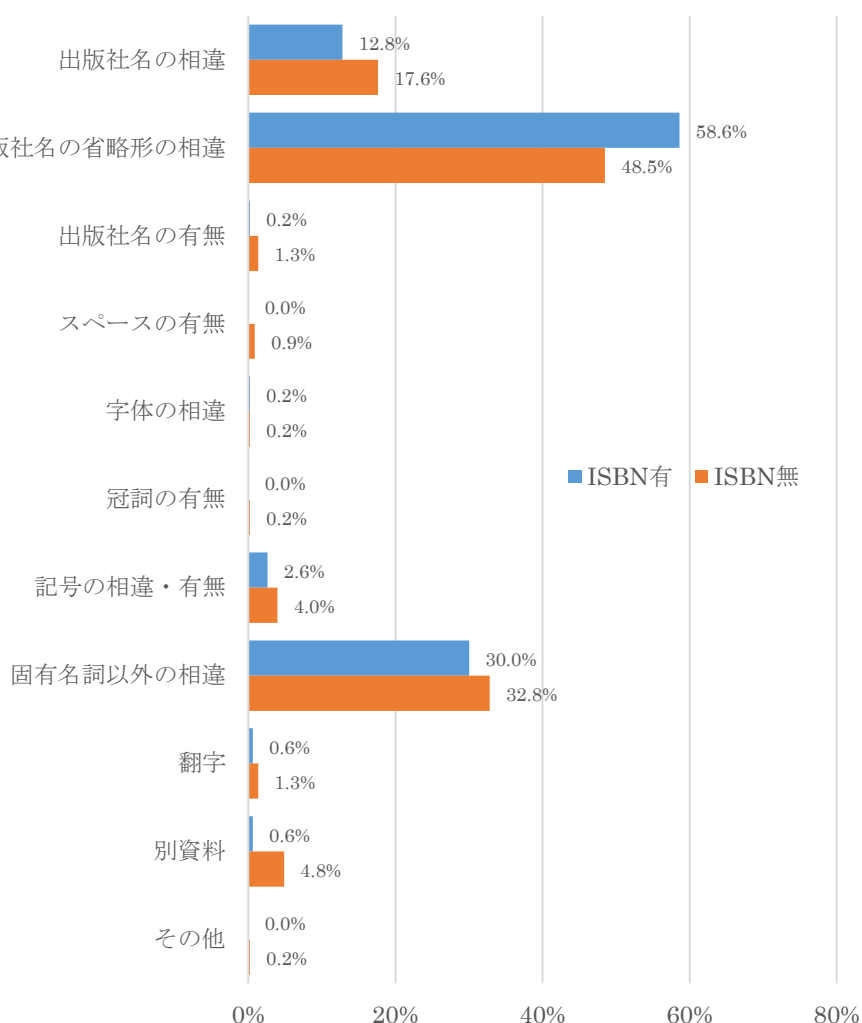


図 4-33 PBLC_1 の不一致内容の割合

PBLC_1 については、他 2 つの海外 MARC とは違った傾向が見られた。もっとも不一致の割合が大きい内容は、「出版社名の省略形の相違」である。一方で、「出版社名の相違」による不一致は少なく、また、「出版社名の有無」による不一致はほとんどなかった。

(3) PBLCDT_1 の不一致内容

3.4.(4)のとおり、項目ごとにランダムに抽出した不一致レコード 500 件について、目視により不一致内容による分類を行い(付録 5)、その結果から不一致内容の割合を導き出した。PBLCDT_1 の不一致内容の割合は図 4-34 のとおりである。

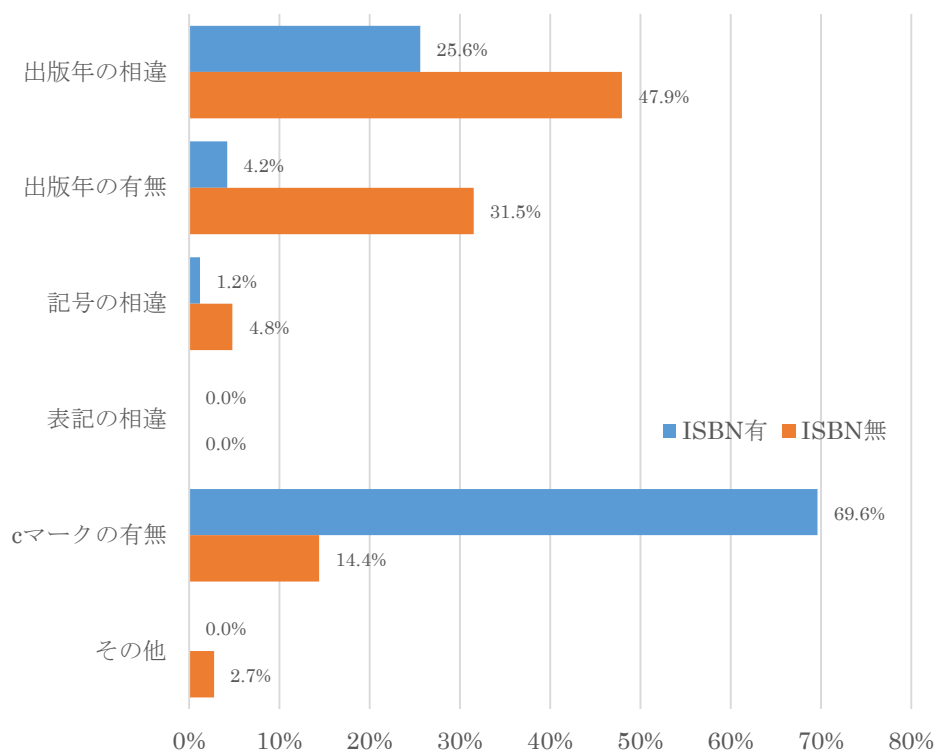


図 4-34 PBLCDT_1 の不一致内容の割合

PBLCDT_1 については、ISBN 有の「c マークの有無」による不一致の割合が最も大きい。これは、他 2 つの海外 MARC と共通した特徴である。また、ISBN 無の「出版年の有無」による不一致の割合が大きいのも海外 MARC の全体的な特徴である。

