

成果報告会発表[1班]

国立国会図書館 安藤大輝
神戸大学附属図書館 末田真樹子
九州大学附属図書館 萩原泰子
秋田大学附属図書館 原智子

目次

- はじめに
- 外部データ比較調査
- 不一致項目比較調査の結果分析
- 外部データに必要な要件
- 新モデルについて

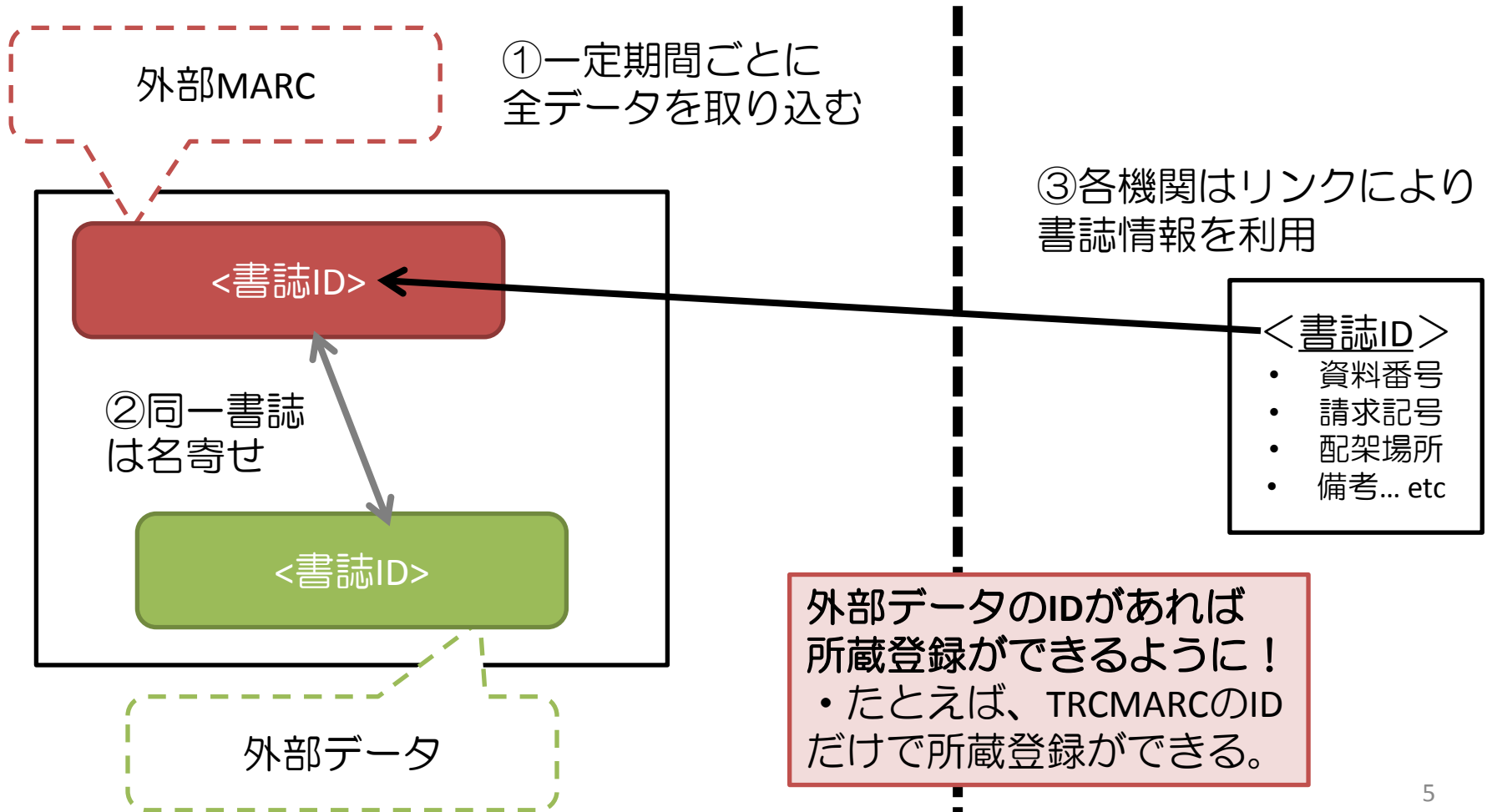
はじめに

テーマ

- テーマ「外部データの活用を含めた所蔵登録の簡素化」
 - 現在は、所蔵登録はすでにある書誌に対して所蔵をつけるのみ
 - 現行の所蔵登録には、簡素化の余地は少ないのでは？
 - 実際には書誌がないときに書誌を作るのが大変
 - 外部データを活用して、現行の所蔵登録を簡素化するモデルが実現できないか？

提案

外部データ自体に所蔵をつけられないだろうか



課題とWSのスコープ

主な調査対象

- 多種の外部データが混在できるか
 - システム処理で対応できる部分
 - 書誌の質が低下しても許容できる部分
 - どこか人が手を加えなければいけない部分

NCにとって一番身近である外部MARCを調べ、
新モデルにおいて外部データを用いる際の

- 外部データに必要とされる要件（データ側）
 - 運用ルールの変更内容（NC側）
- を明らかにすることが最終目標

外部データ比較調査

外部データ比較調査 目的と対象

目的

・ 外部MARCとNC書誌データの差異について、
以下を明らかにする

- ① システム処理で対応できる部分
- ② 書誌の質が低下しても許容できる部分
- ③ どこか人が手を加えなければいけない部分

対象 流用入力件数の多い各MARC(TRC/JP/US/UK/DN)

- ・ 外部MARCから流用登録されたレコードについて
 - ① ロード前の外部MARC
 - ② CATPにロードした参照レコード
 - ③ 流用入力して作られたNC書誌データ
- を比較して、どの段階で相違が生じているか、なぜその項目を編集したかを確認し、差異を分類する。

外部データ比較調査

比較調査項目

- TR タイトル
 - TRR タイトルよみ（JPとTRCのみ）
 - PBLC 出版者
 - PBLCDT_1~3 出版年
 - PHYNM ページ数など
 - PHYSZ 大きさ
-

各MARCのマッピング表、変換ルールも
並行して確認しながら行う

外部データ比較調査 データ抽出条件(1/2)

NCレコードの抽出条件は、以下のとおり

- CRTDT=20100101-20141231
- SOUCE=JP/TRC/US/UK/DN

さらにISBN有無を分ける

- 無：ISBN_1～10のいずれも空白、またはNull
- 有：上記以外

外部データ比較調査 データ抽出条件(2/2)

先の条件で抽出したレコードについて、参照元との一致率、および不一致項目の分類を行う

- 一致率の調査
 - 対象：全レコード
 - NレコードのMARCID=参照MARCのID
- 不一致項目の分類
 - 対象：ISBN有無について、それぞれランダム500件
 - 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）

外部データ比較調査 データの加工

- 1レコード(=1NCID)を1行
- 出力項目
 - 参照MARC：
ID,MARCID,SOUCE,ISBN,ISBNA,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
- 繰り返しのある項目
 - PLC、PBLC、PBLCDT：先頭から順に「PLC1、PLC2、PLC3、PLC4」という項目に出力
 - 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
- 存在しない項目と値が無い項目の書き分け
 - 存在しない項目=空欄
 - 値がない項目=NULL

外部データ比較調査 調査ツール

- paiデータをtsvに変換
 - 調査のためにjavaプログラムを作成し、データ抽出・変換
- 各データの集計
 - Microsoft Excel
 - Microsoft Access

外部データ比較調査 不一致項目の分類 (1/4)

- 調査対象

- TR, TRR, PBLC, PBLCDT, PHYNM, PHYSZ

- ① 目視による分類 (NC書誌⇔参照レコード)

- 不一致内容を目視によって確認し、以下のとおりパターン化する (次頁)

- 機械処理可 (システム処理で対応できるもの)
 - 無視できるもの
 - 要対応 (人手での処理が必要)

- ② サンプルチェック (参照レコード⇔外部MARC)

- ロード前 (外部MARC) とロード後 (参照MARC) のデータを比較する。

調査方法

不一致項目の分類 (2/4)

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	無視できる
・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	無視できる
・責任表示の有無	無視できる	・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)	無視できる	・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	無視できる	・記号の相違	機械処理可
・書名の順番の相違	無視できる	・大文字小文字などの字体の相違	無視できる
・並列書名の相違	要対応	・別資料であるもの	無視できる
・並列書名の有無	無視できる	・翻字	機械処理可
・副書名の相違	要対応	・その他	?

調査方法

不一致項目の分類 (3/4)

不一致項目の種類 (PBLC)	対応の要不要	不一致項目の種類 (PBLCDT)	対応の要不要
・ 出版者名の相違	要対応	・ 出版年の相違	要対応
・ 出版者名の省略形の相違	無視できる	・ 出版年の有無	要対応
・ 出版者名の有無	要対応	・ 記号の相違	表現の相違
・ スペースの有無	無視できる	・ 標記の相違	機械処理可
・ 大文字小文字などの字体の相違	無視できる	・ cマークの有無	機械処理可
・ 冠詞の有無			
・ 記号の相違・有無	無視できる		
・ 固有名詞以外の相違	無視できる		
・ 翻字	無視できる		
・ 別資料	無視できる		
その他	?		

調査方法

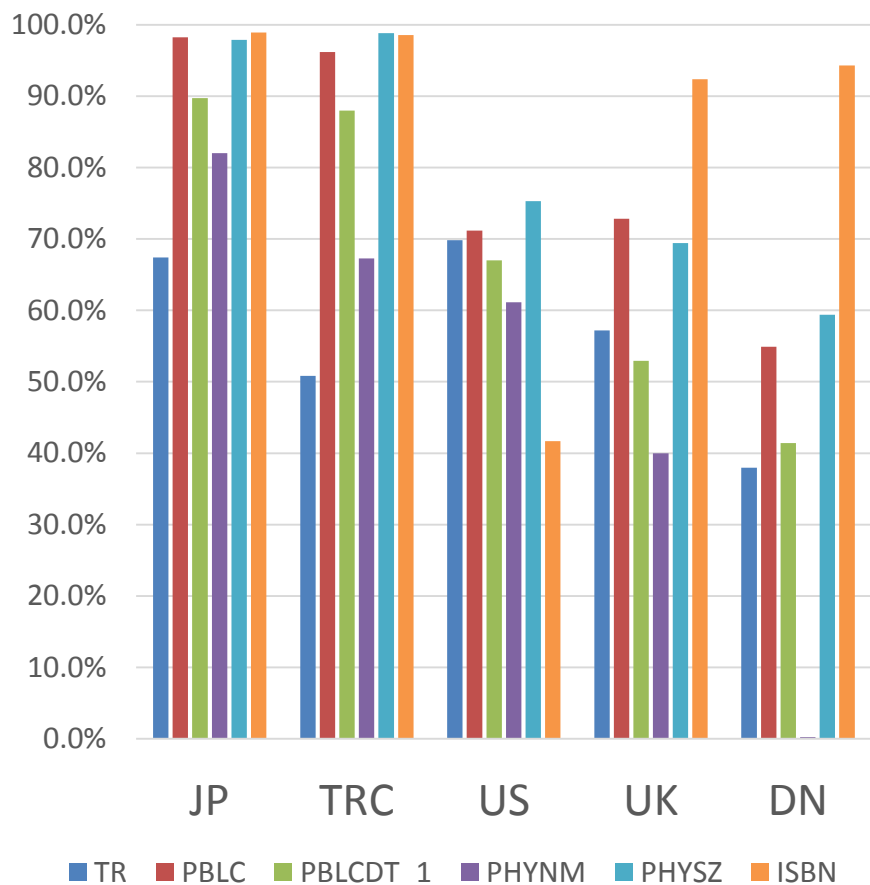
不一致項目の分類 (4/4)

不一致項目の種類 (PHYNM)	対応の要不要	不一致項目の種類 (PHYSZ)	対応の要不要
・数量の相違	要対応	・大きさの相違	要対応
・表記の有無	要対応	・表記の有無	無視できる
・表記の相違	無視できる	・スペースの有無	機械処理可
・前付、後付ページの有無	無視できる	・記号の相違	機械処理可
・単位の相違	無視できる		
・スペースの有無	無視できる		
・記号の相違	無視できる		
・その他	?		

