

国立情報学研究所のメタデータ共同構築計画

米澤誠

国立情報学研究所

〒 101-8430 東京都千代田区一ツ橋 2-1-2

Tel : 03-4212-2355

Fax : 03-4212-2375

E-Mail : yonezawa@nii.ac.jp

概要

国立情報学研究所では、全国の大学等の機関で発信するネットワーク上の情報資源に関するメタデータの共同構築を計画している。これは、既に個々の大学図書館等で行っているメタデータの作成を継承・集約するもので、日本国内で発信するネットワーク情報資源のうち、国立情報学研究所の目録所在情報サービス等を通じて連携協力している機関において、機関内で発信する情報資源のメタデータを網羅的に収集することを目指している。

2002 年前半にはメタデータ共同構築システムプロトタイプを開発し、試行運用を開始する。2002 年後半にはプロトタイプの改善を行い、本格運用を開始する予定である。

キーワード

メタデータ, ネットワーク上の情報資源, 共同分担目録, Dublin Core

1. はじめに

インターネット上に多種多様な学術情報資源が増加している状況で、それらの情報資源に対して効率的・有効なアクセスを保証するために、メタデータの組織化が必要とされている。国立情報学研究所（以下「NII」という）としては、従来の目録所在情報サービス（NACSIS-CAT/ILL）の枠組みの中で、まず、電子ジャーナルを総合目録データベースに収録することを可能とした。[1] しかしながら、これだけでは多種多様なネットワーク上の情報資源を組織化するには不十分であった。

一方、既にいくつかの大学図書館等では、サブジェクトゲートウェイ、リンク集などでのメタデータの収集及び提供を実現している。しかしながら、今後、情報資源の増加に応じて、そのメタデータの組織化には非常に大きな労力が必要となり、単独の機関で維持して行くことは困難になると予想される。

以上の状況を踏まえ、NIIでは2001年度に図書館情報学関係、図書館関係の有識者による検討会議を行った。検討会議では、米国OCLCのメタデータ共同目録であるCORCや英国のBUBL LINKなどの先行事例を参考にしつつ、主に共同構築の意義と役割について審議を行った。その結果、以上の問題を解決するために、複数機関の協力によるメタデータの共同構築が有効であるとの結論となり、以下の計画を立案した。[2]

2. 本計画の概要

本計画では、大学図書館等による共同分担構築方式をとる。分担の仕方の基本は、各機関が機関内で発信している情報資源のメタデータを構築するというものである。例えば、次のような情報資源が収録対象となる。

- 電子ジャーナル、論文、プレプリント、解説書、用語解説、シラバス、貴重書データ（画像、テキスト）、統計、音楽資料、音声データ、データ集、データベース、OPAC、目録、等

これにより構築されるメタデータ群を、Set. Aと呼ぶ。このような方式によるメタデータの共同分担構築は、あまり前例のないものであろう。NIIでは、同様の方式により「学術雑誌目次速報データベース」を構築している。[3] これは、主に大学の紀要等の目次データ（TOC：Table Of Contents）を収録するもので、各大学から目次データの提供を受けて構築するという方式をとっている。これには、大学等397機関が参加しており、3,073タイトルに関して約32万件の記事データを収録するという実績をあげている（2002年2月2日現在）。

Set. A以外に、自機関以外のものであっても、任意の機関グループで、分野毎のメタデータを構築することも可能とする。様々な分野毎のサブジェクトゲートウェイに相当する機能を持つことができる。これはCORCのとっている方式であるが、例えば経済学関係の機関で共同作成している「経済学文献索引データベース」などがこのモデルとなる。（図1）

Set. A, Set. Bともに、参加は任意である。計画当初は、Set. Aでのメタデータ構築の普及を目指し、並行して、Set. Bの構築の実現を検討して行くことになる。

メタデータ構築のためのシステムは、従来の目録所在情報サービス（NACSIS-CAT/ILL）とは別に、新たに用意する。データ記述については、従来型の目録規則ではなく、Dublin Coreに従った記述要素を採用することとした。また、構築されたメタデータを利用者に提供するための検索システムは、現在計画中のNII学術コンテンツポータルシステム（仮称）で実現することになる。（図2）