DNMARC の特徴としては、ISBN 無の「出版年の相違」による不一致の割合が大きいことがあげられる。ただし、これについては、そのうちの 2 割程度が別資料であることが確認できている。(1) の TR の不一致内容でも、ISBN 無については、別資料の割合が他の海外 MARC と比較して高かった。これらのことから、DNMARC は、形式のみを参照するために流用されることが他の海外 MARC に比べて多いと言える。

(4) PHYNM の不一致内容

3.4.(5)のとおり、項目ごとにランダムに抽出した不一致レコード 500 件について、目視により不一致内容による分類を行い(付録 5)、その結果から不一致内容の割合を導き出した。PHYNM の不一致内容の割合は図 4-35 のとおりである。

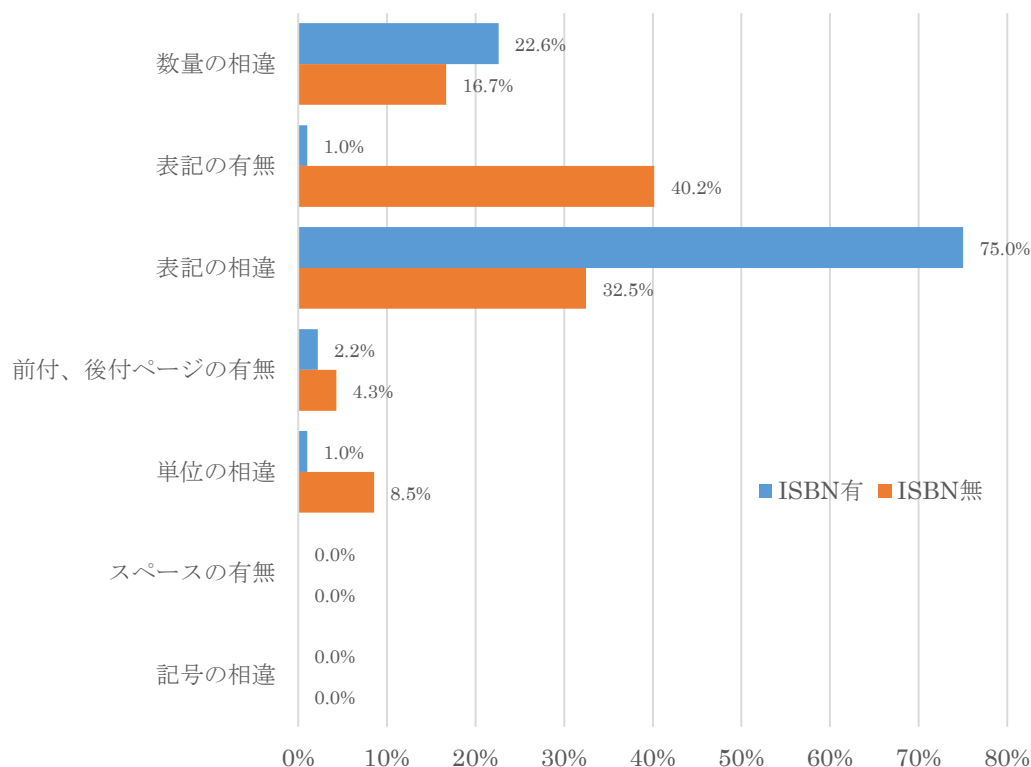


図 4-35 PHYNM の不一致内容の割合

PHYNM の ISBN 有でもっとも不一致の割合が大きい内容は、他 2 つの海外 MARC と同様に「表記の相違」である。ただし DN MARC では、「p.」が「S」と表記されている。ISBN 無でこの不一致が少ないのは、親書誌が多い(v.で表される書誌が多い)ことによる。

DN MARC の特徴としては、「数量の相違」による不一致の割合が大きいことがあげられるが、このうち ISBN 無については、4 割弱が別資料であることが確認できている。また、ISBN 有については、DN MARC の元となるドイツ国立図書館の OPAC で不一致データを再度確認したところ、修正されているものも多く見られた。

(5) PHYSZ の不一致内容

3.4.(6)のとおり、項目ごとにランダムに抽出した不一致レコード 500 件について、目視により不一致内容による分類を行い(付録 5)、その結果から不一致内容の割合を導き出した。PHYSZ の不一致内容の割合は図 4-36 のとおりである。