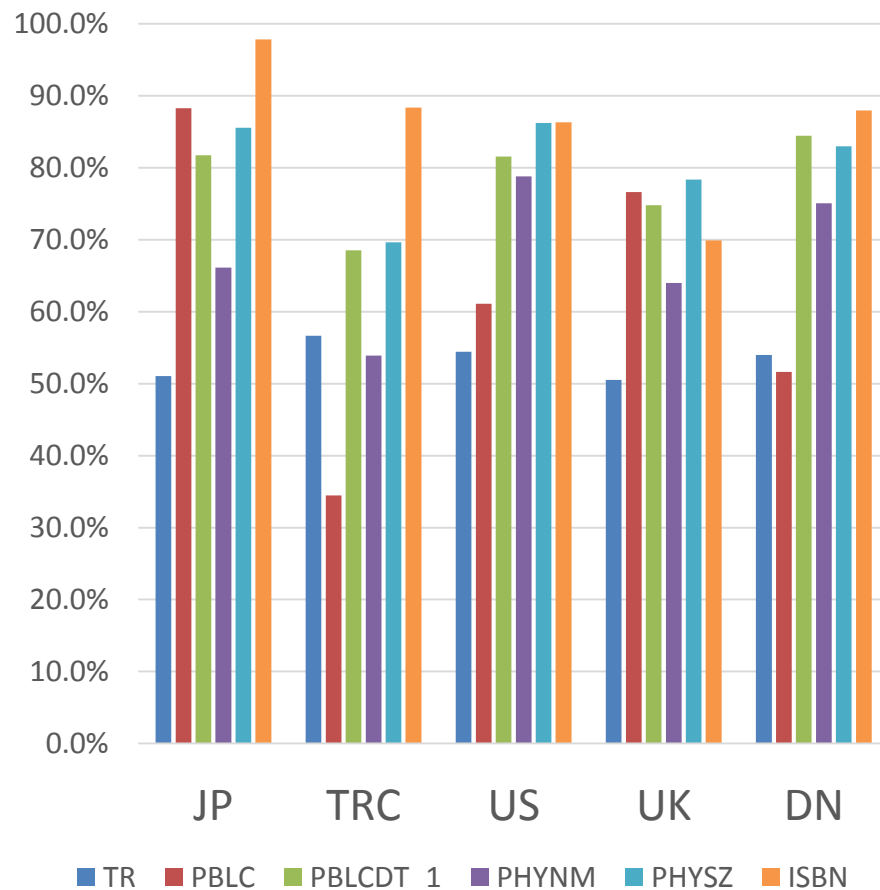
不一致項目比較調査の結果分析

参照元との一致率（2010-2014年）

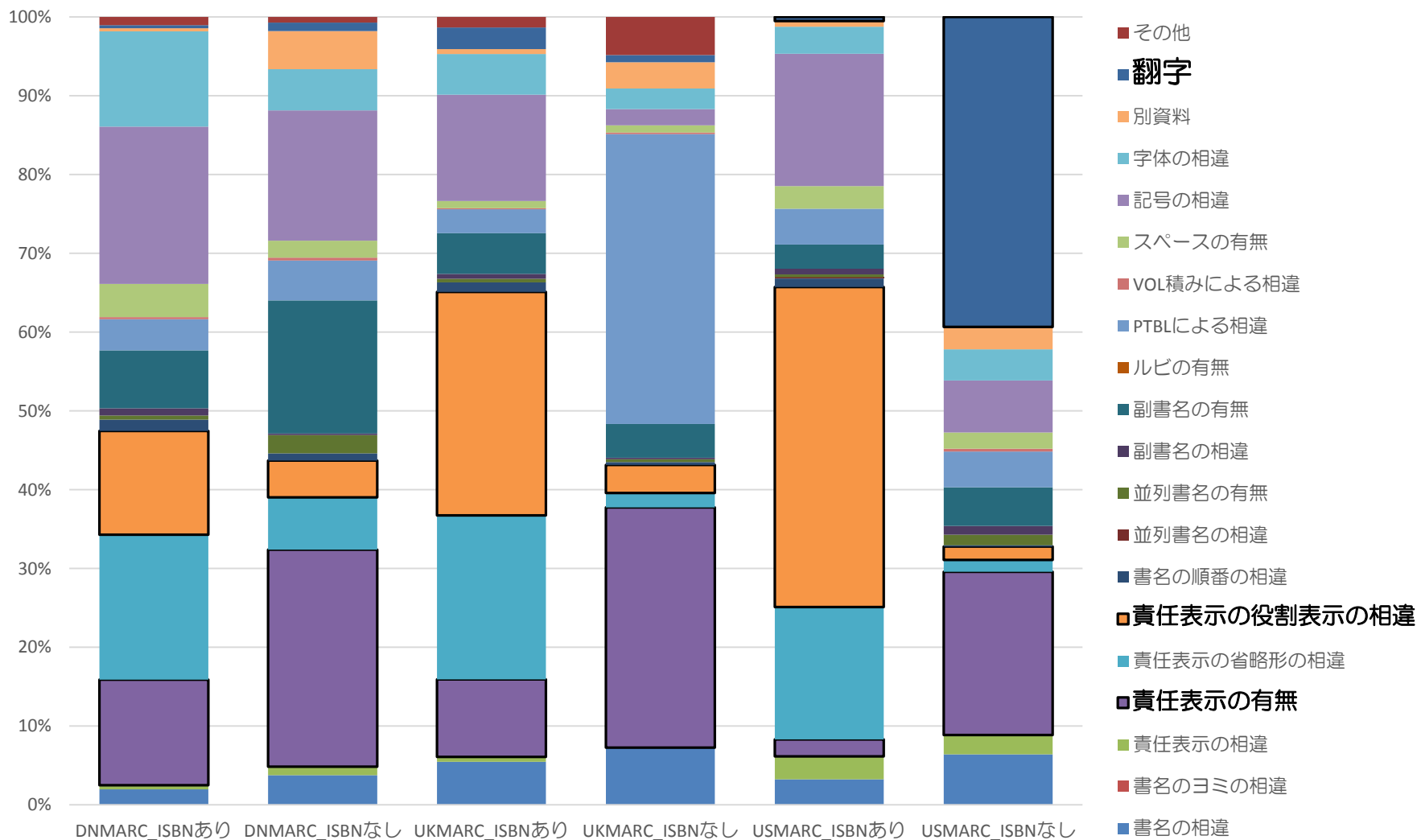
【ISBN有】



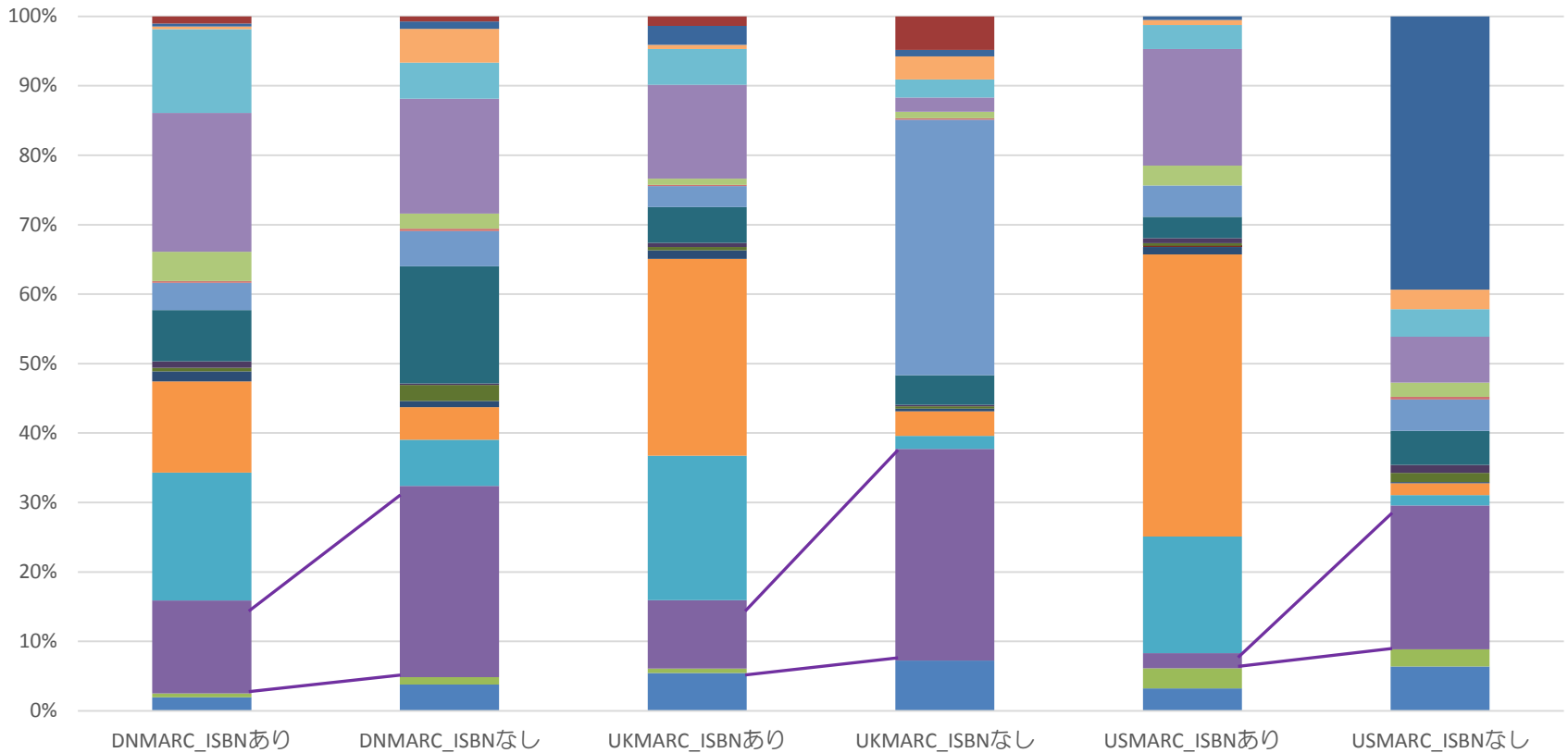
【ISBN無】



不一致項目比較調査の結果分析 (海外MARC TRの不一致項目の分類)

