

デジタルアーカイブと その社会的活用

第7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

評価方法の変更について：再度のお知らせ

- 評価方法の「**定期試験**（50%）」が「**レポート試験**（50%）」に変更になりました。
 - 論題発表：6/22（火）
 - 提出締切：7/22（木）
- コロナ渦の影響で学生が集まって試験をするという形式が難しいということによる方法の変更になります。

コンテンツ

1. メタデータ (つづき)

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

メタデータサンプル

<コーヒー>

<名前>GEORGIA 香るブラック</名前>

<内容量>260 ml</内容量>

<原材料名>コーヒー / 香料</原材料名>

<賞味期限>2022年3月</賞味期限>

<販売者>コカ・コーラ カスタマーケティング株式会社
</販売者>

</コーヒー>



メタデータサンプル

- ペットボトルのラベルの情報が「メタデータ」。
- その表記方法が「メタデータスキーマ」。

メタデータとメタデータスキーマ

- メタデータはあらゆるリソースに記述（Description）され、ユーザはそれを通じてリソースの情報を得ている。
- しかし、その形式は種別ごとに多様。
 - 例えば、パフォーミングアートとおもちゃと清涼飲料水には、それぞれ違うスキーマが必要とされる。

メタデータとは

- デジタルコンテンツを提供するための情報基盤構築が進んでいるが、現時点ではGoogleなどの全文検索によるアクセス提供が主である。
 - 全文検索は有用性が高いが、精密さなどには疑義が残る。
 - 曖昧・ざっくりとした情報要求には全文検索が適しているが、具体的な情報要求の対応には限界がある。
- より信頼できるアクセスを提供するためには、個別のデータベースの検索フォームなどを用いるしかないが、それらは**横断的に用いるには不足**がある。
- 複数のデータベースで用いられる、**一定の規則に従って記述されたメタデータが必要**。

メタデータとは

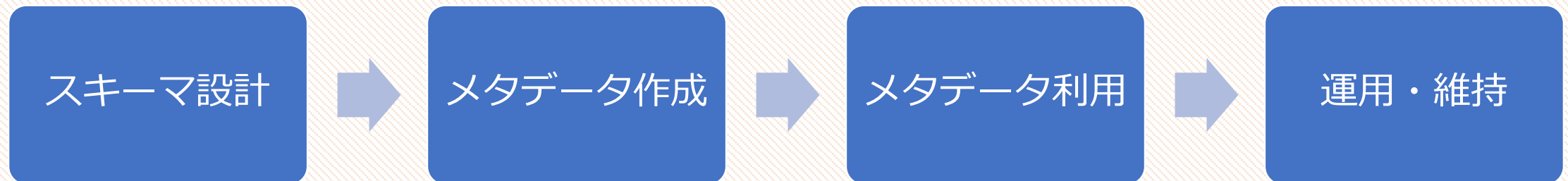
- メタデータの、「メタ」は「高次の」などを意味する接頭語であり、その接頭語を含む用語としてメタ認知、メタゲーム、メタメッセージなどがある。メタデータとは「データに関する（構造化された）データ」という意味。
- 具体的には、メタデータとは「**記述対象となる情報資源に関して、決められた属性についてその属性値を書き表したもの**（メタデータ基盤協議会 2012, p. 6）」。

メタデータスキーマ

- ある対象のための記述項目や記述形式の組み合わせ。
- 目的や分野ニーズに基づき構成される。標準を策定することで、相互運用が可能になる。
- メタデータスキーマを構成する3つの要素
 - 項目の定義：項目の名前と意味、メタデータ語彙
 - 項目ごとの記述要件：必須、推奨、繰り返し可能数などの構造的制約や値・型の制約
 - 実現形式：ウェブで利用する形式の指定