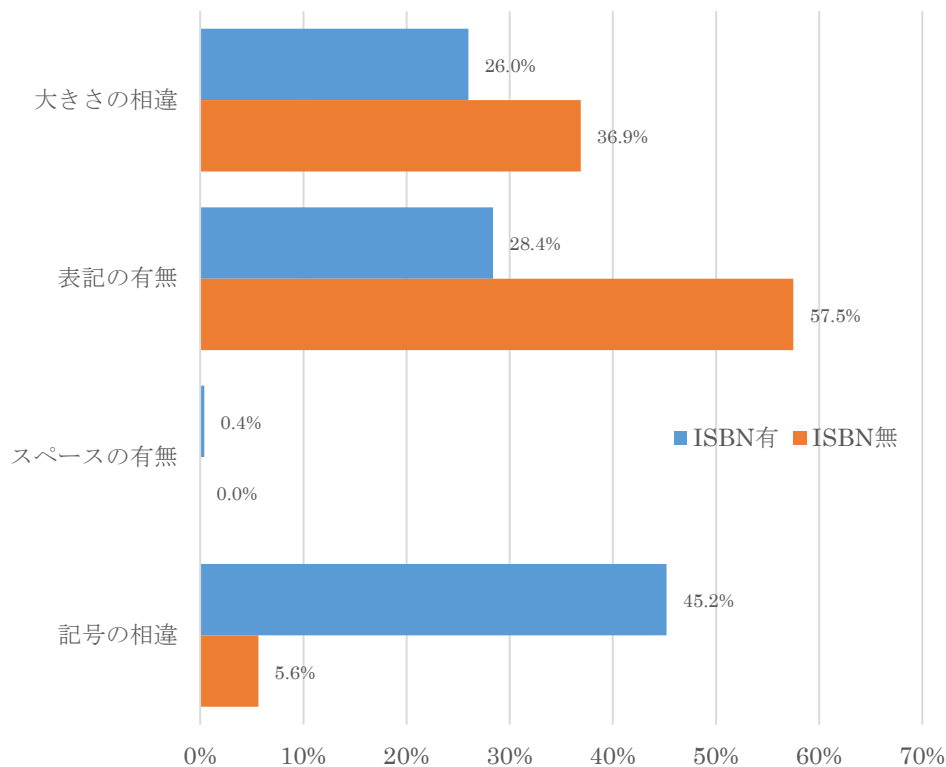


図 4-36 PHYSZ の不一致内容の割合

PHYSZ でもっとも不一致の割合が大きい内容は、ISBN 無の「表記の有無」である。これは、他 2 つの海外 MARC と同様の傾向である。

DNMARC の特徴としては、ISBN 有の「記号の相違」による不一致の割合が大きいことがあげられる。この不一致は主に「cm」と「mm」表記による違いであり、大きさの相違にも影響を与えている。ただし、mm で表記されたレコードについて、DNMARC の元となるドイツ国立図書館の OPAC で検索したところ、ほとんどが NC レコードと同じ大きさの cm 表記になっており、mm 表記のものは確認できなかった。おそらく、どこかの時点で cm に変更されたものと考えられる。また、2015 年 11 月の CAT-P 変換前の DN のデータからは、mm 表記のレコードは確認できなかった。そのため、mm 表記による相違については、現在では考慮する必要がなくなっており、記号の相違および大きさの相違の不一致割合も減少していると予想される。

4.5.3 DNMARC レコードの要対応の不一致 (ISBN 有の場合)

(1) 4.5.2 で分類した結果のうち、ISBN 有の場合について、ISBN が一致するレコードを対象に、3.4 で要対応と判断した不一致の件数を調査した。この調査では、PBLC_1, PBLCDT_1, PHYNM については、階層構造を持ち、親書誌または PTBL リンクを持つ子書誌のいずれかであるものは除外した。結果は表 4-14 と図 4-37 のとおりである。

表 4-14 DNIMARC (ISBN 有) の要対応の不一致の件数

項目	不一致内容	件数	項目	不一致内容	件数
TR	書名の相違	15	PBLCDT_1	出版年の相違	125
	責任表示の相違	4		出版年の有無	20
	並列書名の相違	0		小計	145(29.0%)
	副書名の相違	7	PHYNM	数量の相違	100
	小計 ⁵	24(4.8%)		表記の有無	5
PBLC_1	出版社名の相違	63	PHYSZ	小計	105(21.0%)
	出版社名の有無	1		大きさの相違	128
	小計	64(12.8%)	小計	128(25.6%)	

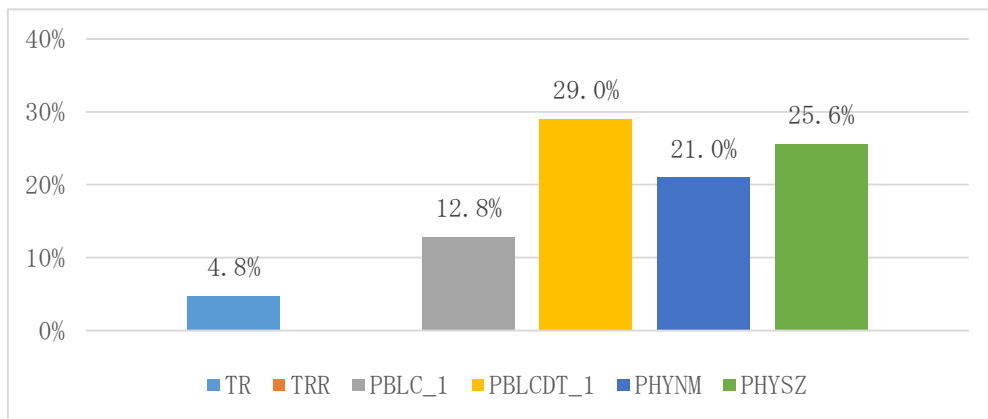


図 4-37 DNIMARC (ISBN 有) の要対応の不一致の割合

(2) DNIMARC レコードの TR の修正が必要となる割合 (ISBN 有の場合)

(1) の結果から DNIMARC の ISBN 有の場合、TR の修正が必要となる割合は、3.5 の計算式により、下記のとおりとなる。

修正が必要となる割合 = 3.43% → 29 冊に 1 冊

⁵ 1 レコードに対して複数の分類項目を付与しているため、不一致内容の総数と異なる

5. 調査を受けての新モデル案

この章では、前章までの調査を受けて、2章で述べた新モデル案が実際にどのように実現できるか考察した。主となる3点（外部データを活用した書誌作成のモデル、所蔵と書誌の対応をどの間に保持するか、書誌と所蔵をどのように分離して持つか）について述べる。他に検討した内容については別紙(b)にまとめた。