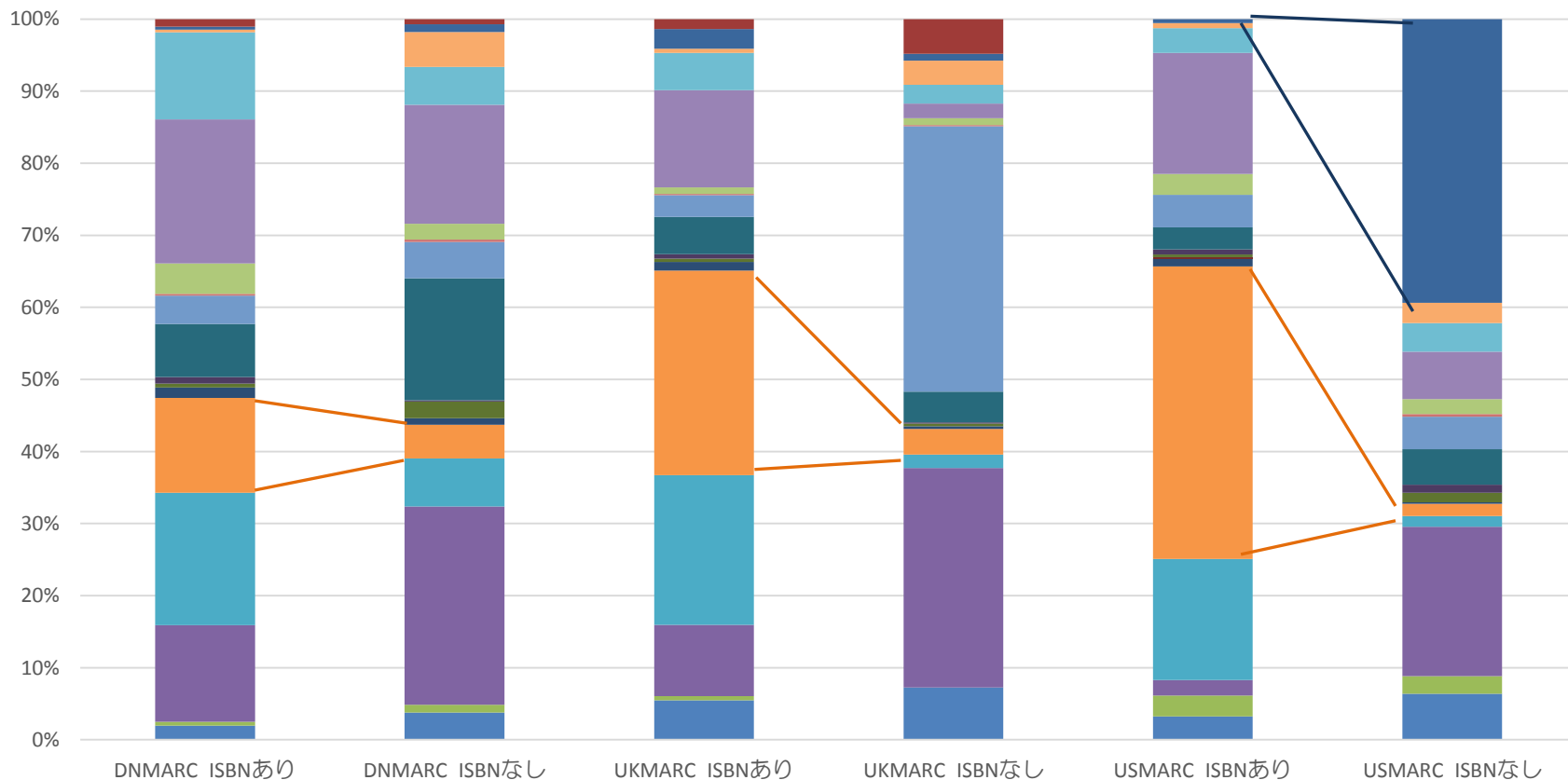


責任表示の有無



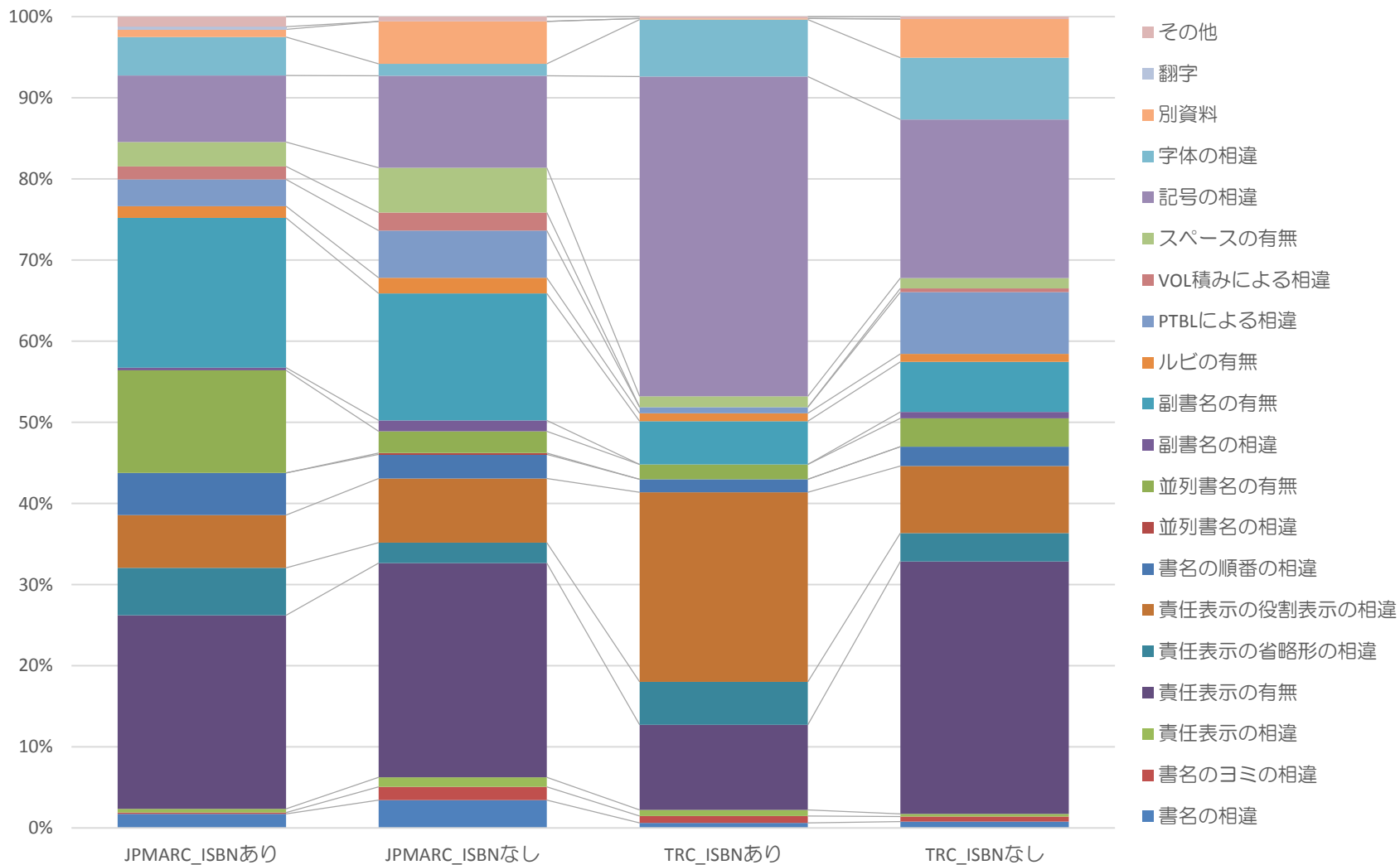
ISBNなし+責任表示なしのデータに
親子関係に依拠するものが含まれる

責任表示の役割の相違 翻字



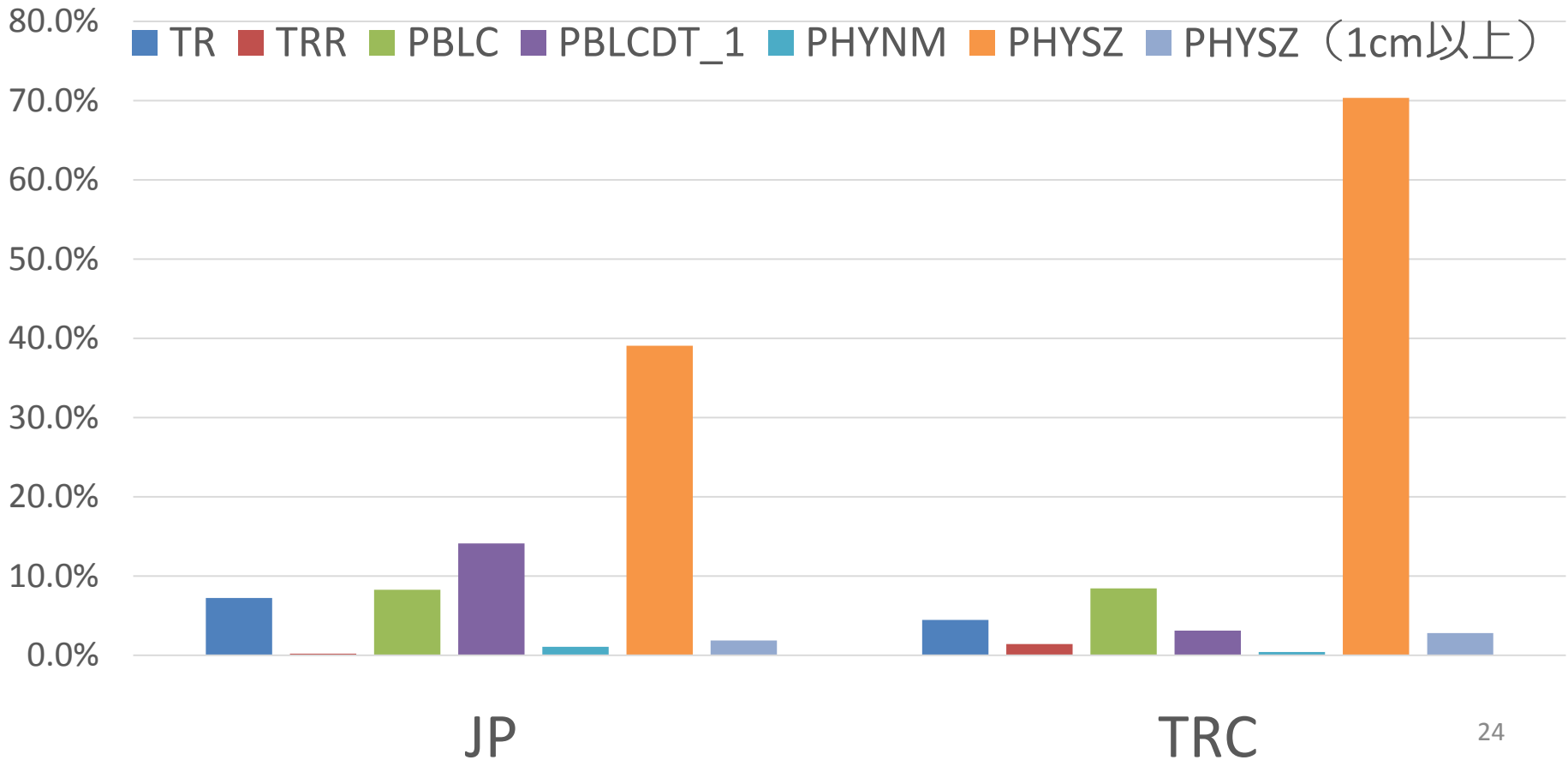
RDA適用 (US:2013.3-) によるルール変更のため、
YR=2014で区切ったISBN有において割合が大きくなっている (UKとUS)

不一致項目比較調査の結果分析 (国内MARC TRの不一致項目の分類)