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。

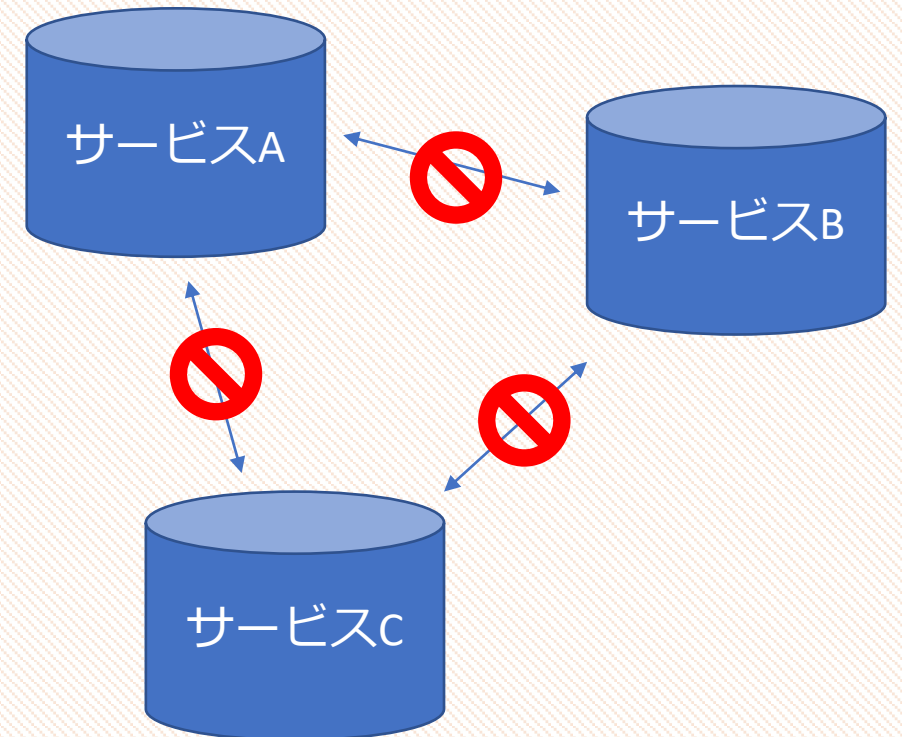


メタデータとウェブシステムの寿命

- ウェブのシステムは短命（5～10年）
- 一方、データセットは長く使われる（50年～）

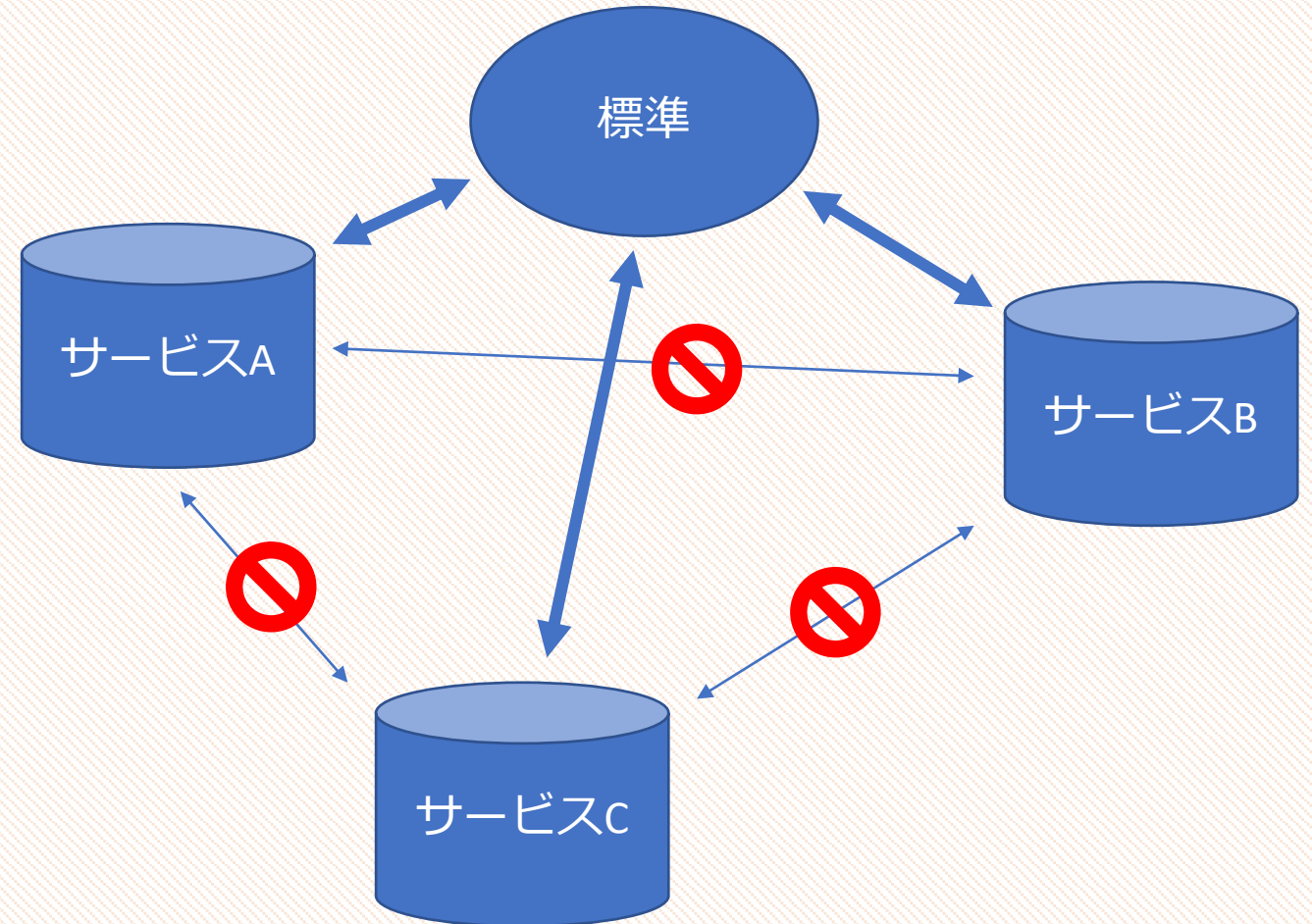
DBサービスの接続における問題

- 各サービス間のデータ接続が難しい
 - 記述項目・メタデータスキーマの違い
 - 記述単位の違い
 - ファイル形式の問題
 - など



標準を用いたデータ接続

- 技術やスキーマ（記述形式）の標準を策定することで、それを介して相互運用性のなかったDB間の**データ流通やデータ接続が円滑化する**。



メタデータ作成演習 その1

- 一人につき一つの「ペットボトル飲料」のメタデータを作ってください。
- 今から**15分**で、データを提出していただきます。
- 手元にペットボトルがあればそれについて記述しても、ウェブでデータを探してデータを作っても、構いません。
- 形式は、先ほどのコーヒーと同様のデータ形式です。＜ペットボトル＞～＜/ペットボトル＞として記述してください。作成したテキストをフォーム（いつものScrapboxの講義ページにリンクがあります）に登録してください。
- 記述項目はいくつ定義してOK。ただし、他の受講生が作るデータと「**つなげる**」ものとする（共通的に用いることができる）ことを意識してください。