5.1 外部データを活用した書誌作成のモデル

当初の検討（2章参照）では、仮に外部 MARC データに一切の修正が必要とならないのであれば、①外部 MARC 等に既存の書誌ある場合には書誌作成を行わずにそのまま用い、②既存のデータがない資料については新規で書誌作成を行う、という2パターンの分岐を想定していた。図で示すと図 5-1 のようになり、入力は外部データ及び NACSIS-CAT の既存書誌ならびに新規で作成する書誌のみとし、これらを名寄せすることで必要な書誌を全て賄う想定であった。

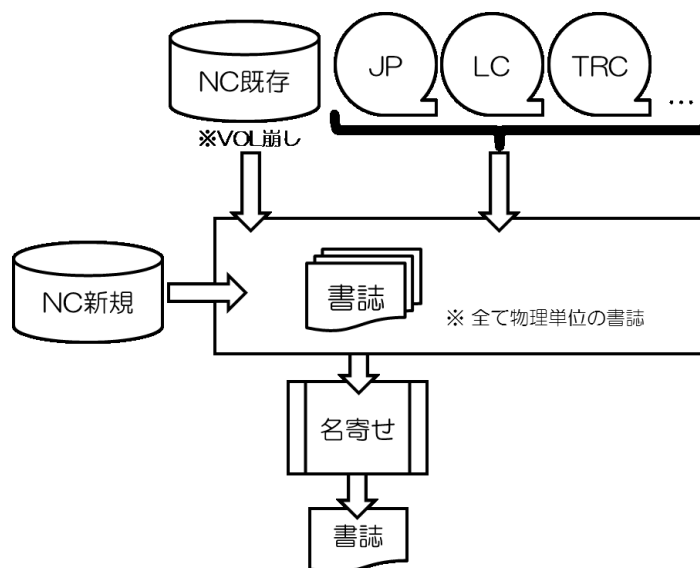


図 5-1 これまでの検討（図 2-2 から外部データ活用に関する部分を抜き出したもの）

しかし、今回の調査の中で、外部 MARC 等に既存の書誌がある場合でも、数%修正が必要となる見込みが得られたことから、ここでは書誌作成のモデルについてさらに検討する。単純には「書誌データを修正する仕組みが必要である」ということだが、どの段階のデータに対して修正するかによりレコード調整の有無などが変わってくるため、各方式を検討しより良い方法を採用する必要がある。

データ修正の方式は表 5-1 の通り、直観的には3つの案が考えられる。

表 5-1 データ修正の方式 3 案

#	修正の方式	メリット	デメリット
1	名寄せした後のレコードを修正し、上書きする方法	修正による効果が最大 管理対象レコード数が最小	レコード調整が発生する
2	名寄せ前の各レコードを修正し、上書きする方法	管理対象レコード数が最小	レコード調整が発生する
3	修正レコードは新規レコードとして扱う方法	レコード調整が起こらない	大量に新規レコードを作成できてしまう

これらの内、調査結果を踏まえると第 3 案がもっとも妥当であると考えられる。以下、各案の概要と、第 3 案が最も妥当であると判断した根拠を述べる。

(案 1) 名寄せ結果のレコードを修正し、上書きする方法

最も直観的な方法として、名寄せ結果として出てくる書誌を修正する方法が考えられる(図 5-2)。これはマスターとなる書誌が一つで、それを編集していくという意味で現行の NACSIS-CAT の仕組みに近い方法であるともいえる。

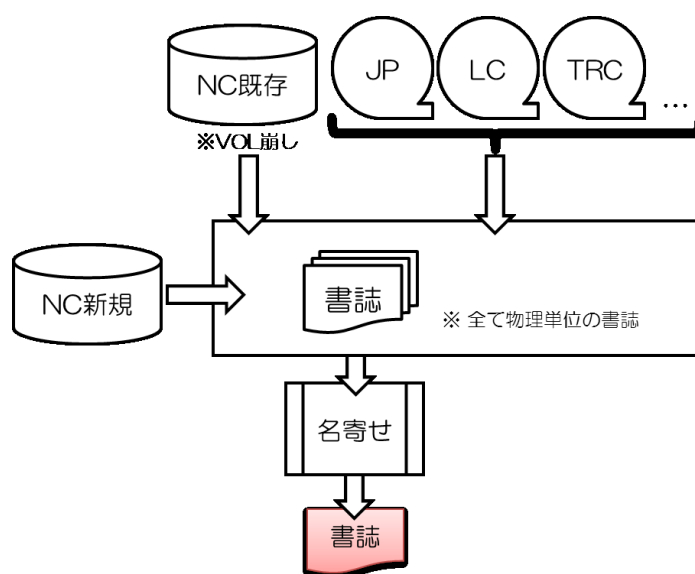


図 5-2 データ修正の方式案 1 の概要図 (網掛け部が修正対象データ)

この方法のメリットとしては、名寄せ済みの書誌を上書き修正することから、修正の結果が確実に名寄せ後の書誌に反映されるため、一度の書誌修正あたりの影響範囲が大きく、後述する 2 つの案より効率的であると言える。また、名寄せ後の書誌のみの編集管理のみをすればよいため、案 3 のようにレコード数が増加する心配もない。ただし、デメリットとして、マスターとなる書誌を修正することから、現行システムの大きな課題でもあるレコ

ード調整が引き続き発生してしまうことが想定される。また、修正対象のデータが名寄せ結果に依存することから、名寄せ結果をどのように保持し、どのように修正データを管理するかなど、複雑なルール定義が必要となる。