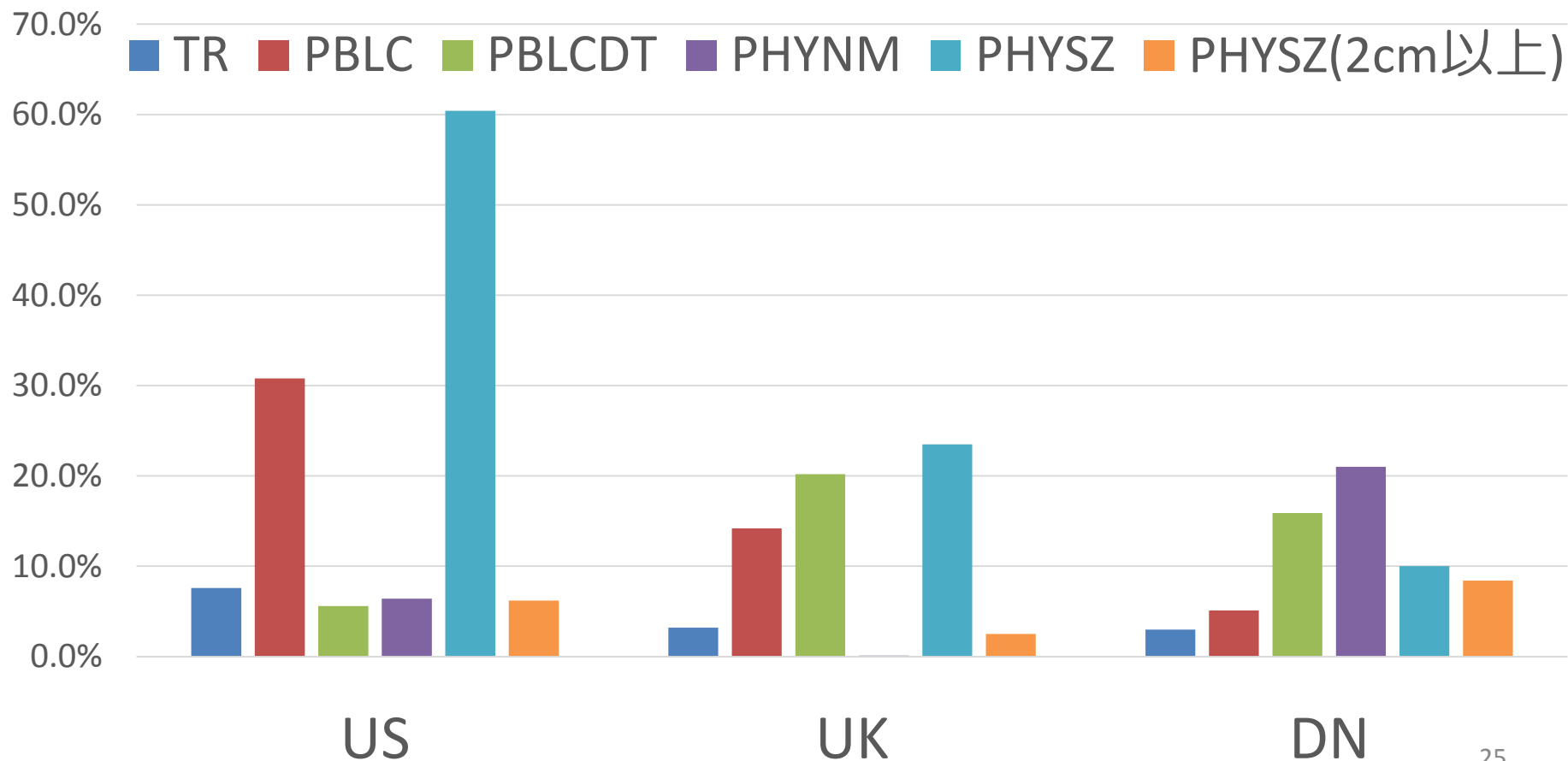
国内MARC

不一致項目（ISBN有500件）のうち
無視できない割合



海外MARC

不一致項目（ISBN有500件）のうちPTBL
による差異を除く無視できない割合



ISBN有と無を比較した 当初の予想との違い

当初

「ISBN有であれば、外部データを無修正で利用できるのでは？」

実際

参照元との一致率比較から、ISBN有が無に比べて一致率が高いわけではない。

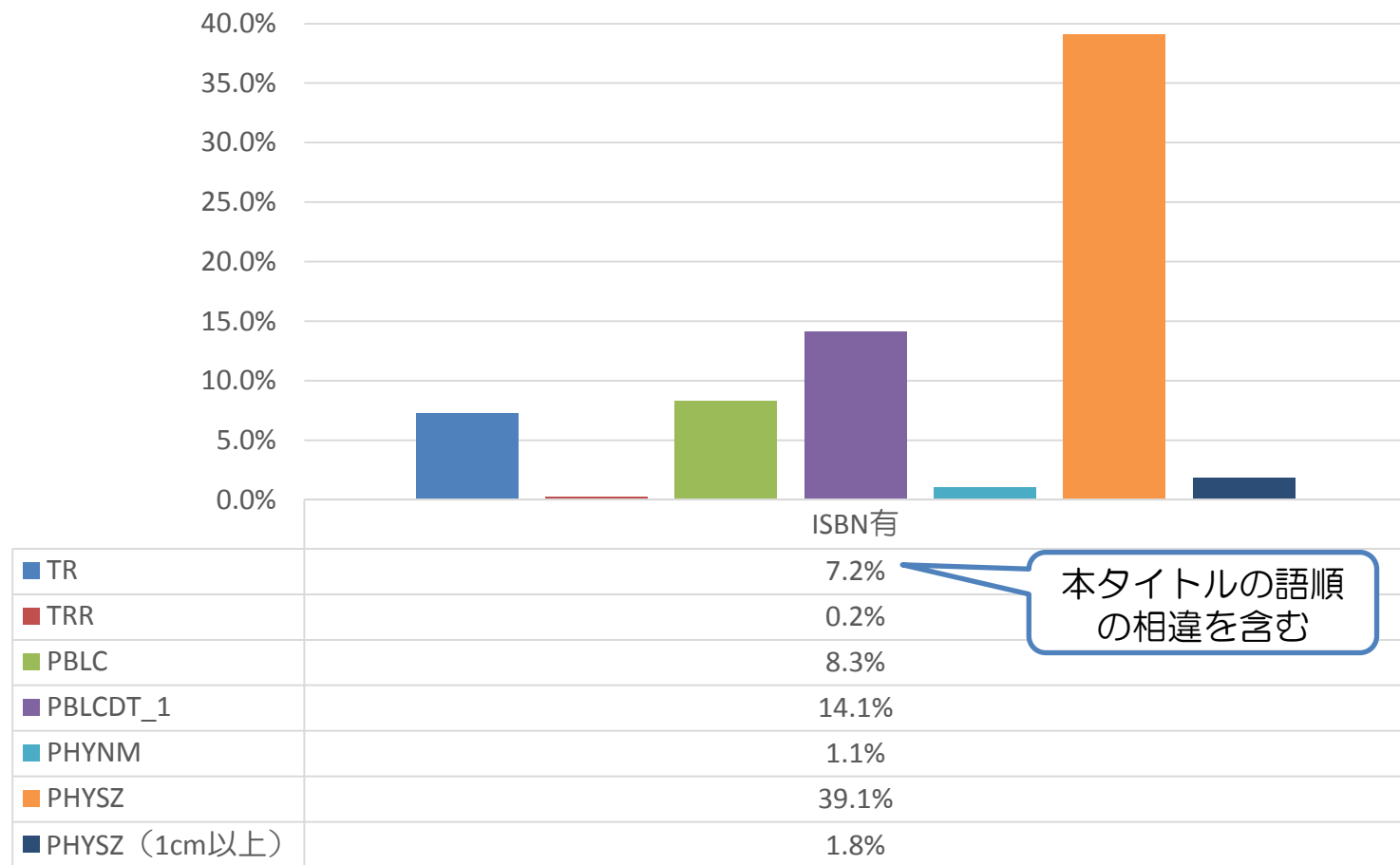
(特に海外MARC)

