

# 識別と記述のフレームワーク

宮澤 彰

国立情報学研究所  
総研大

2011-10-22

情報組織化研究グループ

# 目次

1. 《書誌コントロールを超えて》の内容
2. 登録機関と書誌コントロール機関
3. 識別の考察
4. 知的生産物の分野
5. JPOとEDItEUR  
登録機関候補の例として
6. 書誌コントロール機関のタスク
7. 今後の検討課題

# 《書誌コントロールを超えて》

- ◆ 書誌コントロールの究極の目的は、人類の知的生産物すべてに関して探索および検索（探索したものにアクセスすることを含む）を可能とすることである。
- ◆ 知的生産物の範囲は、文字作品だけではなく、映像、音楽等すべてのタイプ、伝統的出版だけでなく、ネットワークを含む総てのメディアにわたる。

## 《書誌コントロールを超えて》 続

- ◆ メディアの範囲だけでなく、著作や表現形のようなクラスの範囲をどこまで考えるかという点も考慮しなければならない。
- ◆ 主として表示，人間による識別のための情報として使われる記述と，一次資料の情報に何らかの処理を加えて作成され，コロケーション機能を提供するアクセスポイント，という区分は，この中でも意味を持つ。

# 《書誌コントロールを超えて》 続

## ◆ 新しい書誌コントロールの方式

1. 従来の目録作成方式によるネットワーク資源の目録作成
2. サーチエンジンへの提供
3. 登録機関と書誌コントロール機関

# 登録機関と書誌コントロール機関

- ◆ 著作登録機関⇒著作権処理機関
- ◆ 体現形登録機関⇒メディアごと
- ◆ 書誌コントロール機関: 著作と表現形の識別と関連づけ, 個人団体と著作, 体現形の関連づけ, 著作と概念などの主題関連づけなど

# 登録機関と 書誌コントロール機関

2011-10-22

情報組織化研究グループ

7

# 考え方の前提

- ◆ 前提として、これまでの図書館での目録作業とは大きく異なる、様々な機関でのメタデータ作成と、これらの関連づけという形を考えている。
- ◆ また、こういった情報を機械可読形式で公開するセマンティックウェブ的な方法が広がる、という予想をしている。



# 社会のIT化

- ◆ IT化の進んだ現代では、社会的に流通するものは、いずれにしても、何らかの記述をしてどこかに登録するようになっている。
- ◆ 流通するもののメタデータは、ほとんどの場合、どこかに存在する。
- ◆ たとえば、出版業界における近刊情報センターのように、業界団体に組織化する場合がある。

# 社会のIT化 続

- ◆ 個人経営の古書店であったとしても、現在では、EXCEL等での管理はしている、あるいはするようになるだろう。
- ◆ 問題は、これらのメタデータをいかに流通させ組織化して、社会的に利用可能にするかという点にある。

# 知的所有権の確立

- ◆ 知的権利の保護も、社会的に広く認められるようになってきており、この流れは今後も広まっていくと考えられる。
- ◆ このため、著作権の処理も組織的に行われるようになってきている。

# 登録機関

- ◆ 体現形，著作の登録機関を考える理由は，業界団体等で既に整備されている情報は，それを公にする (to publish) ことが，組織化の基礎であるため。
- ◆ 登録機関は，基本的にその業務処理で作成しているメタデータを (業務に差し支えない範囲で) 公開する。

# 書誌コントロール機関

- ◆ 登録機関によって登録されたメタデータをもとに、実体間の関連づけを行う。
- ◆ 伝統的な目録の用語でいえば、典拠作業および主題目録作業ということになる。

# 「識別」の考察

2011-10-22

情報組織化研究グループ

14

# 識別とは

- ◆ 識別とidentification
  - ❖ 「区別すること」と「同じである」こと
- ◆ 記述対象の識別, 関係の識別と属性値の識別
  - ❖ 「これが, ひとつのitem, manifestation, ...である」ということ ⇔ 書誌単位問題。
  - ❖ あるitemがあるmanifestationをexemplifyする等の関係をつけること。
  - ❖ 排他的な属性値のひとつを割り当てること。