(案2) 名寄せ前の各レコードを修正し、上書きする方法

続いて、書誌データを修正する方法として、名寄せ前の各レコードを修正する方法が考えられる。(図5-3)。

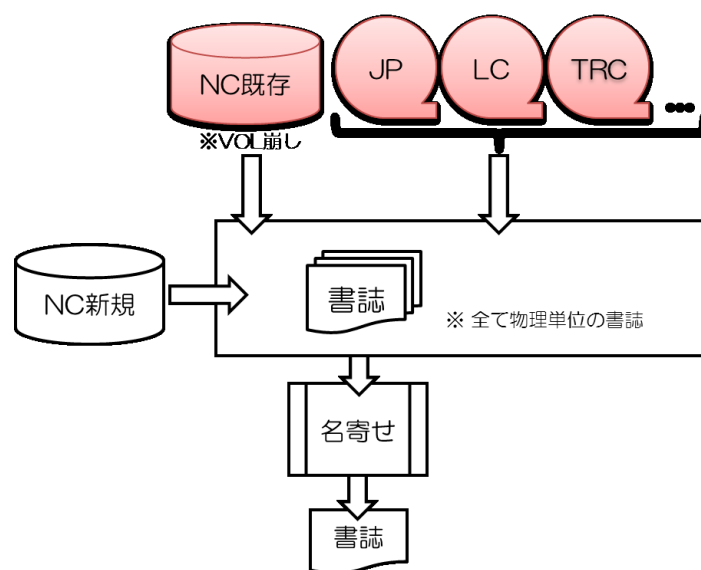


図5-3 データ修正の方式案2の概要図(網掛け部が修正対象データ)

案2では、前述の案1のデメリットのひとつである、「修正済みデータをどのように持つか」という課題を解消したものであるといえる。案1と同じく、修正を行って上書き保存することからレコード数の増加は発生しないメリットがある。しかしながら、あるレコードを修正したときに、他の機関も参照しているレコードを修正することになってしまうことから、案2においてもレコード調整が生じる。

(案3) 修正レコードは新規レコードとして扱う方法

3番目の案として、これまでの2つの案とは異なり、レコードを上書き修正することはせずに、書誌を修正したい場合は新規のレコードとしてデータを作成する方法が考えられる。イメージ図を図5-4に示す。

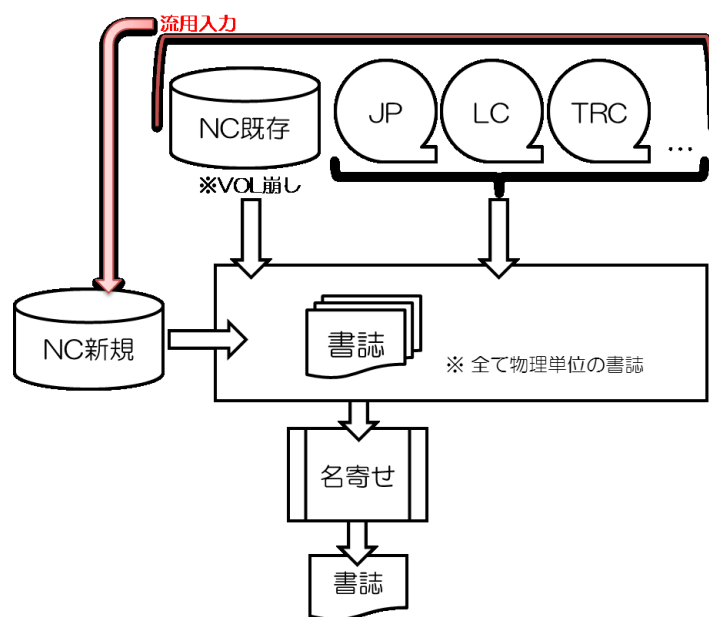


図 5-4 データ修正の方式案 3 の概要図(修正データは新規データとして扱う)

この方式の場合、書誌を修正した場合においても既存の記録には変更が生じず、新しい記録が作られるのみであることから、記録調整は発生しない。ただしこの方式を採った場合には、書誌の修正を行って新規記録が大量に作られてしまうことが懸念される。書誌の修正が多く発生しない場合は案 3 を採用すると良いが、書誌の修正が多く見込まれる場合には案 3 は採用できない。なお、この方式を用いる場合には、修正して作られた新規記録が修正前の記録に名寄せされてしまうと修正した意味がなくなってしまうため、特定の記録とは同定されないようにする上位ルールを適用するといった仕組みが必要となるだろう。

今回調査した、「外部データに修正を加えるとしたら、何%なのか」の調査結果によって、0.87%~6.02%ほどが修正要となるデータの割合であることが分かった。今回の調査結果は期間の限定されたデータではあるが、今後も大きな変動はないものと考えられる。したがって、修正が必要なデータに対して新規記録を作成したとしても記録が無尽蔵に増えてしまうことは考えにくい。本調査における結果から、新モデル案では案 3 (修正記録は新規データとして扱う) を採用することで、記録調整を発生させずに、外部データを活用した書誌作成モデルが実現できることが分かった。

5.2 所蔵登録を行う対象

書誌情報のほかに、所蔵情報をどのように管理するかも重要な課題である。2章で述べた通り、新モデル案では書誌情報を新システム側に、所蔵情報をローカル側に分けて持ち運用をシンプルにしたい。しかしながら総合目録としての機能を持つためには、所蔵情報と