→ISBN有の不一致は何が原因で、どれくらいあるのか

外部データに必要な要件

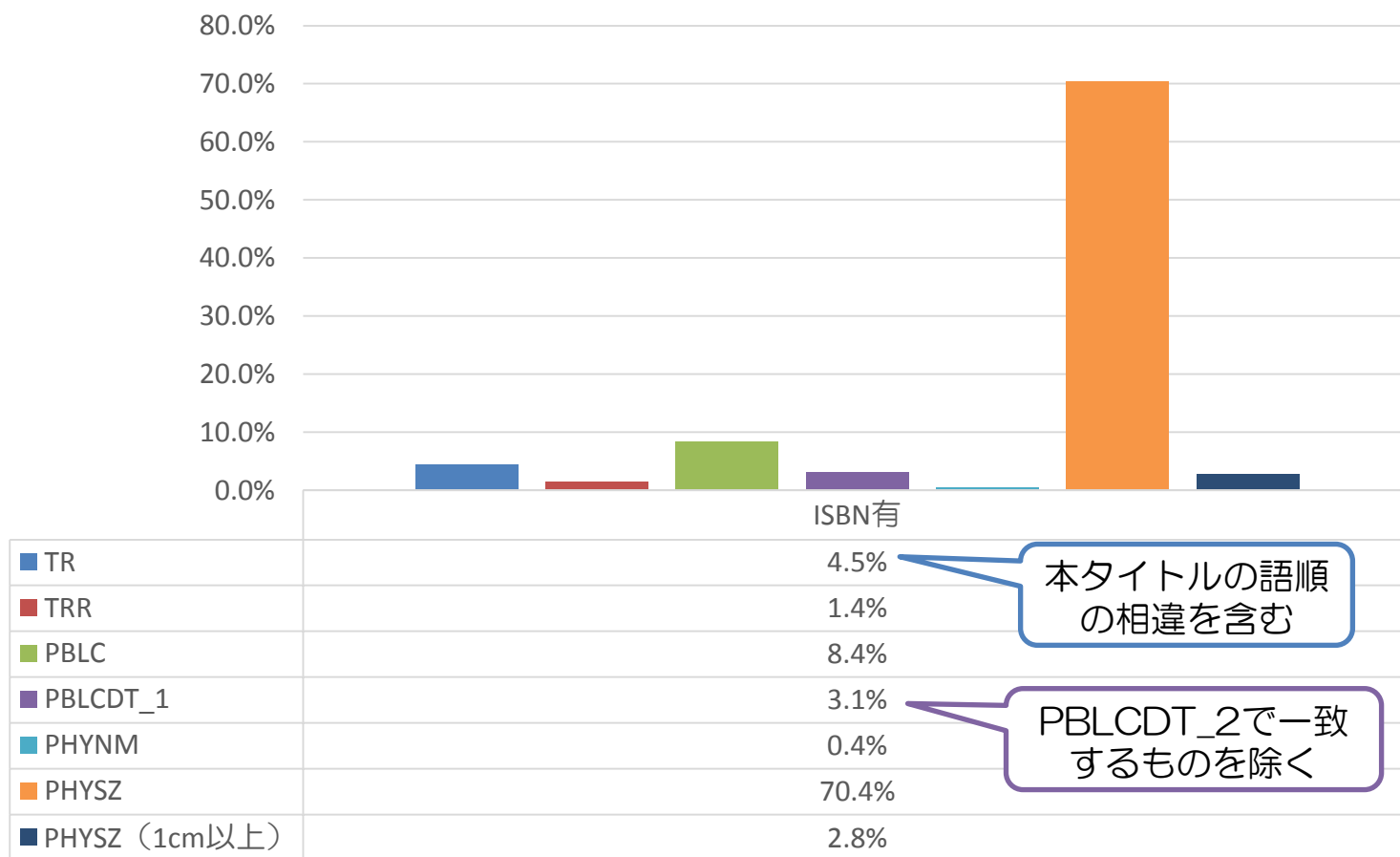
ISBNは十分条件か？（JP）

不一致項目（ISBN有500件）のうち、
無視できない割合



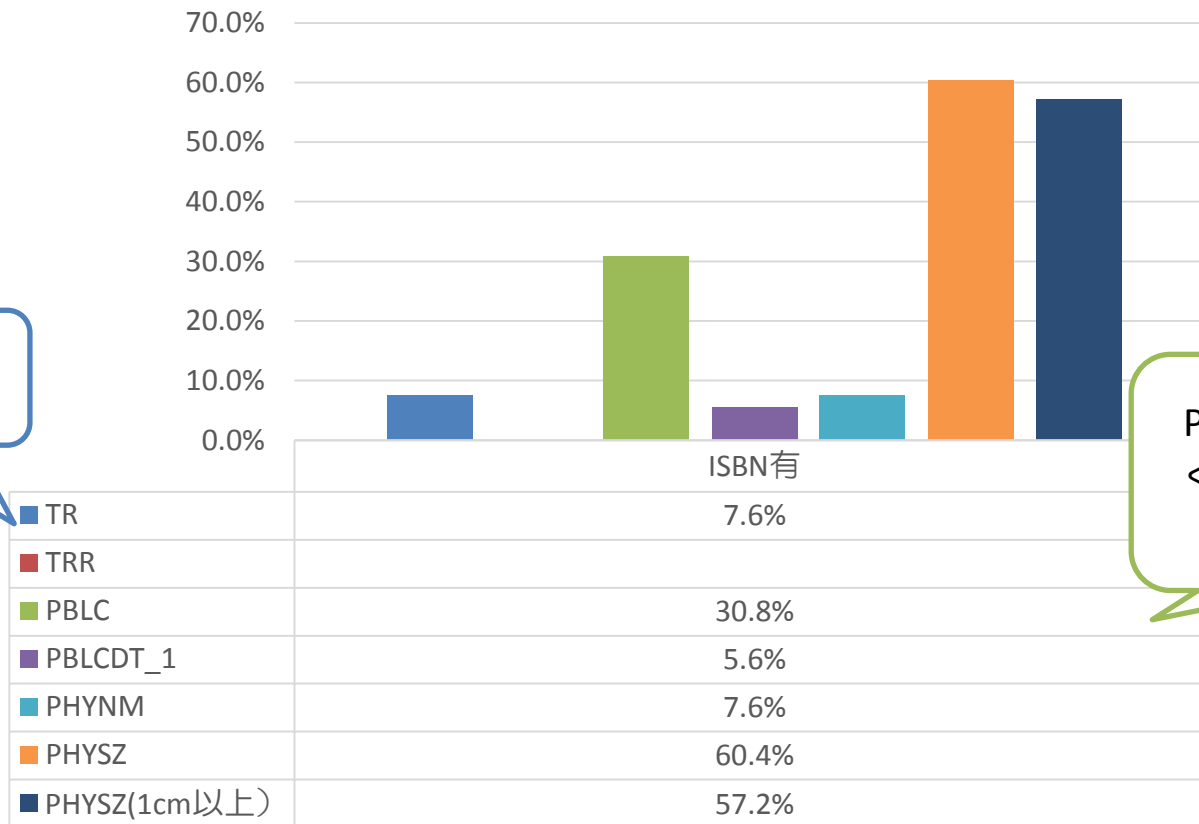
ISBNは十分条件か？（TRC）

不一致項目（ISBN有500件）のうち、
無視できない割合



ISBNは十分条件か？（US）

不一致項目（ISBN有500件）のうち、PTBLによる差異を除く無視できない割合

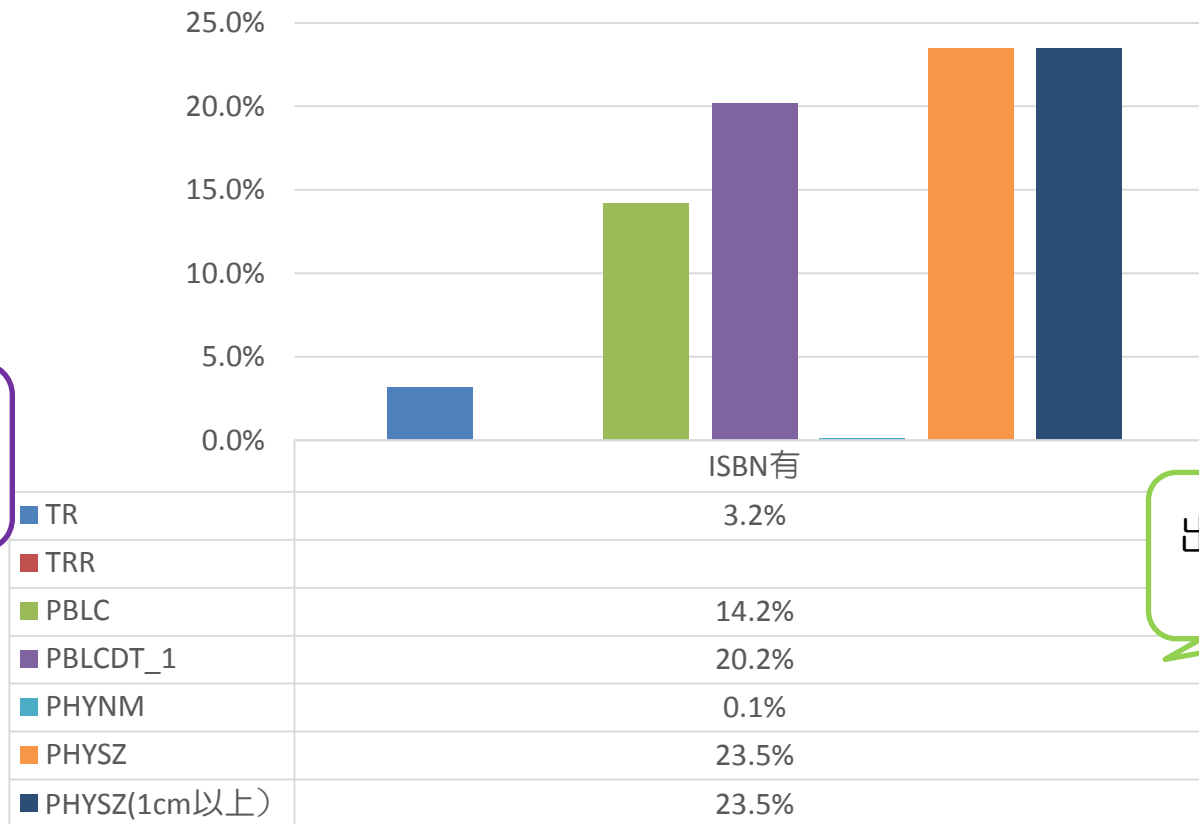


TRは変化なし

PBLCの違いの多くは出版者の変更による

ISBNは十分条件か？（UK）

不一致項目（ISBN有500件）のうち、PTBLによる差異を除く無視できない割合

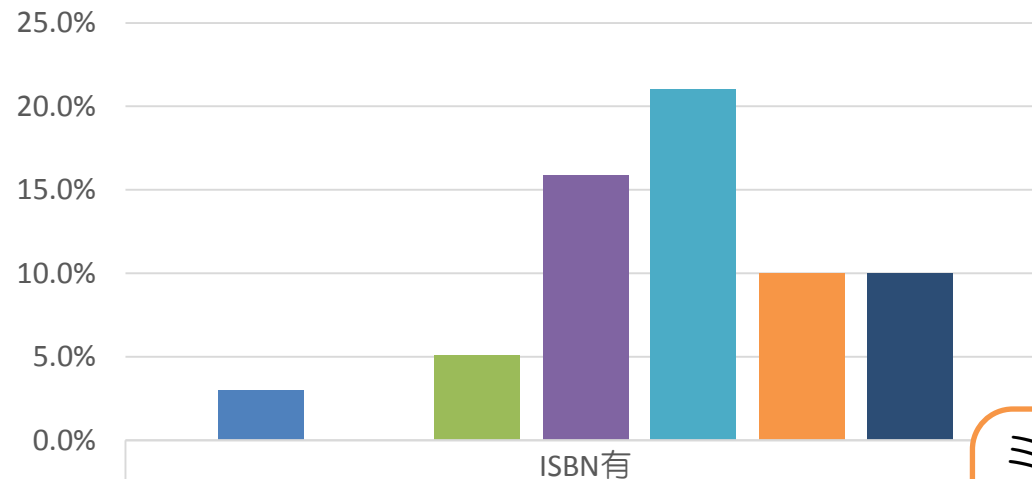


ページ数の相違は極めて少ない

出版年の無い割合が高い

ISBNは十分条件か？（DN）

不一致項目（ISBN有500件）のうち、PTBLによる差異を除く無視できない割合



規則性なく
ページ数の
相違が多い

ミリ表記が特徴
ここでは大きさ
の違いに含めて
いる

■ TR	3.0%
■ TRR	5.1%
■ PBLCDT_1	15.9%
■ PHYNM	21.0%
■ PHYSZ	10.0%
■ PHYSZ(1cm以上)	10.0%

外部データに必要な要件

FRBR(*1) 「Table 6.3 Attributes and Relationships of Manifestations」の**high value**および**medium value**の項目が、LC RDA CORE ELEMENTS(*2)のどこに当てはまるかということ…

1. Title proper
2. Parallel title proper
3. Statement of responsibility relating to title proper
4. Title proper of series
5. Numberig within series
6. Title proper of subseries
7. Nubering within subseries
8. Identifier for the manifestation
9. Carrier type

*1: <http://www.ifla.org/publications/functional-requirements-for-bibliographic-records>

*2: http://www.loc.gov/aba/rda/pdf/core_elements.pdf

外部データに必要な要件

現行NCの項目に当てはめると

- 1～0 は TR
- 4～7 は PTBL
- 8 は ISBN

(9 は図書に限れば固定値のため一旦無視)

→これらの項目は、統制される必要がある
(これまでの「要修正となる相違」とほぼ一致)

外部データに必要な要件 (国内MARC)

ISBNのあるMARCを流用入力した際に、修正が必要となる割合は...

(TRが誤っているときに修正しなければならない割合) × (TRが誤っている割合)
= (無視できない誤り ÷ 誤りレコード数) × (TRが不一致 ÷ (TRが一致 + TRが不一致))

JP = 2.14% → 47冊に1冊

TRC = 2.18% → 46冊に1冊 要修正

外部データに必要な要件

- ISBNのあるUSMARCを流用入力した際に、修正が必要となる割合は...

(TRが誤っているときに修正しなければならない割合) × (TRが誤っている割合)
= (無視できない誤り ÷ 誤りレコード数) × (TRが不一致 ÷ (TRが一致 + TRが不一致))

US = 6.02%	→ 17冊に1冊	} 要修正
UK = 3.97%	→ 25冊に1冊	
DN = 3.43%	→ 29冊に1冊	

新モデルについて

モデルの詳細化

- 各種MARCの無修正利用) 調査の結果
 - 作業の省力化が見込める
 - × 一部はどうしても修正が必要これを踏まえモデルを精緻化する必要がある。
- 書誌データの維持管理モデル
- 所蔵リンクの集約保持モデル
- 書誌と所蔵の分離保持