＜ペットボトル＞

＜項目＞値＜/項目＞

＜項目＞値＜/項目＞

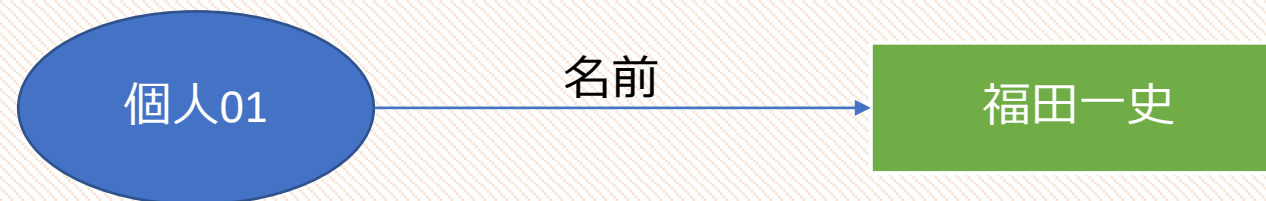
＜項目＞値＜/項目＞

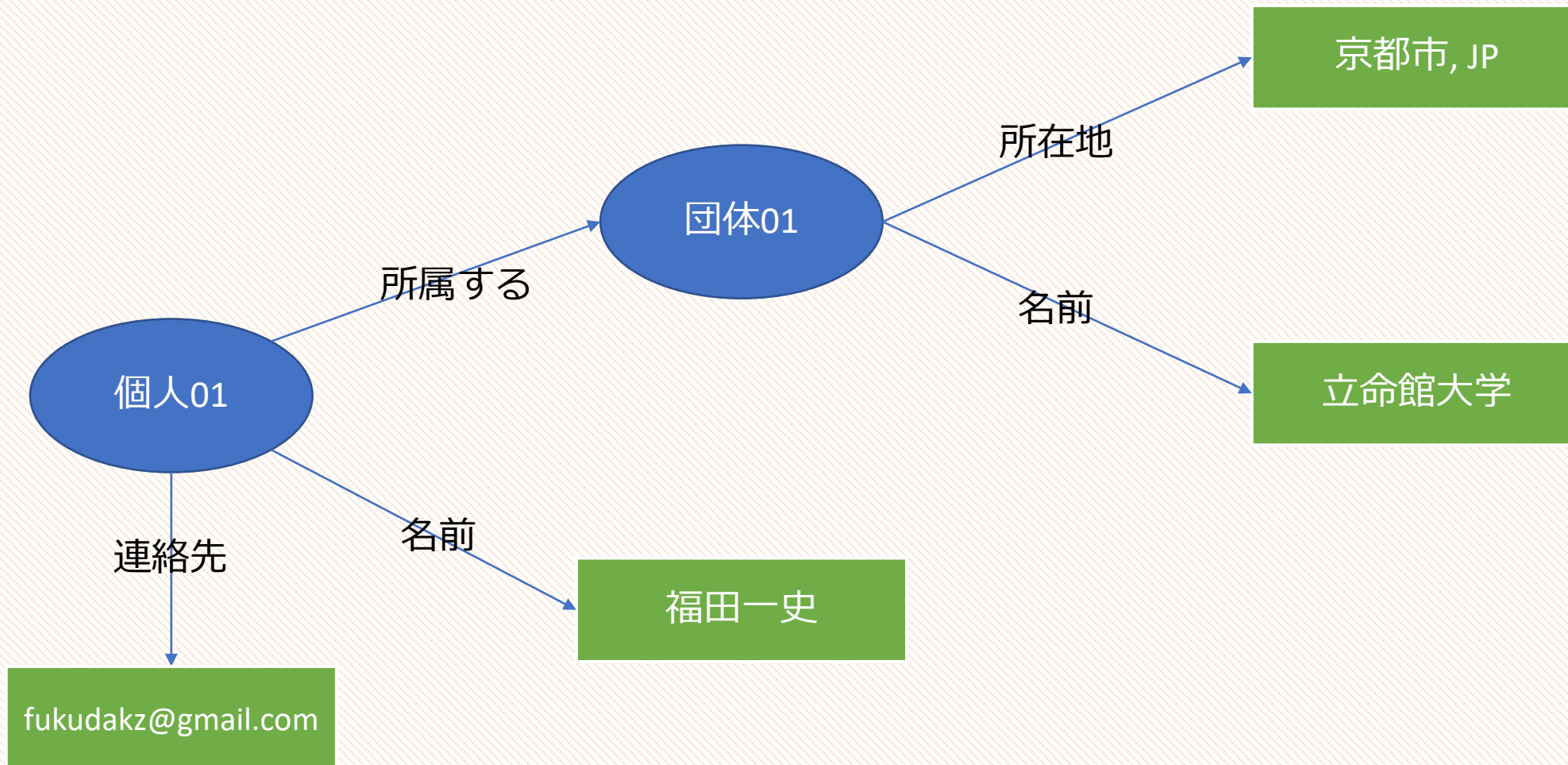
＜項目＞値＜/項目＞

＜/ペットボトル＞

RDF: Resource Description Framework

- ウェブでデータを交換するための標準モデル。メタデータの記述に用いられる。
 - <https://www.w3.org/RDF/>.
- RDFでは「**主語**、**述語**、**目的語**」の三つ組（**トリプル**）でデータを記述する。RDFで記述されるデータは全てグラフ（ネットワーク）で表現することができる。
 - 下図は、「個人01」（主語）の「名前」（述語）は「福田一史」（目的語）であるということを示している。





RDF: データ例

<個人01> <名前> "福田一史" .

<個人01> <連絡先> "fukudakz@gmail.com" .

<個人01> <所属する> <団体01> .

<団体01> <名前> "立命館大学" .

<団体01> <所在地> "京都市, JP" .

RDF: Resource Description Framework

- RDFは抽象構文（抽象的なデータ構造、データモデル）を定義するものであり、実際にデータをやりとりするための形式には複数の選択肢（シリアル化形式）がある。
 - N-Triples
 - **Turtle**
 - RDF/XML
 - JSON-LD
 - など

RDF: Resource Description Framework

- キーワード
 - リソース
 - 記述される対象物。資料、作品、個人、団体、概念など。
 - プロパティ
 - リソースの特徴の各個。ある図書のタイトル、作者、出版者、大きさ、など。
 - クラス
 - 共通の特徴を持つリソースのグループ。図書、ゲーム、版画、絵画、国、法律、スポーツ、スポーツチームなど。
- インスタンス
 - あるクラスの実体リソース。例えば、「カナダ」はクラス「国」のインスタンス、「サッカー」はクラス「スポーツ」のインスタンス。

RDF: Resource Description Framework

- 「RDF 1.1入門」 (日本語版)
 - <http://www.asahi-net.or.jp/~ax2s-kmttn/internet/rdf/NOTE-rdf11-primer-20140225.html#section-use-cases>.
- RDFのデータ変換
 - EasyRDF. <https://www.easyrdf.org/converter>
- RDFグラフの視覚化
 - The Web KANZAKI. <https://www.kanzaki.com/works/2009/pub/graph-draw>