## 識別 続

- ◆ 関係の識別や属性の識別タスクには、時として、関係づけられる実体や属性値の追加、修正を伴うことがある。

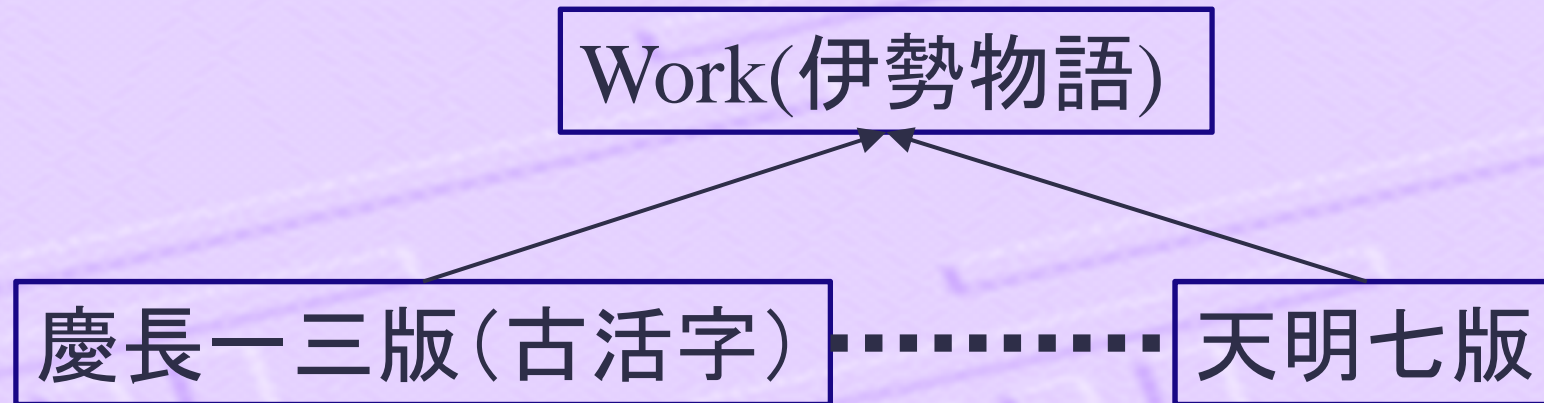
例：書誌レコードと著者名典拠ファイルの関係識別タスク

- ◆ 識別タスクでは通常、属性値の一致をもって識別を行うことになっているが、現実是一般にそれほど単純ではない。

二つの「もの」の総ての属性値が等しければ、この二つの「もの」は同一である？

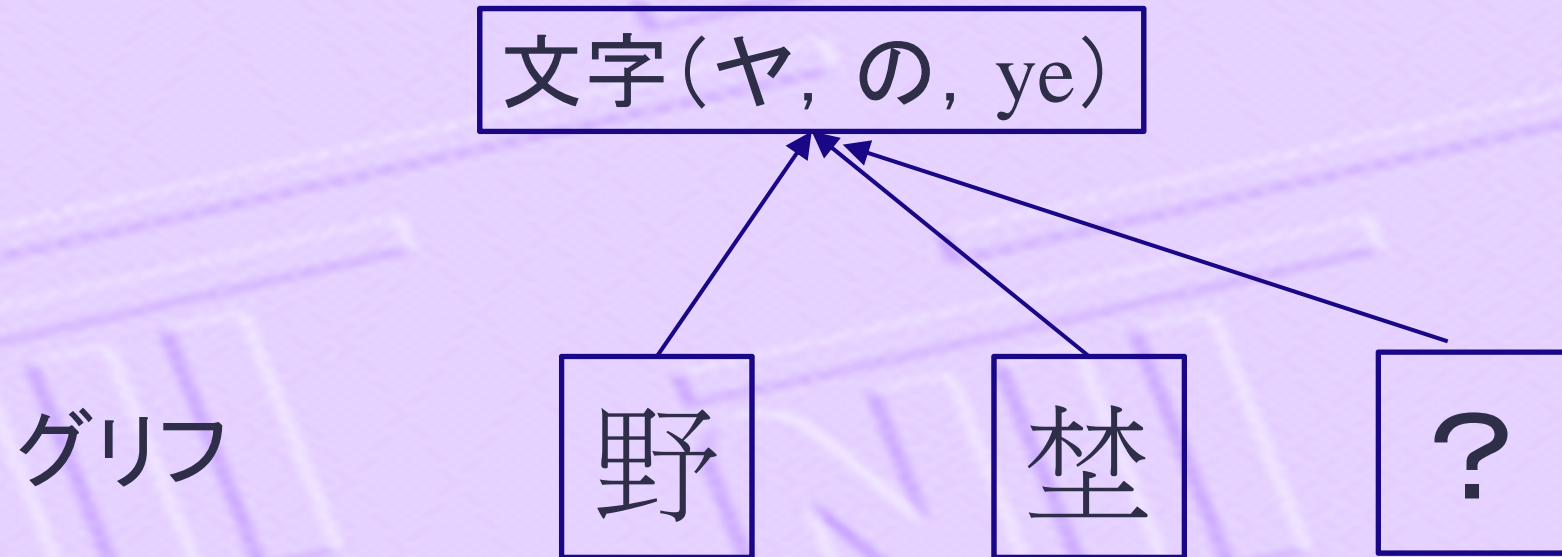


# 例: manifestationの識別





# 例：文字ーグリフモデル



実際の字形をグリフ/文字として識別するタスク

## 識別 続

- ◆ どれだけの属性を比較すればオブジェクトとしての同一性が保証できるかは不明な場合が多い。(アルゴリズムによる識別ができない)。
- ◆ 関係づけられる側の集合も、確定していない場合が普通。
- ◆ だれかが行った識別に多数が従うという、社会的な解決しかとり得ない。

# 知的生産物の分野

# 形態による

- ◆ 言語
  - ❖ 文学, 論文, 記録, ...
- ◆ 静止画
  - ❖ 絵画, 写真, ...
- ◆ 動画
  - ❖ 映画, ビデオ, ...

