

---

# ジャパンサーチ利活用フォーマット 概念モデルと語彙

---

## 目次

<b>1</b>	<b>モデル設計の考え方</b>	<b>1</b>
1.1	統合プラットフォームの目的確認	1
1.2	メタデータの役割と基本的な要件	1
<b>2</b>	<b>基本データモデル</b>	<b>2</b>
2.1	共通アーカイブ情報と利用者タスク	2
2.2	ソース情報の分離と EDM	2
2.3	統合プラットフォームのソース分離モデル	3
<b>3</b>	<b>メタデータのプロパティと構造</b>	<b>5</b>
3.1	メタデータ項目と優先度	5
3.2	資料を記述するメタデータ各項目	5
3.3	資料のアクセスとソース情報に関するメタデータ項目	13
<b>4</b>	<b>語彙の検討</b>	<b>17</b>
4.1	基本記述語彙の選定	17
4.2	関係モデルの構造化プロパティ	19
4.3	アクセス提供情報とソース情報	20
4.4	クラスと概念体系	21
4.5	Linked Data としての共通アーカイブ情報	21
4.6	標準語彙との関係	22

# 1 モデル設計の考え方

## 1.1 統合プラットフォームの目的確認

「知的財産推進計画 2017」に従えば、デジタルアーカイブを活用するための「つなぎ役」として、分野横断の検索機能のほか、各アーカイブ機関やつなぎ役が整理したメタデータを集約・共有化し、活用者による様々な形での利活用に資する統合ポータルを構築する。

- デジタルアーカイブ化によって分野・地域を超えた知を集約し、学術研究・教育・防災・ビジネスへの利活用が期待できることに加え、海外発信機能を付加・強化することにより、インバウンドの促進や海外における日本研究の活性化にもつながりうる。
- デジタルアーカイブが国内外において日常的に活用され、新たなコンテンツやイノベーションを生み出すための基盤となる社会を実現するため、今後、各アーカイブ機関を結ぶ「つなぎ役」と国等が一体となった取組を加速することが必要。
- 「つなぎ役」は、分野内のメタデータ項目を標準化するために分野ごとに標準メタデータ項目を作成していくこと、さらに、その分野において、長期に渡ってデジタルアーカイブ基盤を維持できるよう、アーカイブ機関の技術、法務上の課題等に対応できる人材の育成をサポートしていく役割が求められる。

(知的財産推進計画 2017 より)

ただし統合プラットフォームは、書物の書誌データのように必ずしもデジタルデータではないリソースのメタデータも対象とする。上記のデジタルアーカイブをより広い知的文化財資源(CHO)と読み替え、「各分野の CHO メタデータを集約・共有化し、横断検索をはじめとする様々な活用の促進に資する」ことと考える。

## 1.2 メタデータの役割と基本的な要件

各機関が保有する資料を、検索・表示・提供するための情報。

- つなぎ役(アグリゲータ)との連携が進むと想定されるため、つなぎ役の参加機関の名称等を検索・表示・提供時に出力できるようにする(参加機関のインセンティブ向上)。
- デジタルコンテンツの有無、ライセンス情報の区分等でも絞り込み等ができるようにする。
- 従来の図書館資料で扱ってこなかった各種文化財についても、適切な形で検索・表示・提供できるよう、必要な情報を記録化できることが望ましい。例えば、材質・構造・技法等のモノに関する情報、その由来情報、地理的な情報、いわゆる著者以外の関係者に係る情報、などが考えられる。

## 2 基本データモデル

提供機関から収集したメタデータはそれぞれ異なる目的のために設計されたもので、その意図を十分生かしながら共通のメタデータにマッピングすることは難しい。そこで統合プラットフォームにおいては、収集したソースデータを、(1) 利用者タスクに対応して整理した分野横断モデルのデータ（共通アーカイブ情報）に変換する；(2) 併せて元のデータを収集元などのメタデータとともにソース情報として保持する；ことで共通性と個別性を両立させる。

- 項目を指定した検索、結果の識別・選択、資料アクセスのための情報提示には、共通アーカイブ情報を利用する。
- 自由キーワード検索はソースデータを対象とし、結果表示に共通アーカイブ情報を利用する<sup>1</sup>。
- 検索結果の個別詳細において、共通アーカイブ情報に加えてソースデータの確認を可能にする<sup>2</sup>。資料へのアクセス提供情報の表示を工夫する

集約データ（共通アーカイブ情報）とソースデータを分離したうえで連動させるために、Europeana 型の EDM モデルを念頭に置きつつ、扱いやすいモデルを設計する。