URI: Uniform Resource Identifier

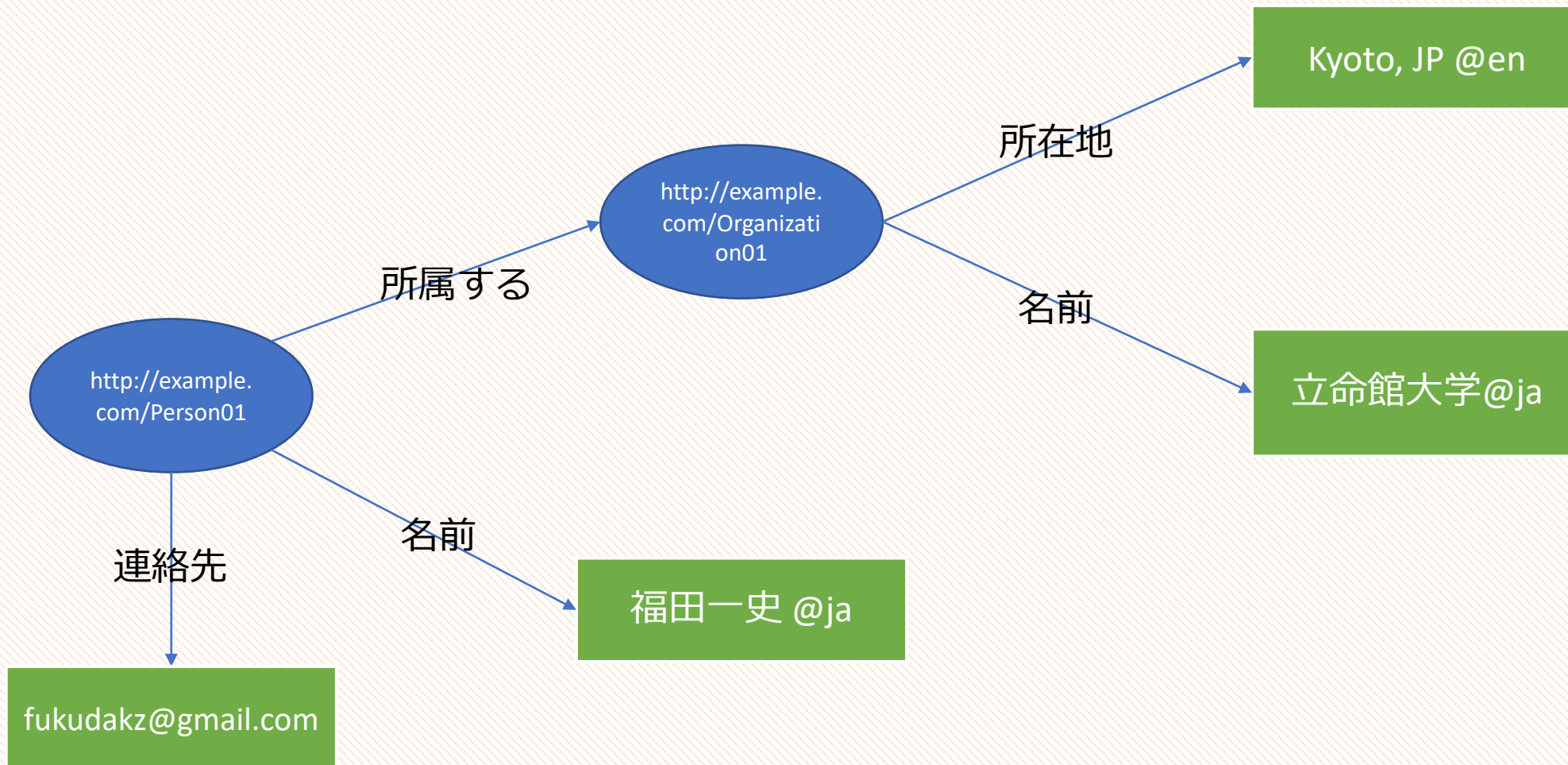
- **データ管理では、とにかく「ID」**（Identifier, 識別子）が重要。
- 個別のデータセットのID（ローカルなID）は、**ウェブで用いる場合他のデータセットと同時に用いる場合、重複が生じる可能性**がある。
 - e.g. 58, article1281, 00115627