# 形態による 続

- ◆ 音響

- ❖ レコード, CD, ...

- ◆ もの

- ❖ 工芸品, 工業製品, ...

これらの混合形態

いずれにしても, 排他的な分類は難しい

# 流通形態による

- ◆ 出版
- ◆ ネットワーク
- ◆ dealerを通じた個別流通

(大量)複製と商業的流通では、  
manifestationの概念が有効となる。



# 業界による

- ◆ 書籍出版
- ◆ 雑誌
- ◆ 映画
- ◆ 音楽(CD, iPod等)
- ◆ 美術品

社会に応じて変化する。

グローバル化の傾向はあるが、まだ国による多様性は大きい

# 知的生産物の種類

- ◆ 今後カテゴリーの確立が必要
- ◆ 図書館資料のわくにとらわれず，広い範囲で考えられるべき。
  - ❖ 美術品，記録類，工業製品等も考慮すべき。
- ◆ 分類にあたっては，生産，流通，保存，管理の体制を含めた社会的合意が必要。

# JPOとEDITEUR

## 登録機関候補の例として

# JPO沿革

- ◆ 2002 日本出版データセンター 設立
- ◆ 2004 日本図書コード管理センター (ISBN) 統合
- ◆ 2006 共有書店マスタ
- ◆ 2007 出版RFID
- ◆ 2009 日本出版インフラセンター に改名
- ◆ 2011 近刊情報センター