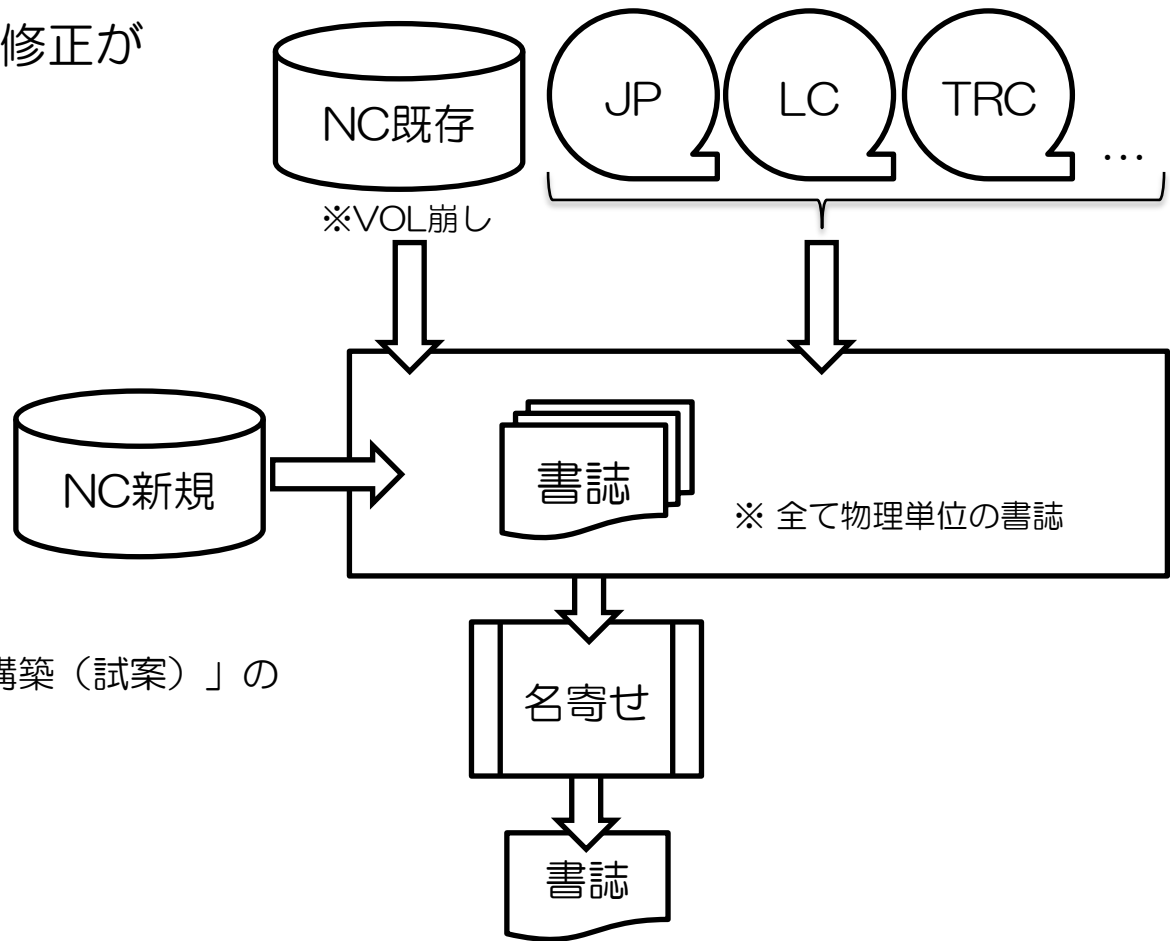
書誌データの維持管理モデル

当初の想定通り一切修正が
いらなければ…

- ①既存の書誌
- ②外部MARC
- ③新規データ

を名寄せすれば
すべて完了

「NAC SIS-CAT/ILL再構築（試案）」の
名寄せ部分までに相当

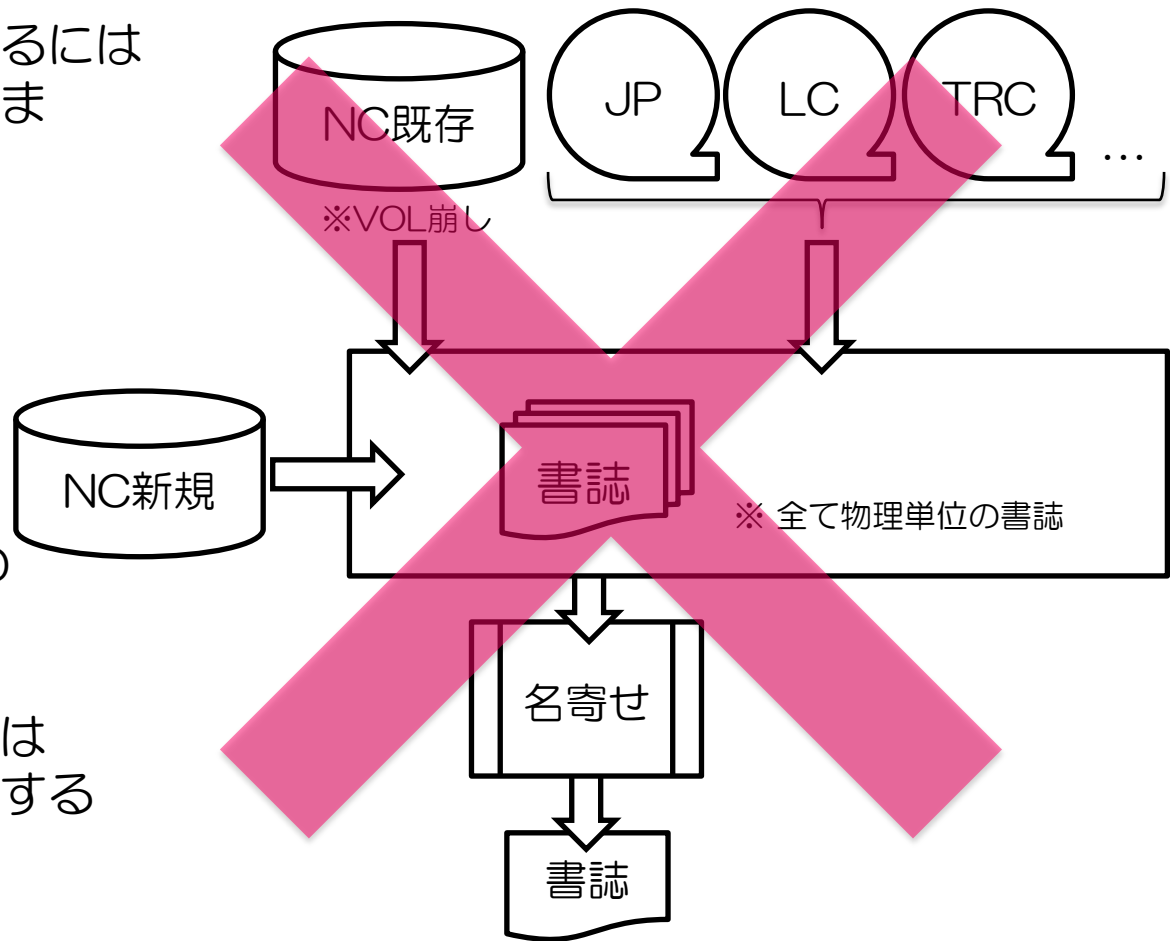


書誌データの維持管理モデル

データ修正に対応するには
このモデルをそのまま
採用できない

候補は3つ

- ①名寄せ結果のレコードを修正する
- ②外部MARCからのレコードを直接修正
- ③修正したレコードは新規レコード扱いとする



書誌データの維持管理モデル

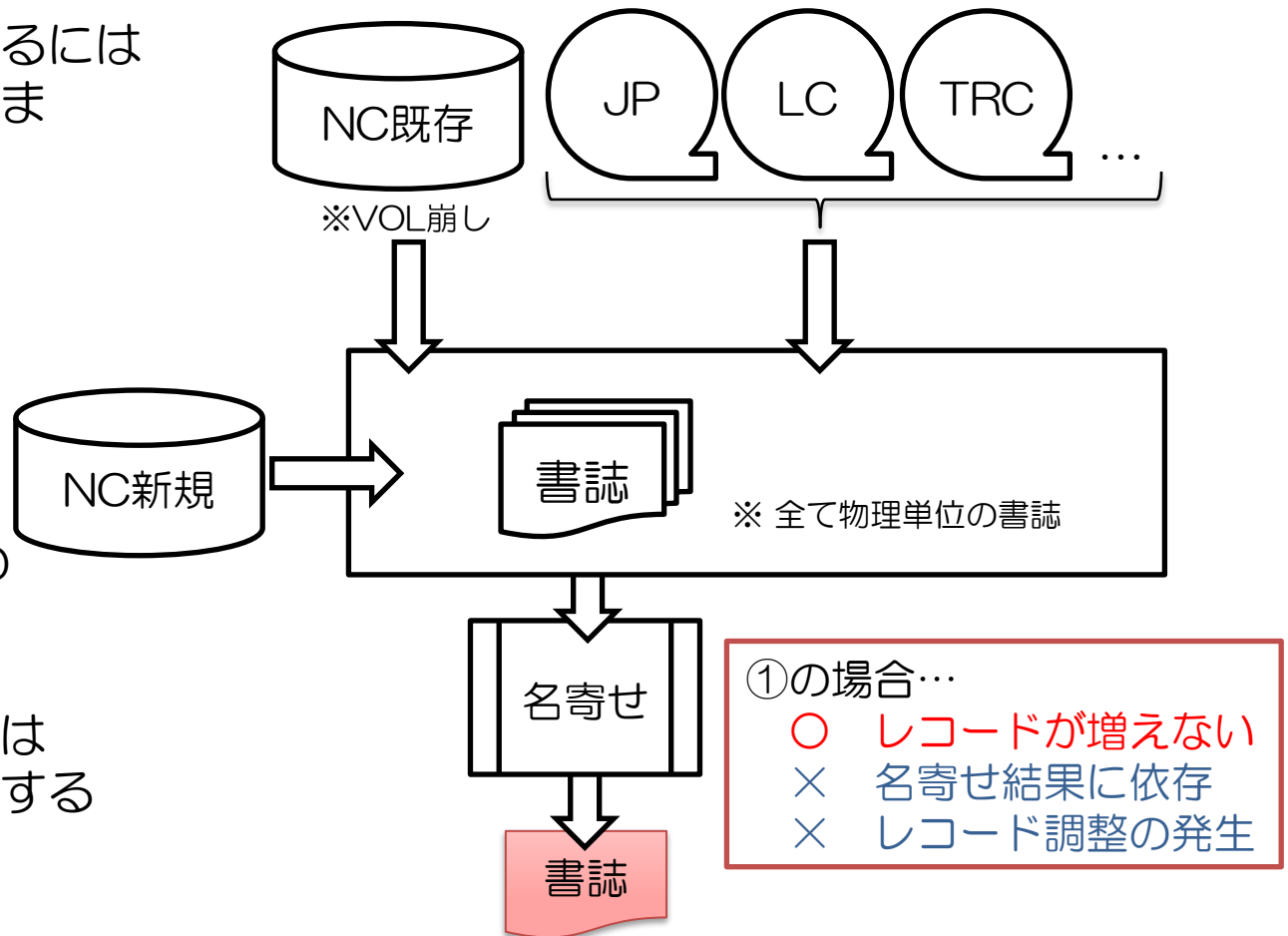
データ修正に対応するには
このモデルをそのまま
採用できない

候補は3つ

①名寄せ結果の
レコードを修正する

②外部MARCからの
レコードを直接修正

③修正したレコードは
新規レコード扱いとする



書誌データの維持管理モデル

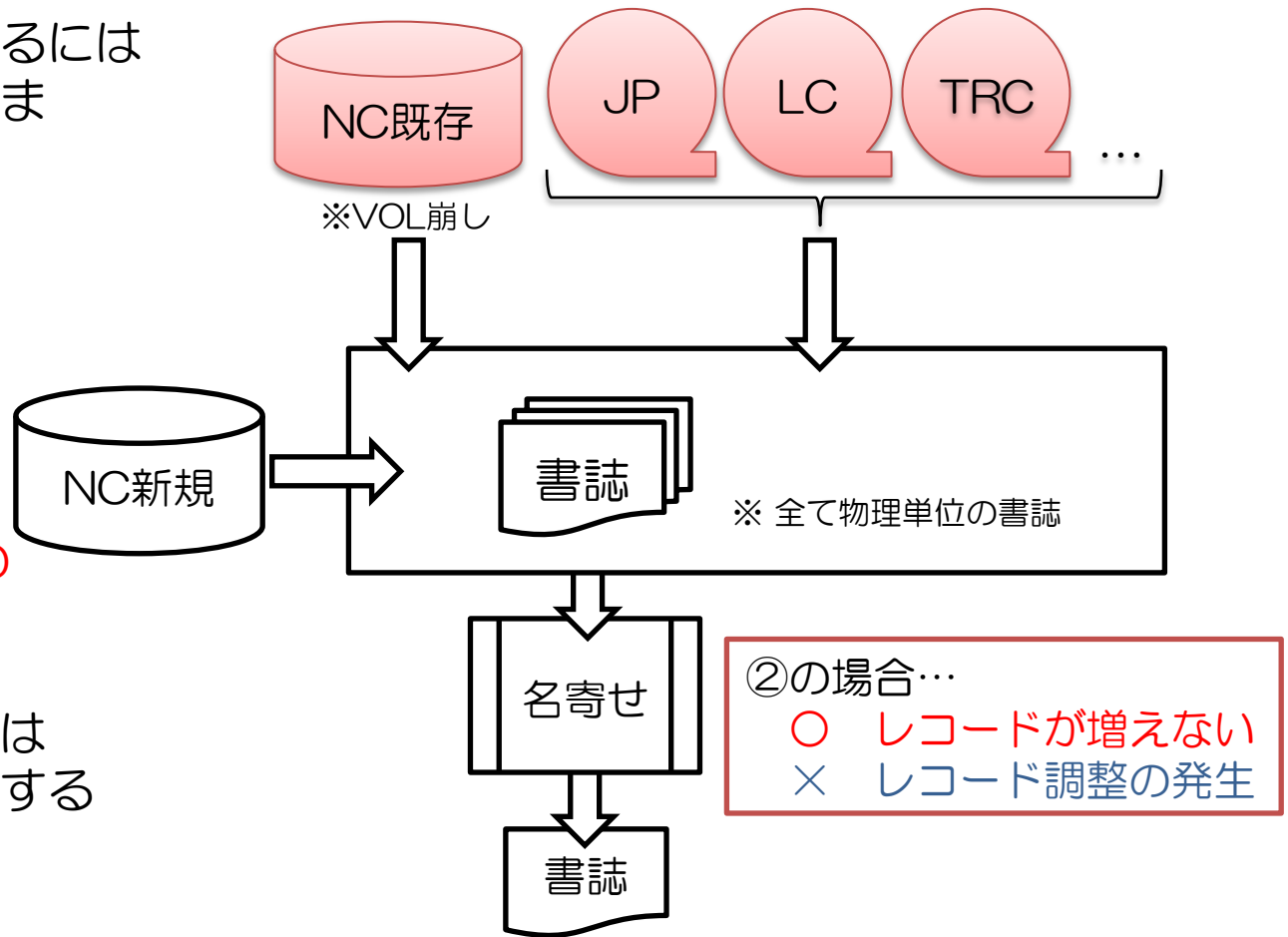