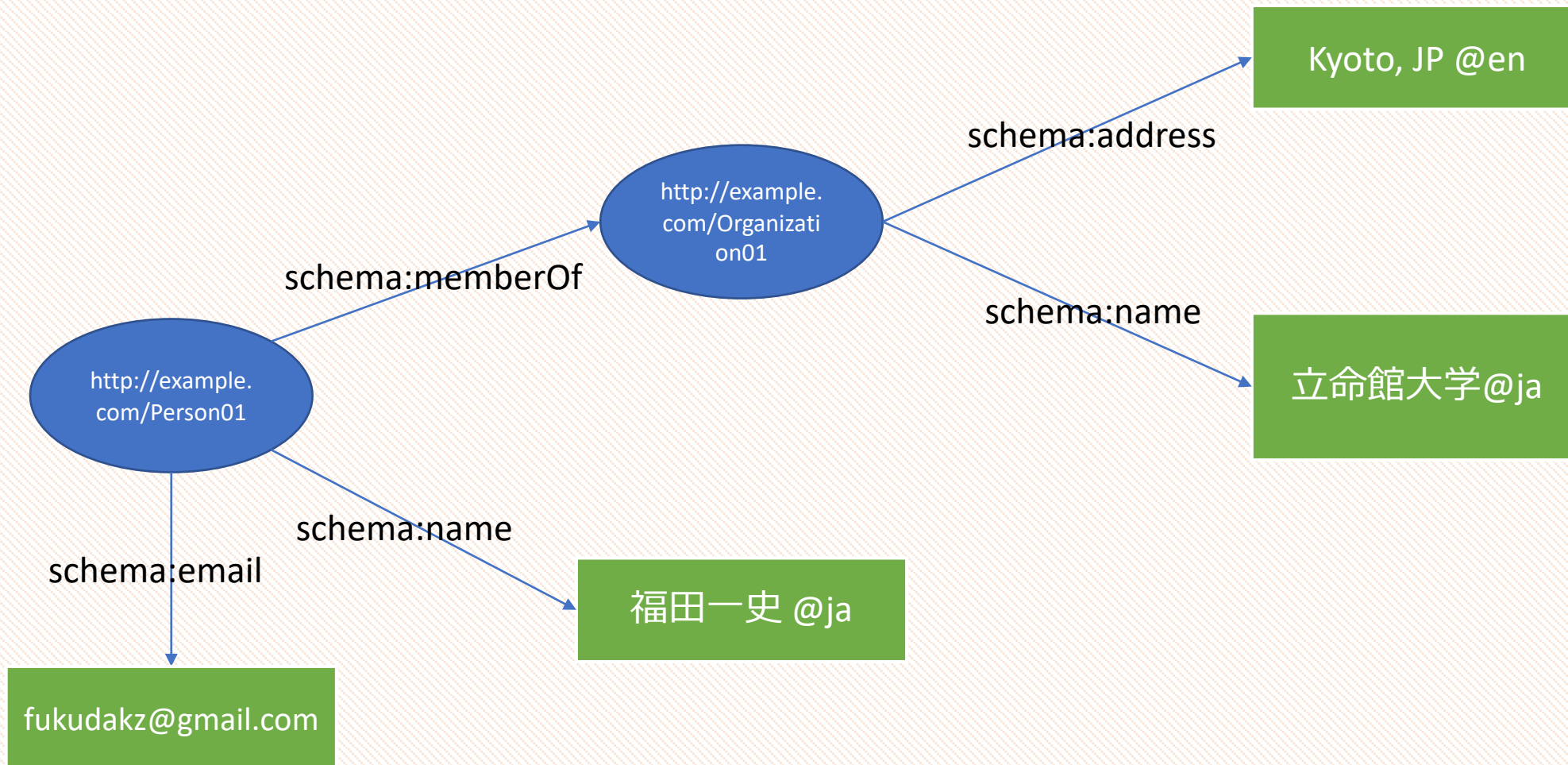
URI: Uniform Resource Identifier

- URIはグローバルな識別子。
- URLを識別子として用いることで、個別のリソースがグローバルに識別可能となる。
 - 機関毎にドメインがあり、それらの中でID体系があれば、グローバルな識別子に
 - アクセスもURIから提供できる
 - e.g. https://jpsearch.go.jp/item/tfam_art_db-1114,
 - e.g. <https://collection.rcgs.jp/resource/PACKAGE0022985>,
 - e.g. <https://id.ndl.go.jp/auth/ndlna/00292934>
 - ref. クールなURIは変わらない. <https://www.kanzaki.com/docs/Style/URI>
 - ref. Cool URIs for the Semantic Web. <https://www.w3.org/TR/cooluris/>
- RDFでは**リソースの識別にURIを用いる**（主語、述語、目的語のいずれにも適用可能）。

リテラル

- URIでない値
- **文字列、日付、数値**など
- RDFの**目的語にのみ出現**する
- RDFのグラフ表現では**四角**のノードで表される
- リテラルには言語タグを付与して言語を指定することが可能