# 事業活動

- ◆ 出版情報の収集配信
- ◆ 出版情報の標準化
- ◆ 出版における各種コードの体系化と管理
- ◆ 流通迅速化の研究
- ◆ 知的財産のビジネスモデル研究
- ◆ 著作権等管理

# 組織

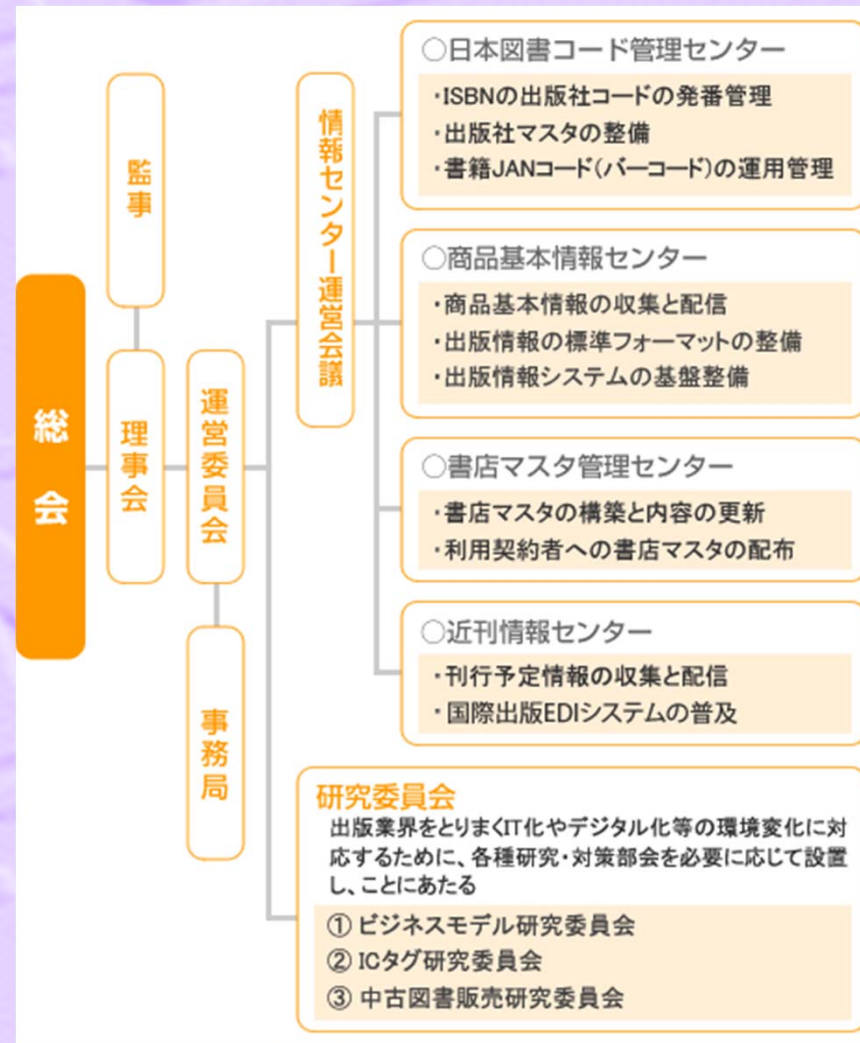
## 設立団体:

日本書店商業組合  
連合会、日本出版  
取次協会、日本書  
籍出版協会、日本  
雑誌協会、日本図  
書館協会

## 社員(会員):

出版社, 取次, 大手  
書店, 新聞社等

専任は4名



# 近刊情報の書誌項目

- 1 ISBN
- 2 取引コード 4ケタ
- 3 発行元出版社
- 4 発売元出版社
- 5 部署
- 6 入力者名
- 7 電話
- 8 FAX
- 9 E-mail
- 10 Cコード

# 近刊情報の書誌項目 続

- 11 ジャンルコード
- 12 書名
- 13 書名 読み
- 14 サブタイトル
- 15 サブタイトル 読み
- 16 レーベル
- 17 レーベル読み
- 18 シリーズ名
- 19 シリーズ名 読み
- 20 巻数