### 2.1 共通アーカイブ情報と利用者タスク

基本要件（§1.2）における「検索・表示・提供」機能を、FRBR §2.2 Scope での利用者目的分析にもとづいて表 1 の 4 つのタスクで捉え、それぞれのタスクの観点から共通アーカイブ情報に求められるメタデータ項目（プロパティ）を分類整理する。

表 1: 利用者の観点で統合プラットフォームに求められる機能

機能	内容
発見（検索）	キーワードなどで検索するための情報（領域の知識を前提とせずに）
識別（表示）	示されている対象が何なのか、求めている（既知の）資料かどうか判断できる情報
選択（表示）	示されている対象が求めている（未知の）資料かどうかを判断でき、比較検討が可能な情報
取得（提供）	個別リソースあるいは詳細メタデータにアクセスするための情報

このタスクに基づく具体的なプロパティ項目は、§3 において詳細に検討する。

### 2.2 ソース情報の分離と EDM

EDM は ORE の集約モデルをベースにいくつかのパターンを提示しているが、Europeana が採用しているのは、集約データにもとづいて Europeana が付与したメタデータと、集約元のメタデータを分離して扱うものである。統合プラットフォームで言えば、前者が共通アーカイブ情報、後者がソース情報に相当する。モデルは 3 つのリソース型を核に構成される。

<sup>1</sup>当初案で検討した全文テキストの RDF グラフは用意せず、全文検索はソース情報側に委ねる。SPARQL クエリはグラフ構造を利用した検索が利点であり、全文テキスト検索のニーズは高くないと思われる。

<sup>2</sup>提供元ページへのリンク、およびソースデータの表示・ダウンロードも提供する（ソースデータはオープンライセンスで公開とする）。ソースデータは、項目名や構造は提供されたままとするが、処理を容易にするため JSON などのフォーマットに変換して保持する。

- 作品（資料）の実オブジェクトを ProvidedCHO とする
- 関連リソース（作品のデジタル化表現や情報ページなど）と作品オブジェクトを Aggregation として集約（グループ化）する
- 複数の視点（Aggregation）からの実オブジェクトに関するメタデータを、それぞれ代理オブジェクト Proxy によって記述する