 - 言語タグはISO639-1の言語2文字コードを用いることが推奨される
 - https://ja.wikipedia.org/wiki/ISO_639-1コード一覧





※ 「schema:」 は 「http://schema.org/」 の省略形

RDF: データ例 (N-Triples)

```
<http://example.com/Person01> <http://schema.org/name> "福田一史"@ja .
```

```
<http://example.com/Person01> <http://schema.org/email> "fukudakz@gmail.com" .
```

```
<http://example.com/Person01> <http://schema.org/memberOf> <http://example.com/Organization01> .
```

```
<http://example.com/Organization01> <http://schema.org/name> "立命館大学"@ja .
```

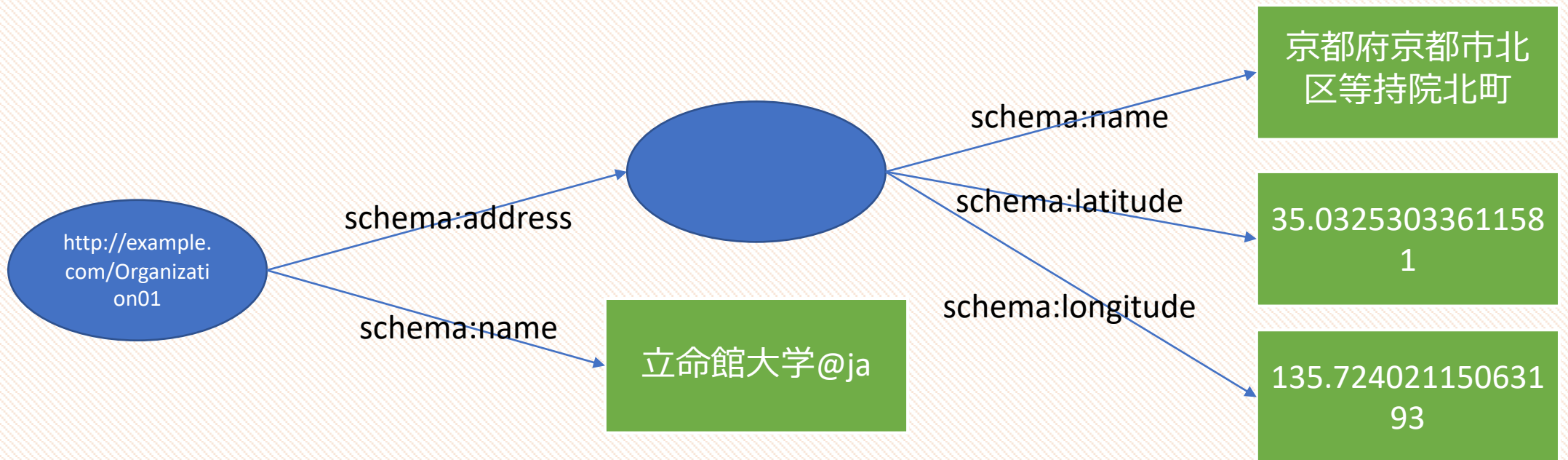
```
<http://example.com/Organization01> <http://schema.org/address> "Kyoto, JP"@en .
```