データ修正に対応するには
このモデルをそのまま
採用できない

候補は3つ

①名寄せ結果の
レコードを修正する

②外部MARCからの
レコードを直接修正

③修正したレコードは
新規レコード扱いとする



書誌データの維持管理モデル

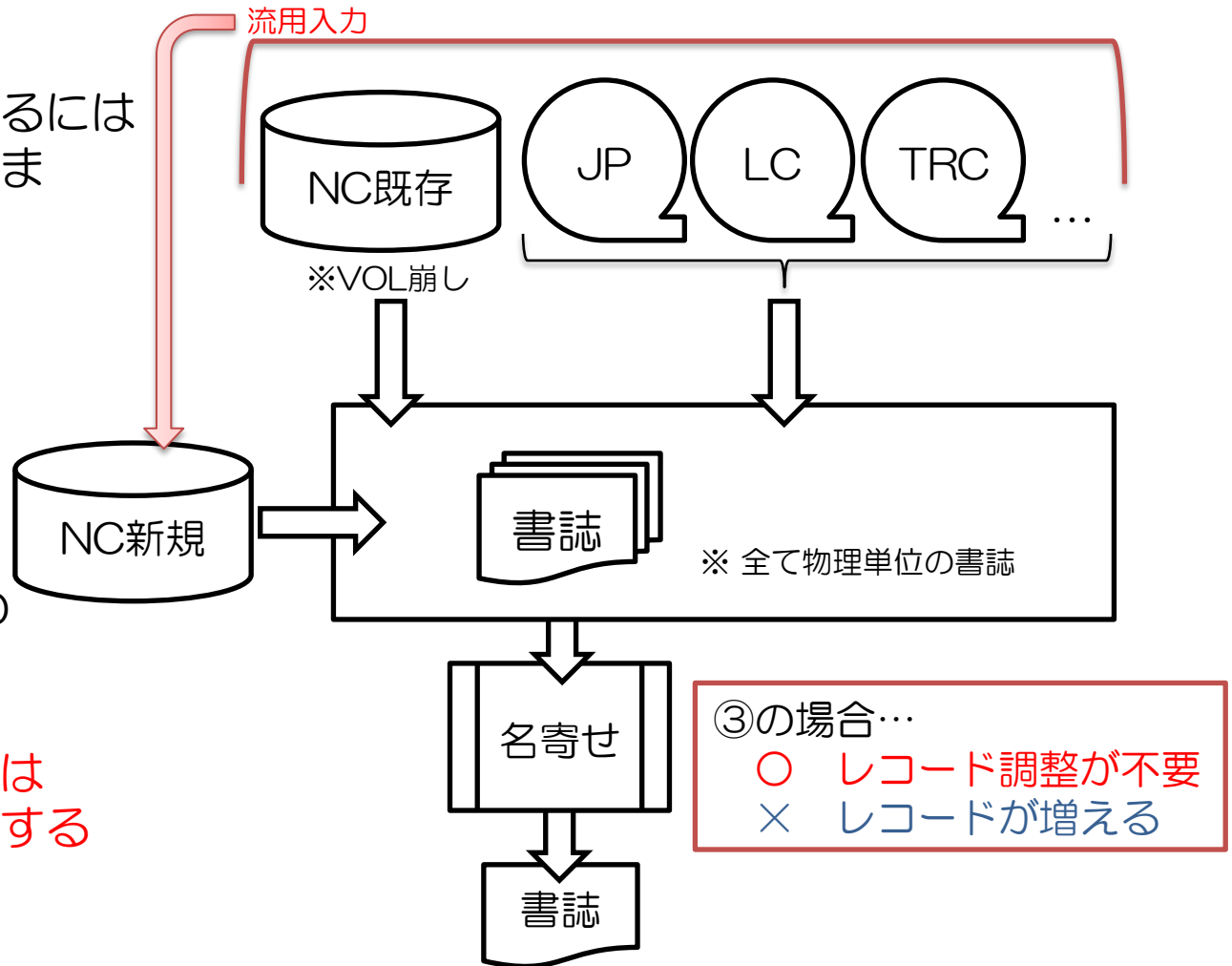
データ修正に対応するには
このモデルをそのまま
採用できない

候補は3つ

①名寄せ結果の
レコードを修正する

②外部MARCからの
レコードを直接修正

③修正したレコードは
新規レコード扱いとする



書誌データの維持管理モデル

③修正したレコードは
新規レコード扱いとする

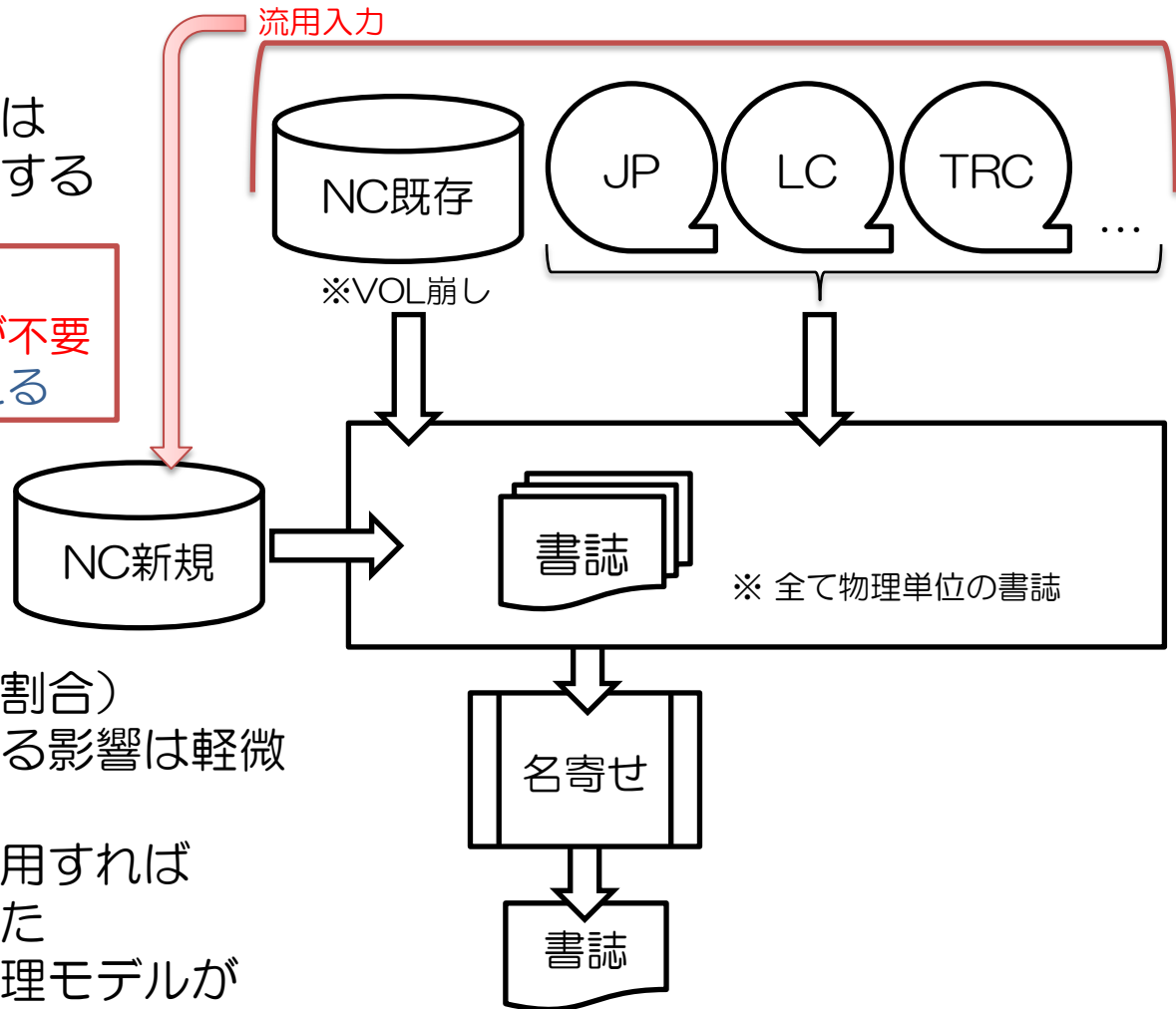
③の場合…

- レコード調整が不要
- × レコードが増える

レコードの増加は
本調査結果より
2%~6%程度

(要修正のレコード割合)
名寄せや運用に与える影響は軽微

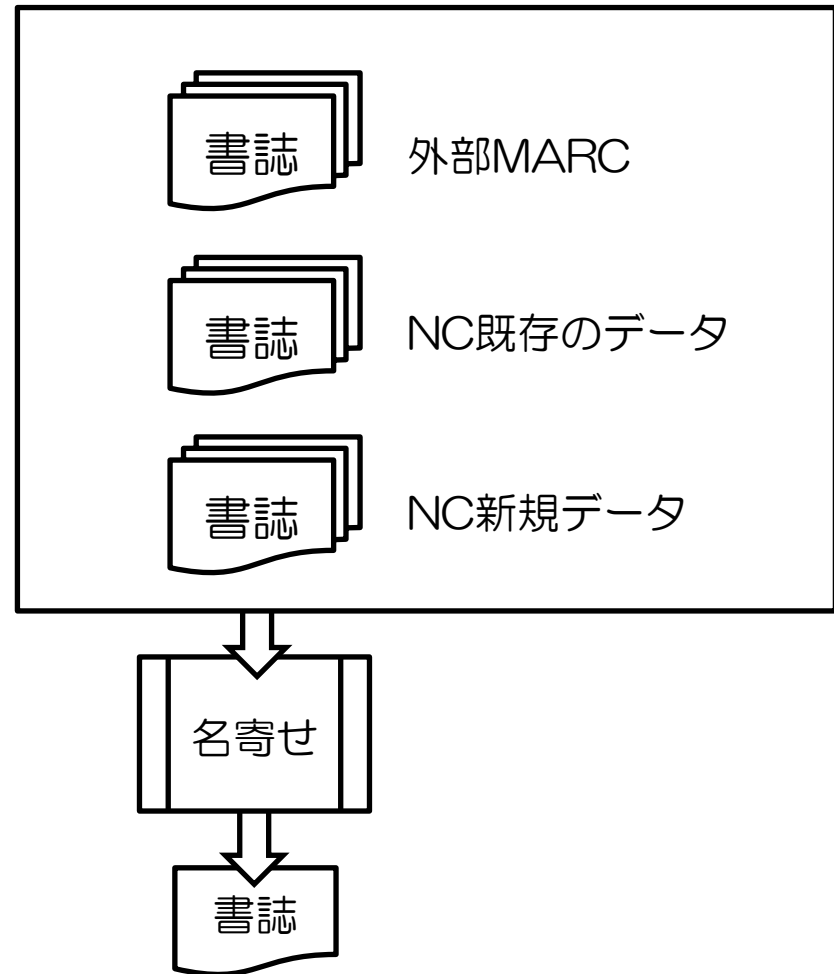
したがって③案を採用すれば
外部データを活用した
書誌データの維持管理モデルが
実現できる



