

デジタルアーカイブと その社会的活用

第8回

立命館大学 映像学部講義
福田一史

<https://scrapbox.io/fukudakz/21デジタルアーカイブとその社会的活用>



manabaRにもリンクがあります

コンテンツ

1. メタデータ（つづき）
2. メタデータ作成フロー

メタデータ

リソースの記述

参考資料

- メタデータ基盤協議会. (2012). “メタデータ情報共有のためのガイドライン”
 - EPUB版／Kindle版
<http://logical-web.com/logibooks/metaguide.html>
 - **PDF版**
<http://www.mi3.or.jp/item/A03.pdf>
 - 本講義ではPDFを参照する

メタデータ語彙

- メタデータ記述に**共通的に用いること**を目的として多くのメタデータ語彙（Metadata Vocabulary）が公開されている
- メタデータ語彙として公開される各用語は**URIを持つ**ため、**Linked Open Dataに用いるのに適している**
 - e.g. <http://schema.org/name>, <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
- 近しい分野や目的にそったメタデータ語彙を用いることで、他のデータセットとの接続が容易に
- 記述にあたっては、一つのメタデータ語彙では不足することもあるため、複数のメタデータ語彙を併用することも多い

RDF語彙

- RDFとRDFスキーマ（RDFS）は、リソースの種別の指定、プロパティ、リソース間の関係を定義するための基本的な機能を提供する
- RDFスキーマはクラスの概念を用いて資料を分類する。
- 解説としては下記などがある
 - <http://www.asahi-net.or.jp/~ax2s-kmtn/internet/rdf/NOTE-rdf11-primer-20140225.html>
 - <https://kanzaki.com/docs/sw/rdf-schema.html>

Dublin Core Metadata Element Set, V. 1.1

- <https://www.dublincore.org/specifications/dublin-core/dces/>
- 図書館コミュニティに端を発するリソースを記述するための語彙。DCMI（Dublin Core Metadata Initiative）により管理される。DCMESは「Simple Dublin Core」とも呼ばれる。
- [ISO 15836:2003](#)
- 15件の記述要素が定義される
 - Contributor, Coverage, Creator, Date, Description, Format, Identifier, Language, Publisher, Relation, Rights, Source, Subject, Title, Type

DCMI Metadata Terms

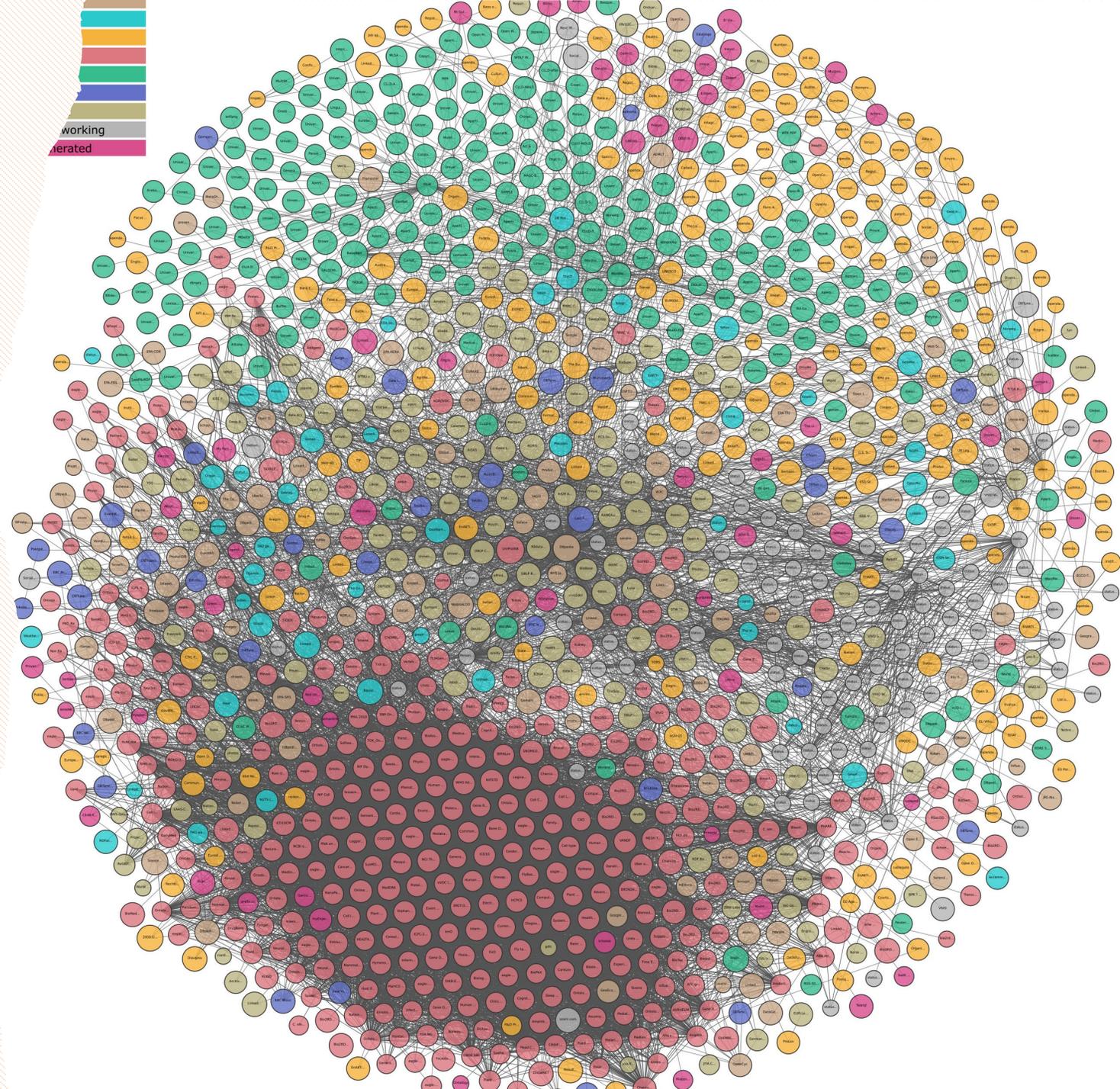
- <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>
- Simple Dublin Coreよりもより詳細な記述に用いることを想定し、拡張されたメタデータ語彙
- プロパティの他にクラスや種別 (Type) の語彙などを含む