統合プラットフォームに対してこれをストレートに適用した場合、おおむね図1のようなグラフを考えることができる。

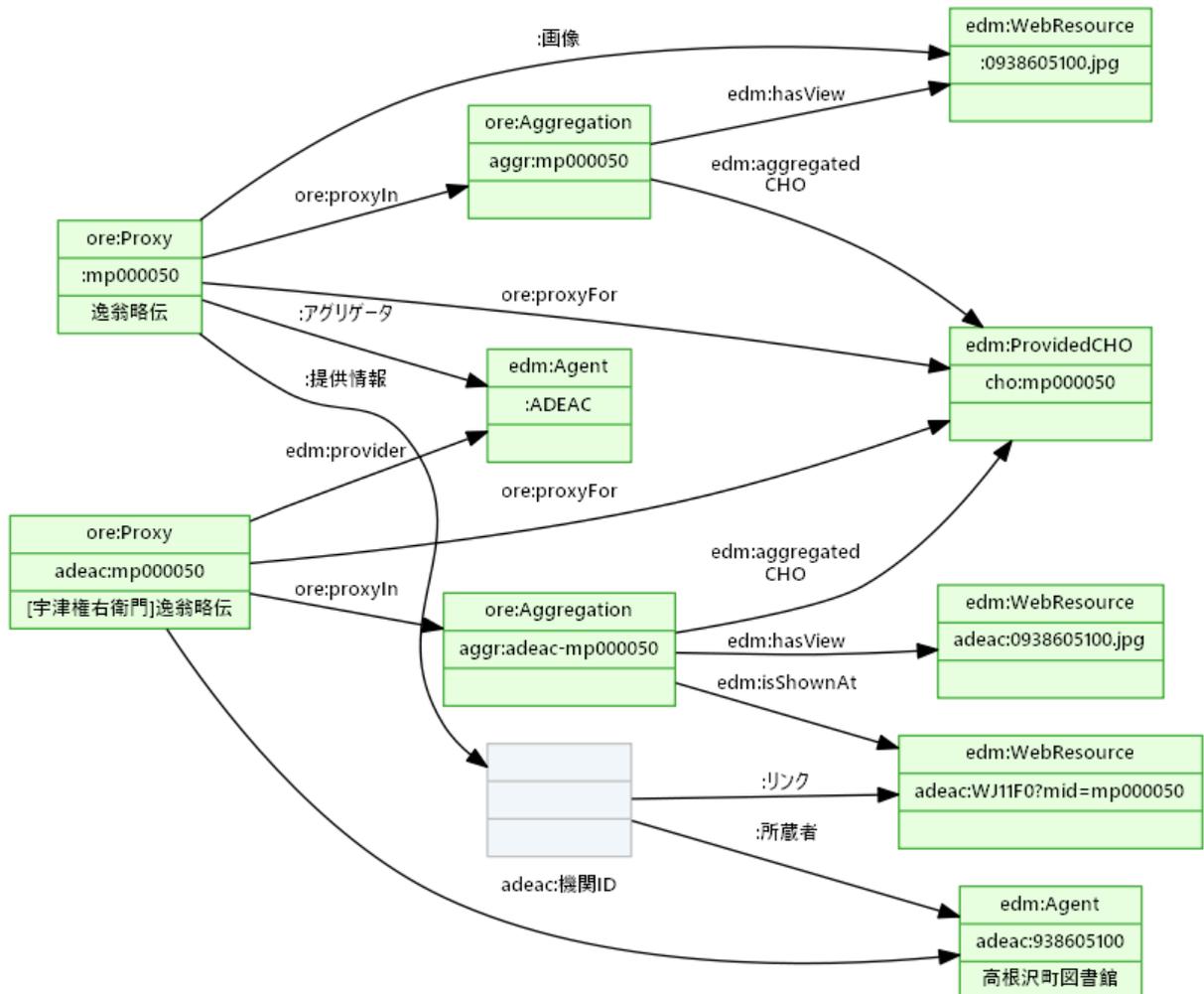


図1: Europeana の EDM に近い形のモデル。この図はノードの上中下段にリソースの型、URI、ラベルをまとめることでグラフをコンパクトにしている。図右上の画像は、右中ほどの画像からサムネイルを生成したと想定。

### 2.3 統合プラットフォームのソース分離モデル

Europeana との互換性を重視する場合は図1を検討する価値があるが、このモデルは利用者にとってかなり複雑で理解し難い。関連リソースの Aggregation は直接収集したもの以外は提供側にあるので、統合プラットフォームのメタデータとして扱うのは複雑である。

そこで、EDMのリソース型概念を前面に出さず、収集元の視点によるメタデータを「ソース情報」という枠組みで分離し、構造を簡潔にしたモデルが図2である<sup>3</sup>。

統合プラットフォーム側の Aggregation に相当するノードは「提供情報」とし、収集するオブジェクトに限らず資料アクセス関連の情報を集約する<sup>4</sup>。ソース情報側は、メタデータとリソース集合を分離していないので ORE/EDM のモデルには直接対応せず<sup>5</sup>、全体で提供元の視点によるメタデータと関連リソースを表す<sup>6</sup>。

