

メタデータの相互運用等に関して

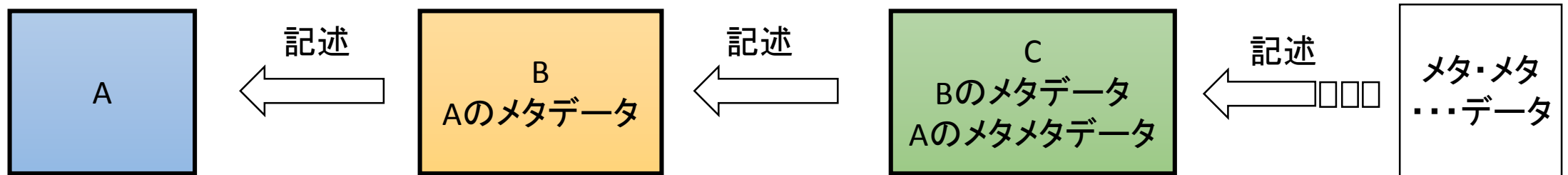
杉本重雄
筑波大学・図書館情報メディア系
sugimoto@slis.tsukuba.ac.jp

内容

1. メタデータの基本概念
2. インターネット上でのメタデータの相互利用の形態
3. まとめ

メタデータの基本概念

- メタデータとは「データに関するデータ(data about data)」、「何らかの対象に関する記述」



- 本の著者やタイトルなどの書誌情報は本のメタデータ
- 書誌情報を書いた人や書き換えた日付は書誌情報のメタデータ
- 書誌情報の書き方を決めた規則は書誌情報のメタデータ
- 書誌情報のメタデータは、本から見ればメタメタデータ
- メタデータの相互運用にはメタメタデータが重要な役割を持つ

メタデータの基本概念

- メタデータの基本モデルとしての1対1原則 (1-to-1 principle)

→ 1個の記述対象に対して1個のメタデータ

例：本の著者と書誌情報の作成者を一緒に入れたメタデータは非1対1

→ 何に関するメタデータなのかを明確化するための基本

- メタデータのスキーマ(記述規則、記述形式)を決める際には、記述対象を明確にとらえた、メタデータのためのデータモデルが必須

→メタデータの設計上の基本であり、相互運用性を考える上での基本

→何を記述したメタデータであるのかが明確でなければ、たとえ他のメタデータと共通記述項目(たとえば、Dublin CoreのCreatorやTitle)を用いていたとして、記述項目名によるメタデータ同士の関係付けは困難

インターネット上でのメタデータ相互利用形態

- アーカイブ間の相互結合（ポータルと特定アーカイブ間結合）
 - ポータル： 何らかのドメインにおける多数のアーカイブへの入り口
 - 特定サービス間の相互結合： プロトコルを共有するアーカイブ同士をつないで互いにサービスを利用できるようにする
- 複数のアーカイブを結ぶための技術
 - 横断検索： 複数のデータベースを横断的に検索するプロトコルを使用して横断検索サービスを実現する（プロトコル SRU/SRW）
 - メタデータハーベスティング（メタデータ収集）： 複数のデータベースからメタデータを収集するためのプロトコルを使用してメタデータを収集し、収集側ではメタデータを加工処理してサービスを提供できる（プロトコル OAI-PMH）
 - Linked Open Dataデータセット（SPARQL endpoint）： RDF形式で構成されたデータのデータベースに対する検索の標準プロトコルを使用して、横断的な検索やメタデータの収集を行い、サービスを実現する（プロトコル SPARQL）

インターネット上でのメタデータ相互利用形態

- サービスの作り方の進歩
 - 出現してきた時間順序： 横断検索⇒メタデータハーベスティング⇒LOD
 - 後の方ほどメタデータの自由度は高く、異種アーカイブをつなぐのに適する
 - 逆に見れば、メタデータの記述項目定義を共有する仕組みが必要
⇒ URIによる記述項目の識別とその基盤ができてきた
- メタデータを集めて利用する際の課題
 - 同じ(とみなせる)対象に対して、複数のメタデータが作られることは稀ではない
 - Europeanaのデータモデル(EDM)は、同一文化財に関する、異なる組織から集められたデジタルコンテンツのメタデータをひとまとめに結び付ける仕組みを持つ
 - メタデータを集めるだけでなく、結び付け、集約する仕組みが必要
⇒ metadata aggregation

まとめ

- 基本モデルがしっかりしているメタデータはつなぎやすい
- LOD化の潮流： Web上で多様な種類のリソースを結び付ける技術基盤ができてきたことはデジタルアーカイブの利活用性向上にとって重要
 - リソースには、デジタル実体のみならず、物理的な実体、概念の実体を含む。Wikipedia等の辞書、件名標目や十進分類などは重要な概念のリソース
 - 提供者はインターネット上での第3者による利用を意識する必要があるが、具体的な利用方法を最初から意識することは容易ではない
- これまでの経験から
 - 技術的には、メタデータ情報を共有し、メタデータの流通を支える情報基盤が必要
 - 社会的には、デジタルアーカイブの利用を支援する組織的な仕組みが必要
 - 多様な利用者やコミュニティがあり、提供者視点だけでは解決不能な問題が多い