Schema.org

- <https://schema.org/>
- 大手の**検索エンジンサービス運営者**ら（Google, Yahoo!, Microsoftなど）により開発されたメタデータ語彙
- 検索効率の向上を目的に、主に**ウェブページの運営者らによりマークアップしてもらうことを想定**して開発された

LOD: Linked Open Data

- ウェブをデータベースに作り変えようというプロジェクト
 - 「文書のウェブ」から「データのウェブ」へ
- The Linked Open Data Cloud. <https://lod-cloud.net/>



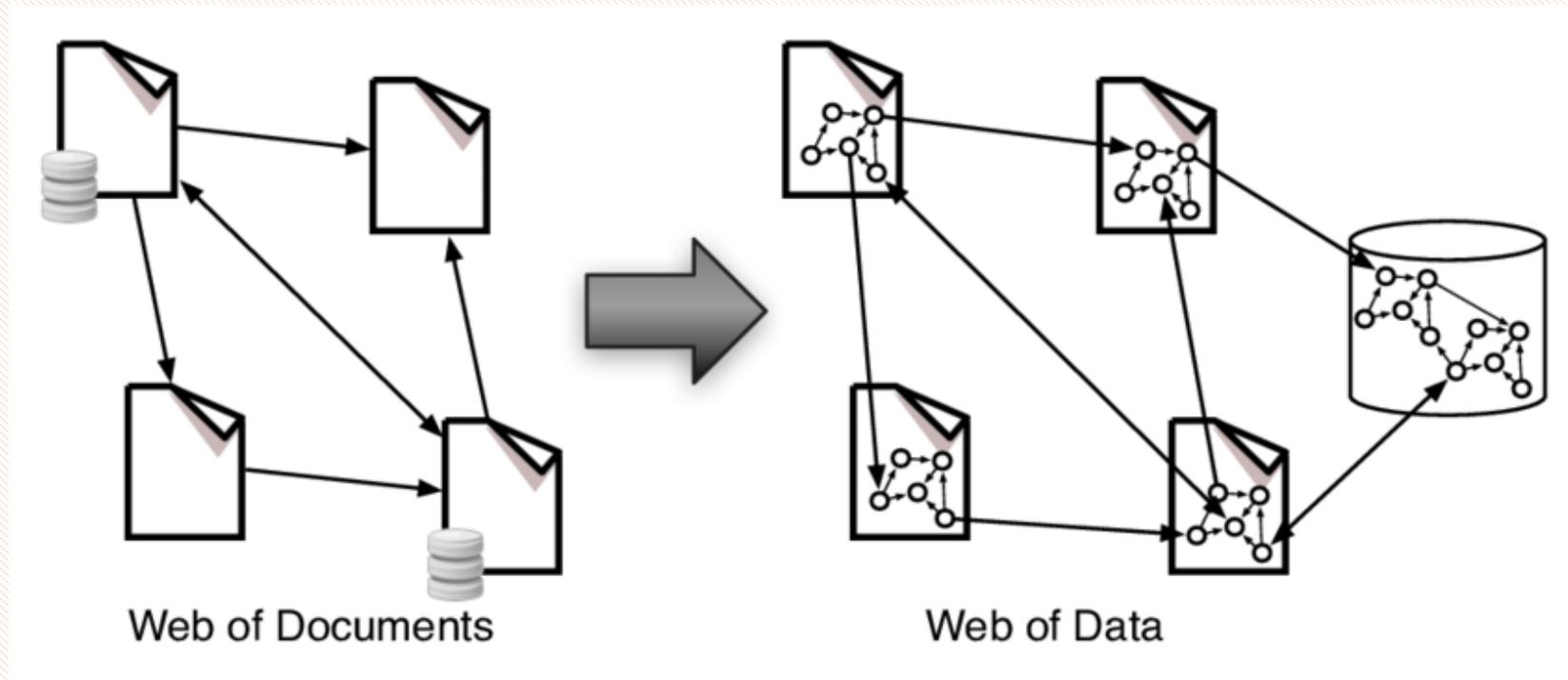


図. 文書のウェブからデータのウェブへ

Emanuel Berndl. (2019). "Embedding a Multimedia Metadata Model into a Workflow-driven Environment Using Idiomatic Semantic Web Technologies". Thesis for Doctoral.

https://www.researchgate.net/publication/334784695_Embedding_a_Multimedia_Metadata_Model_into_a_Workflow-driven_Environment_Using_Idiomatic_Semantic_Web_Technologies

LOD: Linked Open Data

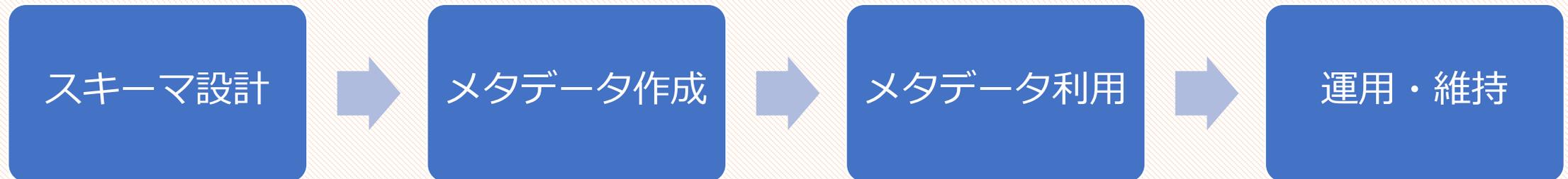