RDF: データ例 (Turtle)

```
<http://example.com/Person01>  
  <http://schema.org/name> "福田一史"@ja ;  
  <http://schema.org/email> "fukudakz@gmail.com" ;  
  <http://schema.org/memberOf> <http://example.com/Organization01> .  
<http://example.com/Organization01>  
  <http://schema.org/name> "立命館大学"@ja ;  
  <http://schema.org/address> "Kyoto, JP"@en .
```

RDF: 空白ノード

- RDFでは、IDを持たない（グローバルに識別する必要がない）ノードを経由して、データを記述することができる。ただし、空白ノードを連続して2つ経由することはできない。



メタデータ作成演習 その2

- 作成演習その1で作成したデータをRDF: Turtleのデータ形式で記述しなおし、そのテキストを提出する。主語は「drink-[学生証番号]」としてください（下のサンプルを参照のこと）。
- Turtleの仕様については [RDF 1.1 Turtle \(CyberLibrarian\)](#) を参照のこと。本データではリソースやプロパティをURIで記述する必要はありません。

```
<drink-1910064XXX>  
  <種別> "缶コーヒー";  
  <原材料> "コーヒー", "香料";  
  <名前> "GEORGIA 香るブラック";  
  <販売者> "コカ・コーラカスタマーマーケティング株式会社".
```


メタデータ作成演習 その2

- Scrapboxのリンクから、Googleフォームで提出してください。
- **提出は必須です！！**
- 提出前に、バリデータなどを用いて、Turtleのテキストデータが有効な形式であることを確認すること
 - <http://ttl.summerofcode.be/>
 - <https://www.kanzaki.com/works/2009/pub/graph-draw>

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/>
- Dublin Coreは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。DCMESは「Simple Dublin Core」とも呼ばれる。
- 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

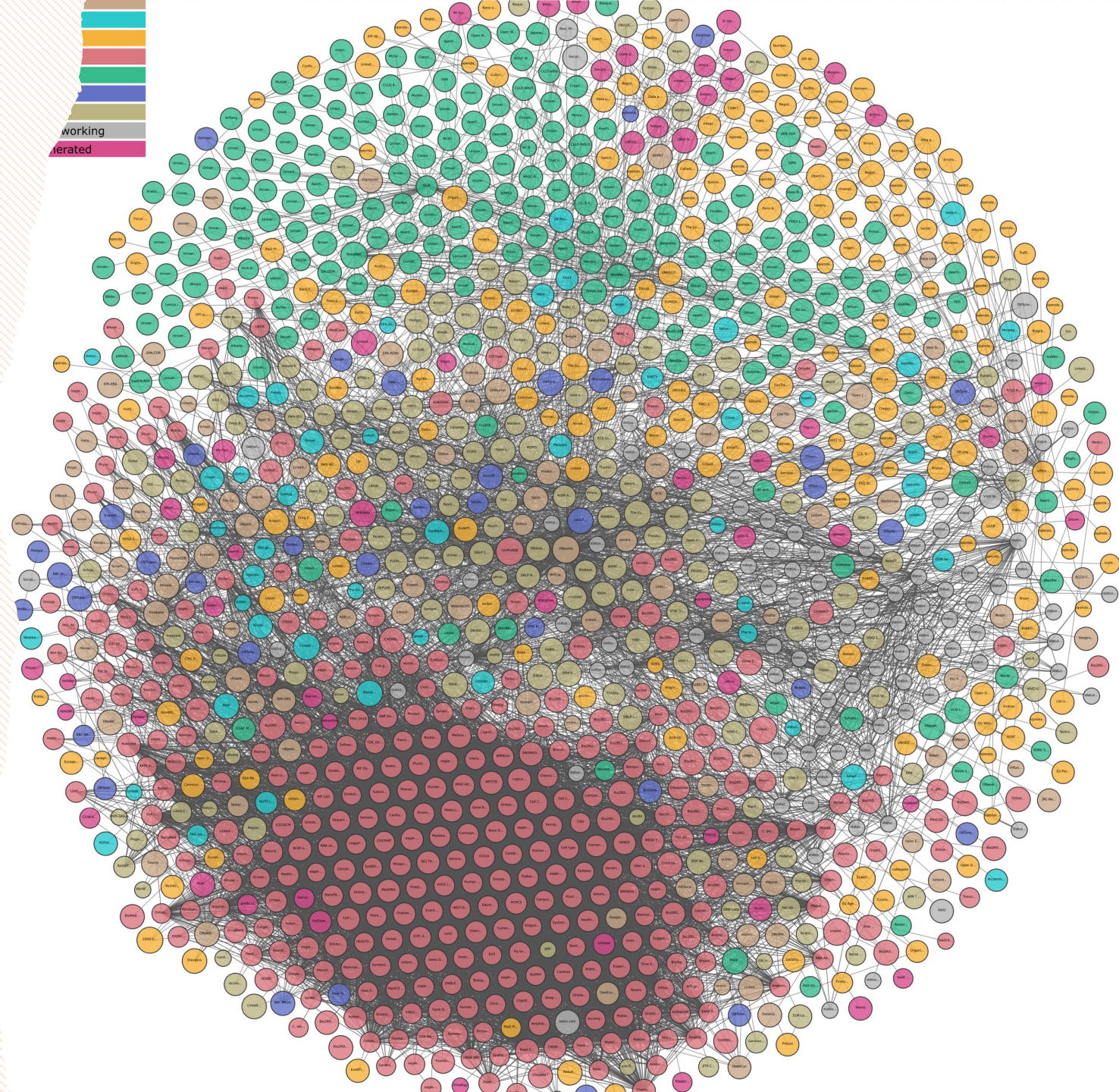
- <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>
- Simple Dublin Coreよりもより詳細な記述に用いることを想定したメタデータ語彙
- プロパティの他にクラスや種別 (Type) の語彙などを含む