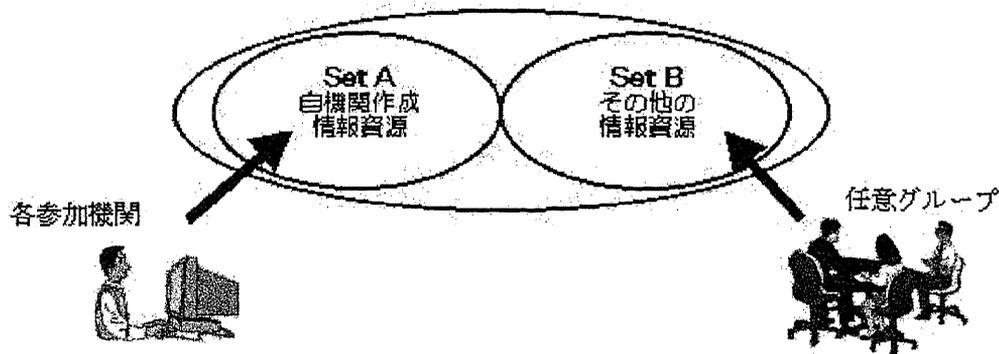


図 1

3. メタデータ共同構築の意義と役割

ネットワーク上の情報資源を検索するためには、Google 等のサーチエンジンは確かに有用であるが、教育研究上必要な情報が必ずしも効率的に検索できるわけではない。収録するメタデータが研究上有用な情報資源に限られていれば、利用者は多くの情報から選別する手間を省くことができる。また、収録するメタデータに的確な主題情報をもたせることにより、利用者は効率的な検索を行うことができる。これらの点に、ロボットで収集したサーチエンジンのメタデータとは別に、サブジェクトゲートウェイや本計画により、「人間の手を介した」メタデータを構築することの意味がある。

また、本計画により、日本国内の各機関で行っている情報発信を支援することになる。既にいくつかの大学図書館では、大学内で出版される学術資料に関して、情報の収集と提供活動を行っている。大学図書館が、大学の Institutional Server の役割を担っている訳である。各機関できめこまかな情報収集を行うことにより、ロボットでは収集できない、より深い位置に隠れている有用な情報資源が日の目を見るというメリットがある。また、各機関が責任をもって機関内の情報発信を行うという制度により、継続的なメタデータ構築が保証されることになる。

4. 構築システム

4.1 基本機能（プロトタイプ）

構築システムの基本機能として、共同分担機関が利用する Web ベースでの検索・登録インターフェースを開発している。メタデータ構築を分担する機関は、このインターフェースを使って、次のような手順でメタデータを作成する。

- ネットワーク上の情報資源を確認した後、URL 等でデータベースの検索を行う。この時、既にレコードが登録されていないかどうか、重複チェックを行う。
- 情報資源自体の情報（ページソース）から、ある程度メタデータの自動作成を行い、データ入力の省力化を図る。
- Creator, Publisher 等については、名称を統一するため、NACSIS-CAT の著者名典拠ファイル等を参照利用できるようにする。
- 図書館システム及び Website 等でのメタデータ再利用のために、各種データ形式でのダウンロードを可能とする。

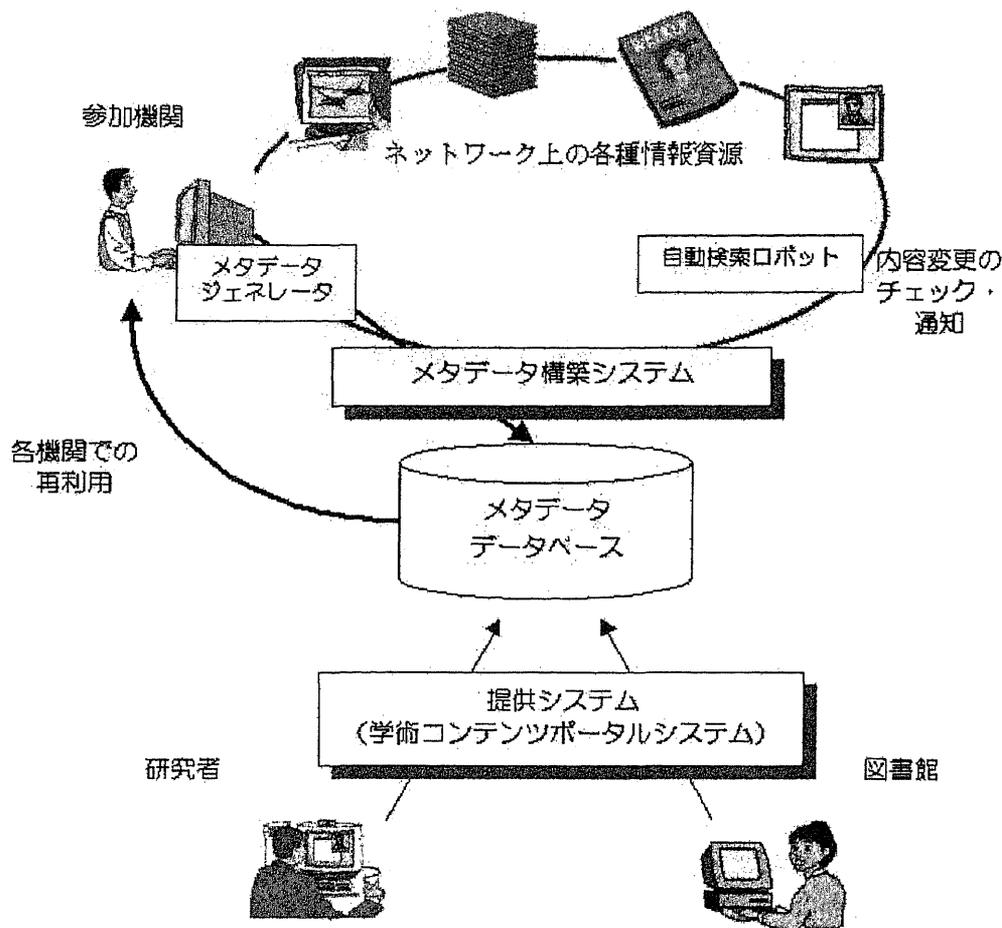


図 2

検索機能については、このシステムが構築を目的とするものであることから、限定的なものとしている。一般利用者のための様々な検索機能は、前述の学術コンテンツポータルシステム側で実現することとなる。