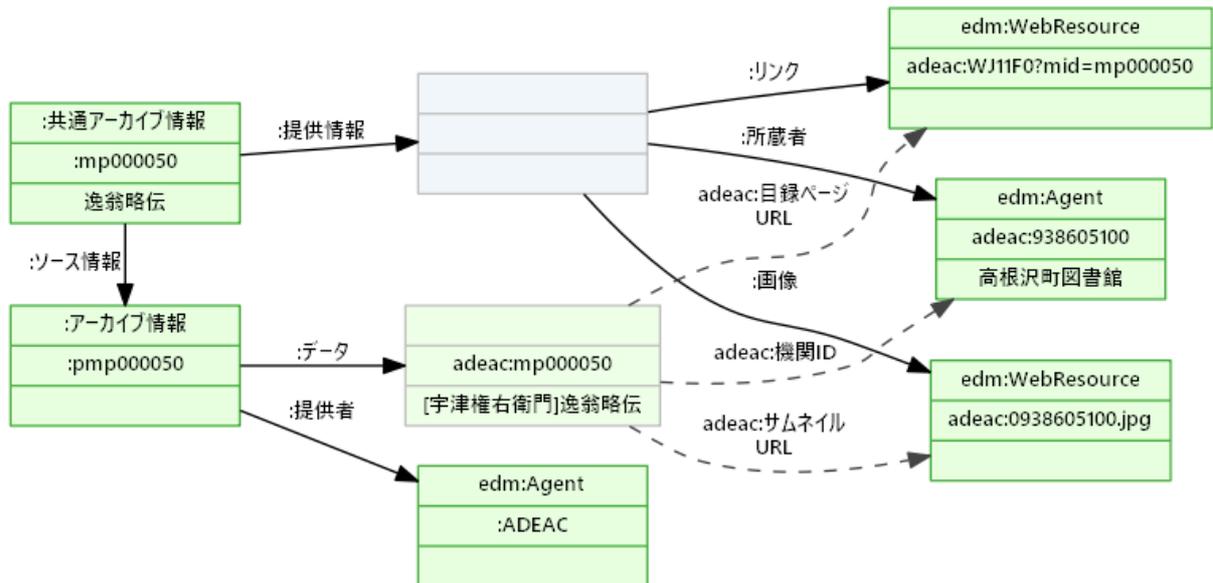


図 2: 共通アーカイブ情報から「ソース情報」プロパティで収集元のプロキシに結びつけるモデル。

<sup>3</sup>当面 ProvidedCHO に相当するリソースは独立させず、共通情報側にそれぞれの内容に対応した基本型（画像など）を併記している。

<sup>4</sup>リソースの説明も含むので、ORE では Aggregation の説明記述である ResourceMap に近い位置づけになる。

<sup>5</sup>EDM との対比のためには、提供元の視点を「代理する」という意味で「ソース情報リソース」を Proxy、画像などのリンクを含むという意味で「ソースデータ」を Aggregation に当たるものと考え、対応関係を描くことはできる。

<sup>6</sup>DPLA では、収集側の Aggregation を起点にして収集側メタデータを `sourceResource` で関連付け、元データは RDF 化されないそのままのデータを `originalRecord` として関連付けている。

## 3 メタデータのプロパティと構造

### 3.1 メタデータ項目と優先度

検討会や質問・意見をを受けて検討したメタデータ項目を表 2 に示す。

#### 3.1.1 単純プロパティと構造化プロパティの併用

利用者の発見タスクのためには、プロパティ項目は単純であることが望ましい。一方で識別や選択タスクのためには、プロパティは適切な詳細度が必要である。

たとえば映画作品の情報には監督、脚本、撮影、音楽などさまざまな役割がある。制作に関与した人を調べるためにはこれらを同一の単純プロパティ（寄与者）で検索できると便利である。しかし結果を適切に識別し選択するためには、その人がどのような役割で関与したのかも知る必要がある。

時間・場所についても同様に、制作（creation）だけでなく採集、撮影、主題などさまざまな種類がある一方、利用者はそうした違いを意識せずに時間・場所を検索したい場合が多いと思われる。

そこで、これらの「いつ」「どこで」「だれが」に関しては、まず単純なプロパティで記述し検索などの利用を容易にすると同時に、提供データの詳細を構造化して記述するプロパティを併用する（表 2 においては単純プロパティを「ショートカット」としている）。また資料の提供情報とソースデータ情報の関係を整理し、それぞれを構造化した。

以下においてこのメタデータ項目を、資料自身の記述に関するもの（§3.2）と、そのアクセスに関するもの（§3.3）に大別して検討する。なおソースデータについての記述である「ソース情報」の内容も、便宜上 §3.3 の一部として扱う。

### 3.2 資料を記述するメタデータ各項目

#### 3.2.1 タイプ

資料の基本区分を表す型を付与する。資料の基本区分は、複数種類の資料情報が混在するアーカイブにおいて、情報を大きく区分するために重要である。ただし収集するデータだけではバランスの取れた型を付与できないかもしれないので、個別検討が必要。

また共通アーカイブ情報自身を表す型も（EDM での Proxy のように）必要だが、「文化財」でもあり「アーカイブ情報」でもあるというリソースは混乱を招く恐れがあるので、具体的なプロパティ設定においてその違いを考慮する。