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/>



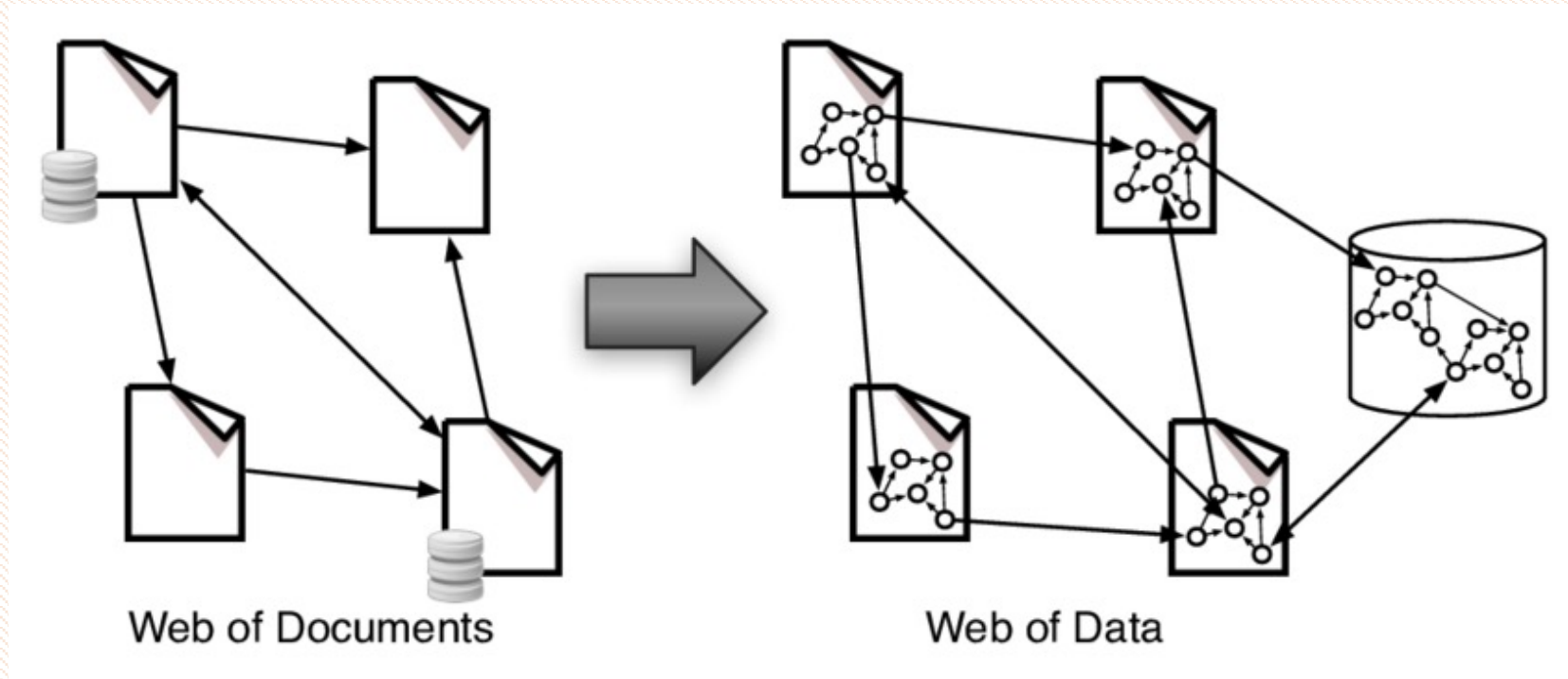


図. 文書のウェブからデータのウェブへ

Emanuel Berndt. (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

スキーマの選択・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）

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、ひとつよう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）