紐付けるための情報を新システム側で保持する必要がある。書誌情報と所蔵情報をどのように紐付けて持つべきかについて本節で述べる。

所蔵情報を紐付ける方法として、名寄せ前の書誌に所蔵を紐付ける方法と名寄せ後の書誌に所蔵情報を紐付ける方法が想定される（図 5-5）。

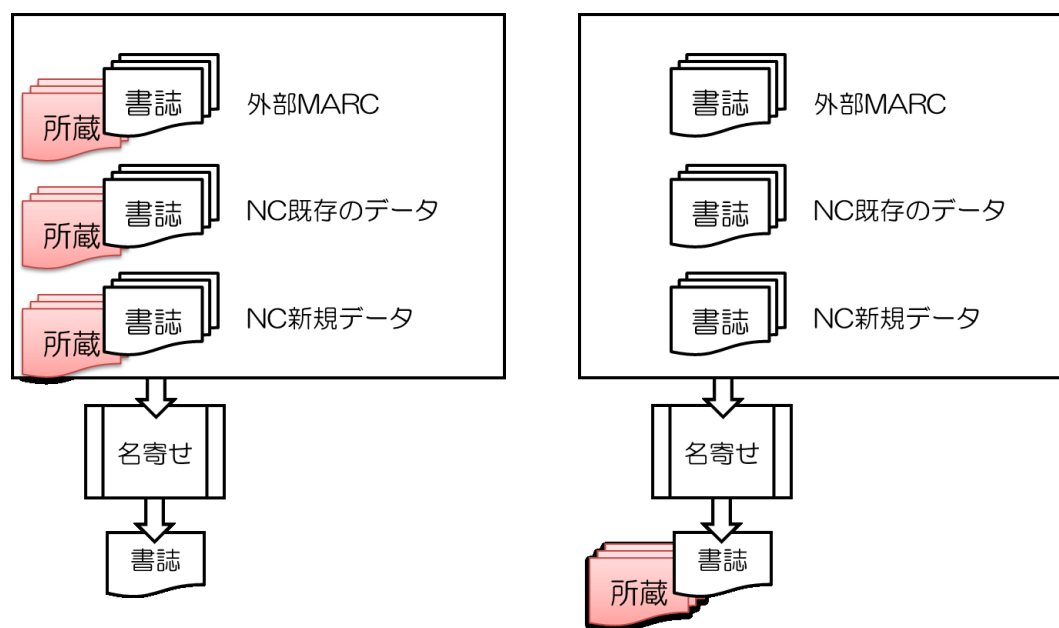


図 5-5 書誌と所蔵の紐付け方式のイメージ図。

(左:名寄せ前の書誌に所蔵を付ける場合、右:名寄せ後の書誌に所蔵を付ける場合)

(案 1) 所蔵情報を名寄せ前の書誌に紐付ける方法

名寄せ前の書誌に所蔵を紐付ける場合には、名寄せ結果に依存せずに所蔵を付けたい書誌に対して所蔵情報（へのリンク）を付与することができる（ユーザに提示する段階では所蔵をまとめて一元的に表示する）。一方で、名寄せ前の書誌から対象となるデータを検索するための仕組みが必要となる。外部 MARC 等に所蔵を付ける場合は、対象となる MARC 番号などから一意に結果を特定可能なことから検索性に問題は生じないが、新規データを検索対象とする場合には検索性の低下が懸念される。したがって、もし、似たような新規データが乱立してしまうと、その中から所蔵を付ける対象がどれかを選ぶことに大きな時間的コストが発生してしまうこととなる。

(案 2) 所蔵情報を名寄せ後の書誌に紐付ける方法

名寄せ後の書誌に所蔵を紐付ける場合には、案 1 の問題点として挙げていた検索性の問題は発生しにくいと考えられる。名寄せ済みの書誌を検索するため、大量のレコードの中から対象レコードを探し出す、といった作業は軽減される。しかしながら案 2 の場合には、

所蔵を付ける対象が名寄せ結果に依存する点や、名寄せが解消された場合に元の名寄せ済み書誌についていた所蔵情報（へのリンク）をどのように扱えば良いか、という課題が残る。

今回の調査結果及び前節の外部データを活用した書誌作成のモデルでの検討結果から、新モデル案ではレコードの修正に伴って新規レコードがどうしても発生してしまう。したがってそのまま案1を用いることはできない。そこで、案1の検索性の問題を解消するため、案2のように名寄せ済みの書誌を検索したうえで、自動的に名寄せ前の書誌に所蔵を付け替え管理する方法が考えられる。すなわち、所蔵を付ける対象として、(1) MARC 番号のように一意に名寄せ前のレコードを特定できるような情報がある場合には、当該書誌に所蔵の紐付けを行い、(2) タイトル等から検索する場合には名寄せ済み書誌から検索して、名寄せ済み書誌に所蔵を付けると自動的に名寄せ前の適当なレコードに所蔵を紐付けるような仕組みを用意すればよい。したがって新モデル案では、上記の仕組みを備えた案1を採用する。

5.3 書誌と所蔵の分離

新モデル案では原則として、書誌情報は新システム側に、所蔵情報はローカル側にのみ持つこととするが、5.1、5.2でのモデルを実現するためには所蔵のリンクはシステム側にも一部持つ必要がある。既存の NACSIS-CAT/ILL や CiNii Books で実現している機能を維持するために、新システム側/ローカル側の両方で以下の情報を保持する必要がある。

- ・ 書誌レコードの ID (MARC 番号など。現行システムでの BID に相当)
- ・ 図書館 ID および機関 ID
- ・ 各参加館での資料を一意に特定するための ID

上記以外の所蔵情報（例えば排架位置や請求記号）は新システム側で保持せず、ローカル側のみで保持する。ローカル側から新システム側への書誌情報の参照は、書誌レコードの ID から書誌を呼び出して利用することができ、新システム側からローカル側への参照は、各参加館での資料を一意に特定するための ID からローカル側に問い合わせることで所蔵情報を参照し、利用・表示する。上記にない排架位置等は CiNii Books では検索には用いておらず、表示を行っているのみであるため、上記の項目さえあれば検索インデックスは構築可能である。

また、5.1で検討した通り、新規レコードを既存レコードに名寄せさせないためのルール情報を付与する必要が生じるため、新システム側には書誌情報に加え、書誌同士の関連を示すためのリンク情報が必要となる。これは編集可能である必要があるため、書誌項目に加えて新規でデータを追加する必要がある。この項目を用いることで、機械名寄せのためだ