最後の、図書館側でのメタデータ再利用には、図書館 OPAC にネットワーク上の情報資源のメタデータを収録する、もしくはメタデータとリンクするというアイデアが考えられる。多くの図書館では、電子ジャーナルの提供は OPAC とは別のリンク集などである場合が多いが、OPAC と連携することにより、アクセスルートを一元化することができる。電子ジャーナル以外にも、ネットワーク上の有用な参考ツールを、OPAC 経由で利用者に提供することも考えられる。

このほか、メタデータからのアクセスを保証するために、自動巡回ソフトウェアにより定期的に URL をチェックし、URL が変わった場合に各作成機関に通知する機能をもつ。また、既に各機関で作成しているメタデータを一括アップロードするために、既存メタデータからの一括変換を行う機能も用意する。

以上の基本機能については、2001 年度に開発を行っている。

4.2 拡張機能

2002 年度以降には、以下のようなメタデータ構築を支援する拡張機能の実現を検討する。

- 分類・件名データベースによる、データ作成補助機能
- ハーベスティング・プロトコルによるデータ収集・蓄積機能
- メタデータジェネレータ機能

5. メタデータ記述要素

メタデータ記述要素については、Dublin Core に準拠するという方針で、CORC, EU の DESIRE プロジェクト, 国立国会図書館 WARP (仮称) 等の実装状況を踏まえて、検討を進めている。ただし、日本の情報資源を取扱うために、Dublin Core Qualifiers もしくは Library Application Profile の定義に対して、修正の要望を提示する必要があるかもしれない。従来、日本の図書館の現場から、国際的な規格としての Dublin Core に対しての要望が少なかったようである。本計画が、今後、日本の機関からの意見・要望を DCMI 側に伝えるための受皿として機能できないものかと考えている。

適切な検索を実現するために、分類、件名等の主題情報の付与は必須とする予定である。どのような分類表、件名表を採用するかについては、現在検討中である。

著者名の形をある程度正規化するため、NACSIS-CAT の著者名典拠レコードの標目形を基準とする。個人名については、NII の研究者ディレクトリの研究者氏名を基準とすることも検討している。著者名については、それぞれのデータベースとの厳密なリンク関係付けは行わない。また、典拠の標目形に従うことを必須としない。どのような著者名で記録するか迷った場合の、判断材料として利用する程度の「ゆるやかな典拠コントロール」とする。冊子体資料に比べて、メタデータにおける著者名検索の重要度は低い。必須とすることにより、典拠レコードの新規作成という労力を軽減する。

既存メタデータとの連携を可能とするためには、それぞれ NII メタデータとのマッピングを行っている。各既存メタデータとどのような形で連携するかについては、定型的な連携パターンを決めるのではなく、個別に対応を検討することとしたい。

6. 情報資源の評価基準

内外のゲートウェイサーバでは、それぞれネットワーク上の情報資源の評価基準を公開している。現在、それらの評価基準を比較検討し、各機関での情報資源の選択を容易にするための基準を検討している。情報資源の内容、著者の信頼性、情報の鮮度、ナビゲーション、デザイン等のポイントから、評価基準をとりまとめる予定である。

7. 今後のスケジュール

2001 年 4 月以降に構築システムプロトタイプによる、試行運用を開始する。この試行運用では、数機関の協力により、各機関で発信する情報資源 (Set.A) のメタデータを分担構築する。試行運用でプロトタイプの評価を行い、必要な改善を行った後、2002 年後半には本格運用を開始する予定である。

8. おわりに

NII のメタデータ共同構築計画は、ようやくスタートしたばかりである。現段階では内容未定の部分も多いが、海外でのメタデータ整備に立ち遅れることのないよう、多くの機関の協力を得て、計画を進めていきたい。

また、本計画は図書館だけをパートナーとするものではないが、伝統的に、カタログिंगという情報の組織化を専門としていた図書館（図書館員）の果たす役割に、期待するところは大きい。既に世間的に大きな評価を得ている Webcat のように、日本の図書館員の手による有用なデータベース構築を実現したいと考える。多くの機関の賛同と御支援をいただきたい。

参考文献

- [1] 総合目録データベースにおける電子ジャーナルの取扱い（暫定案）. NACSIS-CAT/ILL ニュースレター, 1号, 2000, <http://www.nii.ac.jp/CAT-ILL/PUB/nl2/No1/contents.htm>
- [2] ネットワーク上の情報資源の取扱いの検討状況. NACSIS-CAT/ILL ニュースレター, 5号, 2001, <http://www.nii.ac.jp/CAT-ILL/PUB/nl2/No5/index.htm>
- [3] 学術雑誌目次速報データベース. <http://www.nii.ac.jp/sokuho/>