- Linked Data - Design Issues
 - <https://www.w3.org/DesignIssues/LinkedData.html>
- Linked Dataの条件（5 star linked data）
 1. あらゆるモノのIDに**URI**を用いる
 2. IDに**HTTP URI**を用いる
 3. URIにアクセスすると**標準**（RDF, SPARQL）を用いた**構造化データ**が得られる
 4. 構造化データには**他のモノのURI（外部リンク）**を含む
- Tim Barners-Leeが示す次のウェブ（Youtube - TED）
 - https://www.youtube.com/watch?v=OM6XIIcm_qo

メタデータ作成フロー

メタデータ的设计・作成・公开・维持

メタデータのライフサイクル

- メタデータは下記のフローで生成・利用される。
- 一貫して共通性や標準性を検討する必要がある。



スキーマの設計

- スキーマの選択・設計と公開の指針（メタデータ情報基盤構築事業 2011, p. 9） ※括弧内のA, B, Cは優先度
 1. スキーマを相互運用可能な形で選択・設計する（A）
 2. 新たなスキーマを設計する場合、仕様語彙や参照記述規則の定義を尊重し、メタデータを相互運用できるように設計する（A）
 3. 独自スキーマを設計する場合も、特定領域の知識なしに理解し、交換可能なフォーマットに変換するための情報・規則を用意する（B）
 4. スキーマ定義を、コンピュータ処理可能な標準方法でも表現し、公開する（C）