けでなく、書誌間や著作間での関連を示すことが出来る。したがって新モデル案では外部 MARC の記述事項自体は必要ない限り変更しないが、異なる情報源に記載されているメタデータの関連性を充実させていくことができ、低いコストでより充実したメタデータを作成することも可能となる。

6. まとめ

1 班は NACSIS-CAT の再考に向けて、「外部データの活用を含む所蔵登録の簡素化」というテーマのもと、NACSIS-CAT レコードおよび参照 MARC の差異に関する調査を行い、調査結果にもとづき、既存の NACSIS-CAT の枠組みを発展させ、外部データをより効率的に活用した書誌作成・所蔵登録管理の新モデル案を提案した。

6.1 NACSIS-CAT レコードおよび参照 MARC の差異に関する調査

この調査では、以下のことが明らかになった。

(1) ISBN 有無別の一致率

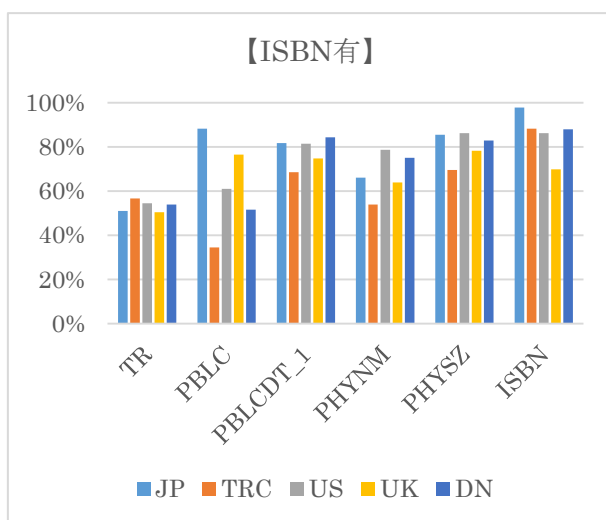


図 6-1 各項目の一致率(ISBN 有)

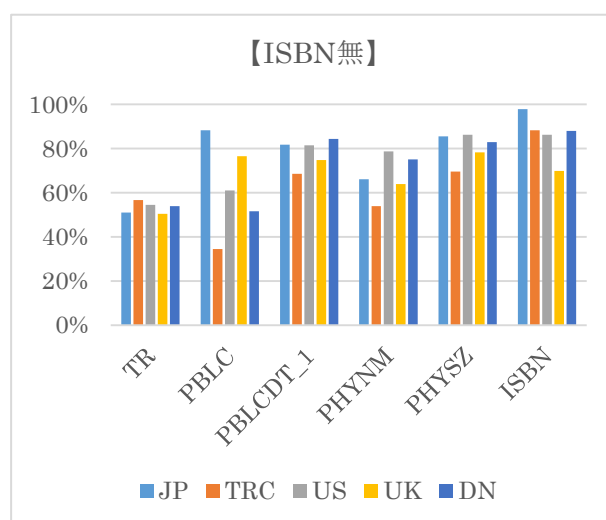


図 6-2 各項目の一致率(ISBN 無)

どの MARC においても ISBN の有無によって、一致率に変化は見られなかった。また、TR や PBLC_1 などの調査項目についても、割合の大小はあるにせよ、一定数以上の不一致が見受けられる。

1 班は当初、「ISBN をキーにすれば、外部 MARC をほぼ無修正で利用できるのではないかと予測していた。しかしながら、ISBN だけでは十分条件にはならないことが分かった。そこで、不一致となったデータを確認しながら、要対応の（人手による修正が必要な）不一致を算出した。

(2) 不一致項目の分類結果

(1)の結果と同様に、ISBN の有無によらず、どの項目にも要対応のデータが存在した。そこで、最初の予測「ISBN をキーにすれば、外部 MARC をほぼ無修正で利用できるのでは

はないか」を、「ISBN 有るデータには、要対応の不一致数はどのくらいあるか？」に変更して、検討を続けた。

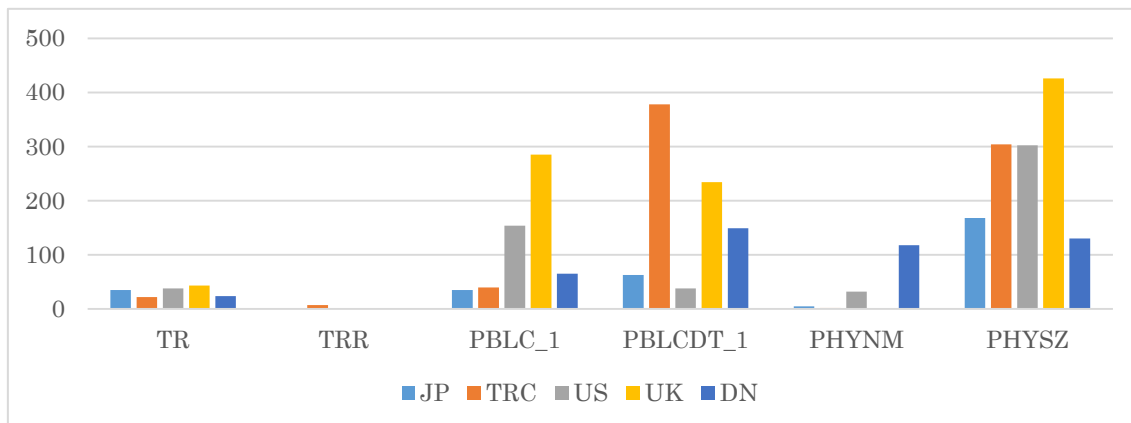


図 6-3 ISBN 有の各項目の無視できない不一致数

(3) 無視できない不一致内容およびその割合

さらに、書誌検索および同定の際に重要となる TR に焦点をしばり、TR が誤りであり、かつその誤りが無視できないため要修正である割合を算出した。その結果、ISBN 有の場合、修正が必要な割合は、**0.87%~6.02%**であることが明らかになった。