# 近刊情報の書誌項目 続

- 21 配本回数
- 22 セット商品分売可否
- 23 著者名1
- 24 著者名1 読み
- 25 著者名1 区分
- 26 著者略歴(紹介) 1。
  - 著者名2
  - 著者名2 読み 全角カタカナ
  - 著者名2 区分 著、編、訳、監修等
  - 著者略歴(紹介) 2

# 近刊情報の書誌項目 続

27 判型

28 判型(実寸)

# EDItEUR 概要

- ◆ 1991年設立
- ◆ EDIFACT (国連のEDI標準)を出版流通の世界で実装, 普及を目的。
- ◆ 名称はEDI to Europeを意識してつけられた。英国中心に主としてヨーロッパで活動。
- ◆ 会員企業は, 書籍流通, 図書館, システムベンダ等数十社, 日本からはJPO, 丸善, 紀伊国屋
- ◆ 2006-04 International ISBN Agency

# 活動

## ◆ 識別子

- ❖ International ISBN Agency (ISBN RA)
- ❖ ISTC
- ❖ 研究・提言 component identification等

## ◆ 出版流通における標準化

- ❖ ONIX
- ❖ ICEDIS
- ❖ Rights Management standards
- ❖ EDI

# 活動 projects

- ◆ ヨーロッパのいくつかの活動に参加
  - ❖ ARROW: Accessible Registries of Rights Information and Orphan Works towards Europeana
  - ❖ Linked Heritage
- ◆ 出版流通におけるRFID: ISO 28560-
- ◆ 電子出版EPUB (IDPF)への意見注入など

# 組織

- ◆ 専任3名 (2.5 full time)
- ◆ Projectや, 標準に対応して, Expert 5名, Consultant 3名を契約ベースで雇用
- ◆ 迅速, 柔軟な対応が可能

# 書誌コントロール機関のタスク

2011-10-22

情報組織化研究グループ

39

# 書誌コントロール機関

- ◆ 分野に応じた組織化モデルの作成。

文学作品の類であれば、work, expressionが必要となるだろう。

音楽作品では、編曲による異expressionと、演奏による異expressionの区分のような概念が必要となるかもしれない。

記録類のようにworkやexpressionの無意味な世界もある。

絵画作品のように、workをどのように捉えるか自体が問題となる分野もある。

(=どのようなコロケーションが必要か?)



# 書誌コントロール機関 続

- ◆ モデルに応じた, manifestationより上位の概念部分の識別と関係づけを行う。
    - workやexpressionの区切りをきめ, manifestationを関係づける。
    - conceptやeventを識別し, work等を関係づける。
    - agent (個人や団体)を識別し, manifestation, work等を関係づける。
- (特にagentについては, 分野をこえた作成が必要)

# 書誌コントロール機関 続

- ◆ これらの上位概念について，識別子を管理し，メタデータと主に公開・サービスする。
- ◆ このレベルの情報は，国単位で完結するものではなく，国際的な協力が不可欠となるだろう。

# 今後の検討課題

# 登録機関について

- ◆ manifestation単位の識別を登録機関に任せておけるか? ← ISBN運用の問題点
- ◆ 著作権管理が進みつつはあるが, これらが著作登録機関としての役を果たせるようになるか?
- ◆ 登録機関による識別子の付与とメタデータの公開, という方式が社会的に受け入れられるか?

# 書誌コントロール機関について

- ◆ 分野の検討：メタデータ作成，利用の双方の面から必要。
- ◆ 分野に応じたモデルの検討。
- ◆ 収集・保存機関をもたない分野の問題。このような分野での知的資源の識別・記述に意味があるか？  
(アーカイブされないネットワーク上の資源など)
- ◆ 収集・保存機関に入るまで，識別・記述されない知的資源の存在。