所蔵リンクの集約保持モデル

所蔵をどこにつけるか

- ①名寄せ済書誌に付ける
- ②名寄せ前書誌に付ける



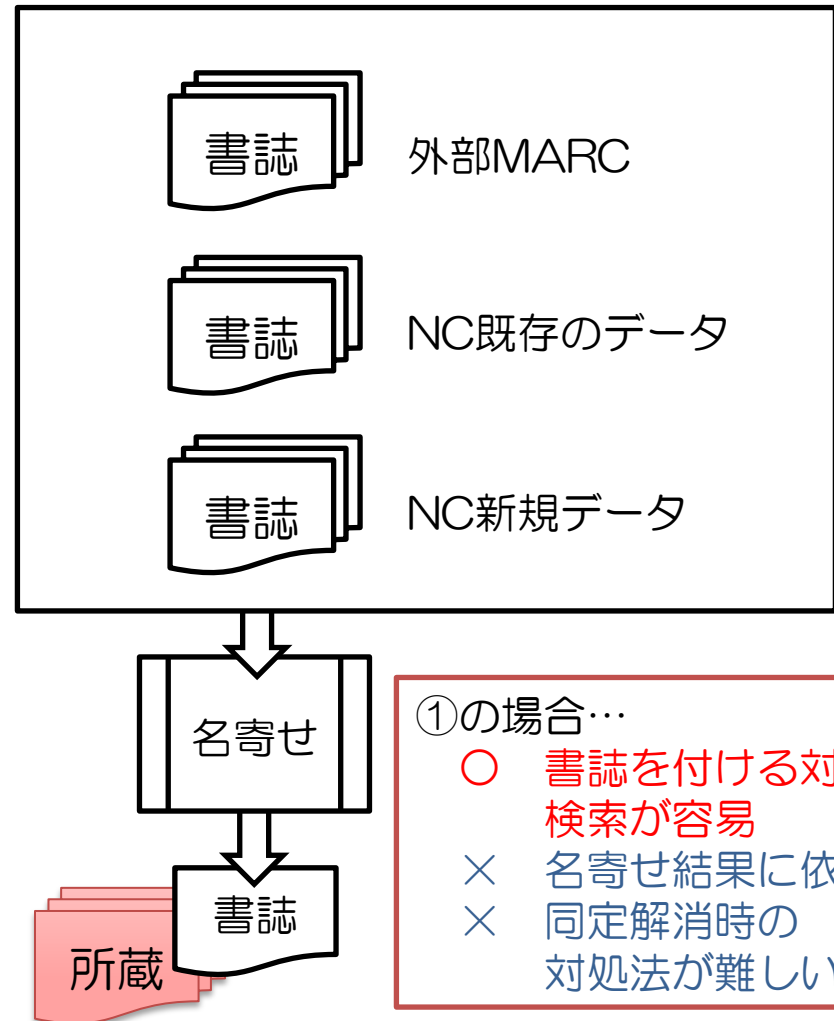
所蔵リンクの集約保持モデル

所蔵をどこにつけるか

- ①名寄せ済書誌に付ける
- ②名寄せ前書誌に付ける

①の場合…
名寄せ済書誌を検索して、
所蔵を付ける対象の書誌を
特定する。

新規データが重複していても
検索性は変化しない



- ①の場合…
- 書誌を付ける対象の検索が容易
 - × 名寄せ結果に依存
 - × 同定解消時の対処法が難しい

所蔵リンクの集約保持モデル

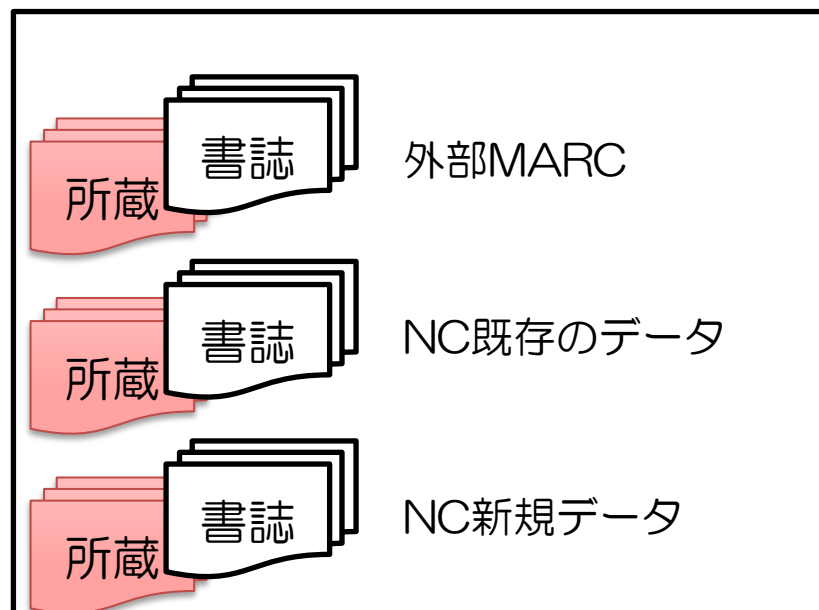
所蔵をどこにつけるか

- ①名寄せ済書誌に付ける
- ②名寄せ前書誌に付ける

②の場合…
名寄せ前書誌を検索して、
所蔵を付ける対象の書誌を
特定する。

新規データが重複すると
検索性が大きく低下する

⇒ 名寄せ済データを検索して
名寄せ前書誌に所蔵を付けられる
仕組みがあれば良いため
②を採用する



- ②の場合…
- 名寄せ結果に依存しない
 - 同定条件に機関毎の所蔵情報を利用可
 - × 新規データの検索性低下

書誌と所蔵の分離保持

書誌（⇒NACISIS-CAT/ILL）と所蔵（⇒ローカル）を分けて持つ

既存機能を実現するためにNACISIS-CAT/ILLで
保持する必要がある所蔵データは、

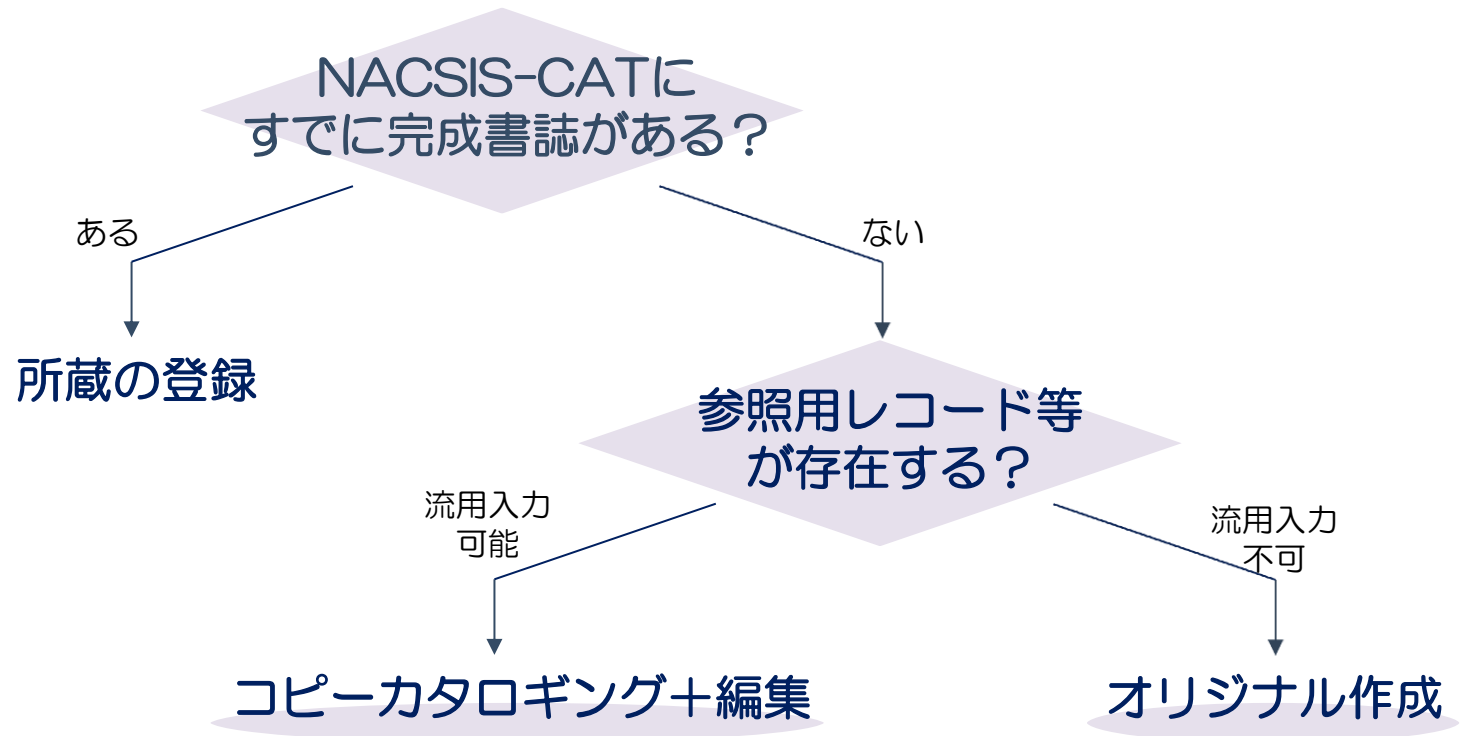
資料へのID （MARC番号など；BIDに相当）
図書館ID・機関ID
RGTN

上記以外はNACISIS-CAT/ILLに保持しない。

ローカル側への書誌ダウンロードは、NACISIS-CAT/ILLから
好きなデータを持ってきてローカルに登録することが出来る。
相互貸借など、NACISIS-CAT/ILL側からRGTNを元に
各館システムに問合せする。

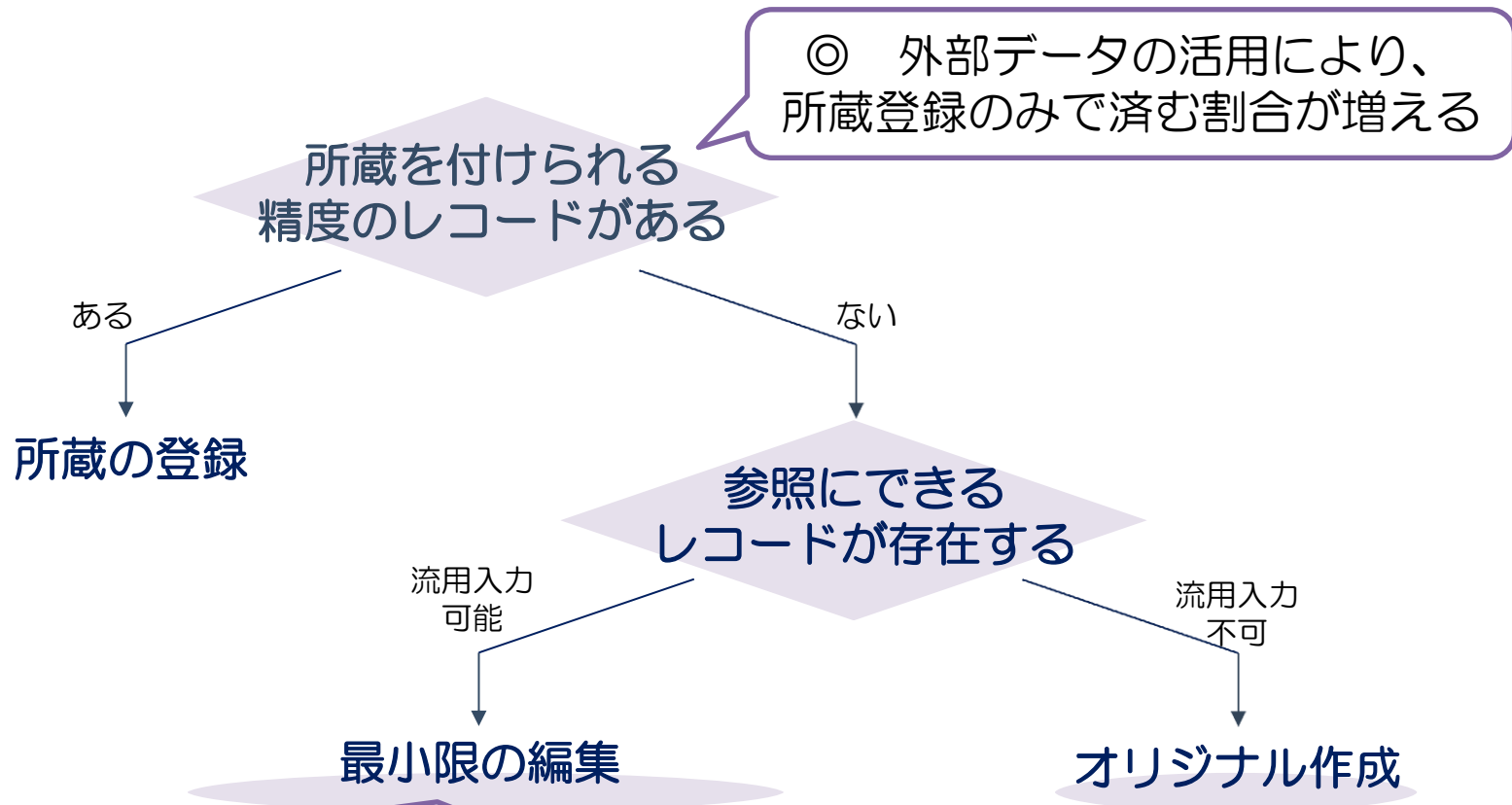
所蔵登録書誌作成作業の変化

現行のフロー



所蔵登録書誌作成作業の変化

新モデルのフロー



◎ 編集対象項目は限定されるため、
これまでよりも省力化できる

新モデルで変わること

所蔵登録書誌作成業務

- ◎ これまでNACSIS-CATの上で所蔵を付ける書誌を探さなくてはならなかったが、新モデルでは受け入れ時のMARC番号などをそのまま活用できるようになる。
- ローカル所蔵メンテナンスに伴うNACSIS-CAT所蔵メンテナンスの作業がなくなる。

CiNii Books

- 見た目上はこれまでと同様のサービスを提供可能。

システム

- △ 書誌と所蔵を分けて持つため、完全なサービスを提供するには、NACSIS-CAT/ILLと各館のローカルサービスが両方常に稼働している必要がある。