6.2 新モデル案の検討

6.2.1 外部データを活用した書誌作成モデル

6.1 の調査結果をうけて、外部データを活用する新モデル案が、どのように実現可能かを考察した。検討の結果、外部データを無修正で利用できるものは利用し、一定数は存在すると予想される修正が必要なデータについては、上書きして利用するのではなく、新規レコードとして扱うのが妥当と判断した。今回の調査で明らかになった **0.87%~6.02%** の範囲であれば、新規レコードが大量に作成されることは考えにくい。

なお、この方式を用いる場合は、修正して新規作成されたレコードが修正前のレコード(外部データ)に名寄せされない仕組み、書誌同士の関連を示すための情報が必要である。

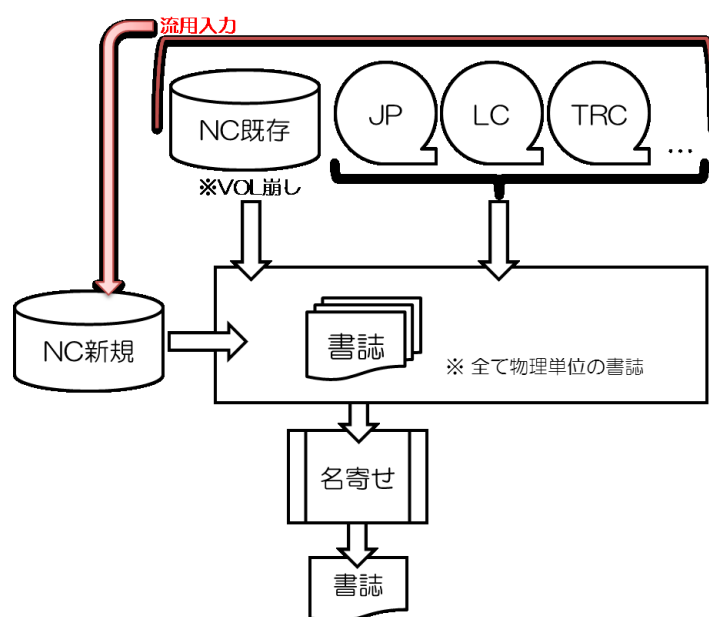


図 6-4 新モデル案 (図 5-4 と同一)

6.2.2 所蔵登録について

新モデル案では原則として、書誌情報を新システム側に、所蔵情報はローカル側に持たせるとする。しかしながら、既存の NACSIS-CAT/ILL や CiNiiBooks で実現している機能を維持するために、新システム側で以下の情報を最低限、持つ必要があると考えられる。

- ・ 書誌レコードの ID (MARC 番号など。現行システムでの BID に相当)
- ・ 図書館 ID および機関 ID
- ・ 各参加館での資料を一意に特定するための ID

6.2.1 で提案した新モデル案では、既存レコードと新規レコードが混在するため、所蔵登録の際、書誌同定に時間的コストが発生するおそれがある。これを解消するために、検索

と書誌同定は名寄せ済みの書誌を，所蔵登録は名寄せ前の書誌に対して行う方法が考えられる。つまり，名寄せ後の書誌を検索したうえで，名寄せ前の書誌に所蔵をつける。そのときに自動的に適切なレコードにつけると，もしどの書誌につけるかを選択したい場合にそれが可能になる仕組みを用意すればよいと考えられる。

以上のとおり，1班では NACSIS-CAT レコードおよび参照 MARC の差異に関する調査結果にもとづき，既存の NACSIS-CAT の枠組みを発展させ，外部データをより効率的に活用した書誌作成・所蔵登録管理の新モデル案として，「修正レコードは新規データとして扱い，外部データと新規データを名寄せにより利用する書誌作成モデル」がもっとも実現可能性が高いという結論にいたった。

6.3 今後の課題

今回の外部データとして，国内外の主要 MARC のみを対象に調査を行った。今後，他の MARC や MARC 以外の外部データについても同様に調査することにより，新モデル案のデータとして活用することも可能である。

参考文献

1. The Library of Congress. “On the Record: Report of the Library of Congress Working Group on the Future of Bibliographic Control”. 2008
<http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf>, (参照 2015-12-25)

邦訳 : 国立国会図書館訳. ” On the Record : 書誌コントロールの将来に関する米国議会図書館ワーキンググループ報告書”.
http://dl.ndl.go.jp/view/download/digidepo_1001859_po_ontherecord_jp.pdf?contentNo=1&alternativeNo, (参照 2015-12-25)
2. 国立情報学研究所 学術コンテンツ運営・連携本部 図書館連携作業部会 (次世代目録ワーキンググループ) . “次世代目録所在情報サービスの在り方について (最終報告) ”. 2009. http://www.nii.ac.jp/CAT-ILL/archive/pdf/next_cat_last_report.pdf, (参照 2015-12-25)
3. 国立情報学研究所. “NACSIS-CAT/ILL 参加館状況調査アンケート結果報告書 (平成23年3月調査)”. 2012
https://www.nii.ac.jp/CAT-ILL/about/project/pdf/enq2011_1_0315.pdf, (参照 2015-12-25)
4. 大向一輝. “特集, データベース構築の今: CiNii Articlesのシステムデザインとデータモデル”. 情報の科学と技術. 2012, vol.62, no.11, p.473-477
5. 第11回これからの学術情報システム構築検討委員会 配布資料 資料No.3-1. “NACSIS-CAT検討作業部会の設置について (案) ”.
http://www.nii.ac.jp/content/korekara/about/document/pdf/korekara_doc20150313.pdf (参照2015-12-25)