メタデータの仕様

- メタデータの仕様を共通化するための方式として、DSP（記述セットプロファイル）がある。ここでは簡易DSPについて紹介する。
- 簡易DSPは、記述規則ブロックと名前空間宣言ブロックで構成される。

メタデータの仕様

- メタデータの仕様を策定・公開するための方式として、DSP（記述セットプロファイル）がある。ここでは簡易DSPについて紹介する。
- 簡易DSPは、記述規則ブロックと名前空間宣言ブロックで構成され、タブ区切りファイル（.tsv）で記述される（次々ページ）。

メタデータの仕様

- 項目記述規則

- 項目規則名、プロパティの修飾名、最小出現回数、最大出現回数、値タイプ、値制約、コメント（説明）、といった要素で構成される。

項目規則名	プロパティの修飾名	最小出現回数	最大出現回数	値タイプ	値制約	コメント
タイトル	schema:name	1	1	文字列		リソースのタイトル／名前
ISBN	schema:isbn	0	1	文字列		リソースのISBN
発行日	schema:datePublished	1	1	文字列	dcterms:W3CDTF	リソースが発行された日付
著者	schema:creator	0	-	構造化	foaf:Agent	リソースの作者
出版者	schema:publisher	0	-	構造化	foaf:Agent	リソースの公開者

[@ns]

schema: <http://schema.org/>

[Book]

#項目規則名	プロパティの修飾名	最小出現回数	最大出現回数	値タイプ	値制約	コメント
タイトル	schema:name	1	1	文字列		リソースのタイトル/名前
ISBN	schema:isbn0	1		文字列		リソースのISBN
発行日	schema:datePublished	1	1	文字列	dcterms:W3CDTF	リソースが発行された日付
著者	schema:creator	0	-	構造化	foaf:Agent	リソースの作者
出版者	schema:publisher	0	-	構造化	foaf:Agent	リソースの公開者

メタデータ記述

- メタデータ記述の推奨指針（メタデータ情報基盤構築事業 2011, p. 9） ※括弧内のA, B, Cは優先度
 1. リソースにグローバルな識別子（**URI**）を与える（A）
 2. **人間に理解可能なラベル**を標準的な方法で与える（A）
 3. 標準的で再利用可能な形で、コンテンツの作者を記述する（B）
 4. 曖昧さのない**標準形式**で日時、位置情報を付与する（B）
 5. 可能ならばキーワードを統制語彙で付与する（B）
 6. ラベルに読みを与える場合は、**言語タグ**を用いて区別するか、ラベルを構造化して記述する（C）
 7. リテラル値のデータ型、言語タグは、目的が明確な場合に限り、スキーマで仕様を宣言して一貫した形で与える（C）

ISO8601

- 日付と時刻の標準形式
- 年月日の順番は地域により違う場合がある。例えばアメリカでは月-日-年の順番で日付を記す場合が多く、日本で広く用いられる形式と違う。
 - e.g. June 19, 2021
- 年月日を「YYYYMMDD」（基本方式）や「YYYY-MM-DD」（拡張方式）により記述する。視認性や年の明記のため拡張方式が頻繁にもちいられる。
 - e.g. 2028-01-19
- その他に時刻や期間の指定などが定義されている。
- ref. https://ja.wikipedia.org/wiki/ISO_8601

メタデータの公開

- メタデータの公開と交換・利用に関する指針（メタデータ情報基盤構築事業 2011, p. 9） ※括弧内のA, B, Cは優先度
 1. メタデータの公開には、**標準的なデータ形式としてRDFを用いる** (A)
 2. メタデータを**正しく理解・利用するためにスキーマを参照し**、必要に応じてプロパティの整合調整を行う (B)
 3. データを公開用などに変換する場合は、情報が失われないように構造と粒度を保ち、利用者がダムダウンする。主要プロパティはあらかじめ単純化値を提供する (B)

メタデータの運用

- 運用に関する指針（メタデータ情報基盤構築事業 2011, p. 9-10） ※括弧内のA, B, Cは優先度
 1. スキーマの管理データを明示し、**バージョン管理**を行う（A）
 2. メタデータには作者、作成日時、準拠スキーマなどの**管理データを付与**する（A）
 3. データを集約して格納する場合、由来情報とあわせて管理する（B）
 4. スキーマを**公開レジストリに登録**し、利用者の発見を助けるとともに、**最新版、旧版を確認できるようにする**（B）
 5. メタデータを作成・公開する場合、スキーマの記述規則と矛盾がないか検証する（C）