表 2: 統合プラットフォームが持つメタデータ項目

基本項目	内容	F	I	S	O
タイプ	資料の基本区分+共通情報を表す型		-		-
ラベル	一覧などに表示し資料を識別する名前	-			-
名称	タイトル、別名、読みなど検索対象とする名前				-
寄与者/関係	資料に寄与した人/組織。出演者等も				-
	寄与者一般としてのショートカット				-
作者	寄与関係 [制作] へのショートカット				-
発行者	寄与関係 [出版] 相当へのショートカット				-
場所/関係	場所に関する構造化情報。制作/内容は関係で区別する				-
	場所に関するショートカット				-
時間/関係	時間に関する構造化情報。制作/内容は関係で区別する				-
	時間に関するショートカット				-
	時間のリテラル値 (標準プロパティがある場合)				-
主題	主題および分類 (キーワード含む)				-
区分	一定範囲、分野の枠組みに基づく区分/カテゴリー				-
識別子	ISBN など体現形レベルの識別子			-	
言語	資料の記述言語				-
画像	資料の特徴を確認するための画像	-			-
資料体記述	資料の物理的特徴・素材等の記述 (リーダー付き)				-
記述	個別項目に収録できない情報 (リーダー付き。表示にはリーダーも項目名に含める)				-
上位資料	現在の資料がその一部である上位資料				-
提供情報	資料の提供に関する情報	-	-	-	-
提供者	資料 (に関する情報) の提供者。保管者は別プロパティで定義	-			
リンク	資料の紹介ページやアクセス情報が記載されたページの URL	-	-	-	
オブジェクト	資料のデジタル画像、音声動画などの URI	-	-		
権利情報	資料利用のライセンスおよび権利	-	-		
個別識別子	提供元が付与する識別子			-	
ソース情報	ソース情報およびその提供者に関する情報	-	-	-	-
提供者	ソース情報の提供者 (アグリゲータ)	-	-	-	
データ	プラットフォームが保持・提供するソース情報データ	-	-	-	
リンク	アグリゲータの情報ページ	-	-	-	
更新日	収集元データの更新日	-	-		

### 3.2.2 ラベルと名称

資料を識別するための名前をラベルとし、汎用的なプロパティで付与する。ソース情報にラベル相当の名前がない場合は、アーカイブで連番などに基づいて生成する。言語タグは与えない。

ラベルは原則としてすべての資料が持つべき必須項目であり、アーカイブ外から利用するときにも取得した情報を確認する基本項目として重要である。

加えて、作品などに与えられた名前を名称とする。原則として全て言語タグを付与し、読みも@ja-Kanaなどの言語タグによって区別する。サブタイトル、別名も名称として並列に扱う（主タイトルはラベルと一致するものを調べることによって区別する）。

必須のラベルとは逆に、ソース情報に名称相当の項目がない場合は、名称は設けない。

タイトルと別名の区別 タイトルと別名などを同じ名称として扱うのは、異なる名前を単一プロパティで検索できるようにするためである。標本の正式名称が「はえ追い」で別名が「払子（蠅払い）」となっているとき、どちらの名称でも区別なく検索できる方が利用者にとっては便利だといえる<sup>7</sup>。

読みを構造化せず、別プロパティにしないのも同じ理由からである。またラベル（よみ）のように読み項目を設けずラベルや名称に付記する形で読みを加えているソース情報もあり得る。

ただし、英文でも主タイトルと別名が提供されている場合、両者を等しく名称として扱うと主タイトルが区別できなくなる。サンプルデータでは見当たらないが、もしこの区別が必要であれば、ラベルにも言語タグを付与する必要がある<sup>8</sup>。

### 3.2.3 作者、寄与者および発行者

作者、発行者などはすべてリテラルではなく実体（Agent）として記述する。NDLA の名称典拠とマッチングさせて典拠 URI で記述することを目指す。典拠 URI が得られない場合でも、ソースデータでの ID などに基づき、仮 URI を付与する<sup>9</sup>。

<sup>7</sup>もちろん 2 つ以上の項目を対象に検索することはできるが、シンプルな検索で見つかるほうが利用しやすい。

<sup>8</sup>なお言語タグの与え方としては、(1) 他言語もしくは読みがある場合は、区別のために各ラベルに言語タグを加え、(2) 単一ラベルの場合には言語タグは付与しない、という方法も考えられる。本来は、言語タグを用いるなら全てのラベルにタグを付与すべきだが、ソース情報の同一ラベル項目に複数言語が混在している可能性があり、これに一律の言語タグを付与するとかえって不正確になる。なお、同じリソースに対するラベルに言語タグ付きと無しが混在するのは、アプリケーションにとって扱いが難しいため、避けるべきである。

<sup>9</sup>ID がない場合でも、機関名 + 氏名のハッシュなどで URI を生成する。最小限でも同一ソース内で同一実体を集約できる他、後日典拠 URI との一致が得られれば、典拠 URI に置き換えた上で旧 URI は転送するなり sameAs を用意することで恒久性を確保できる

寄与者は役割を記述するために図3のようなロール・モデル<sup>10</sup>を用いて寄与関係として記述する。

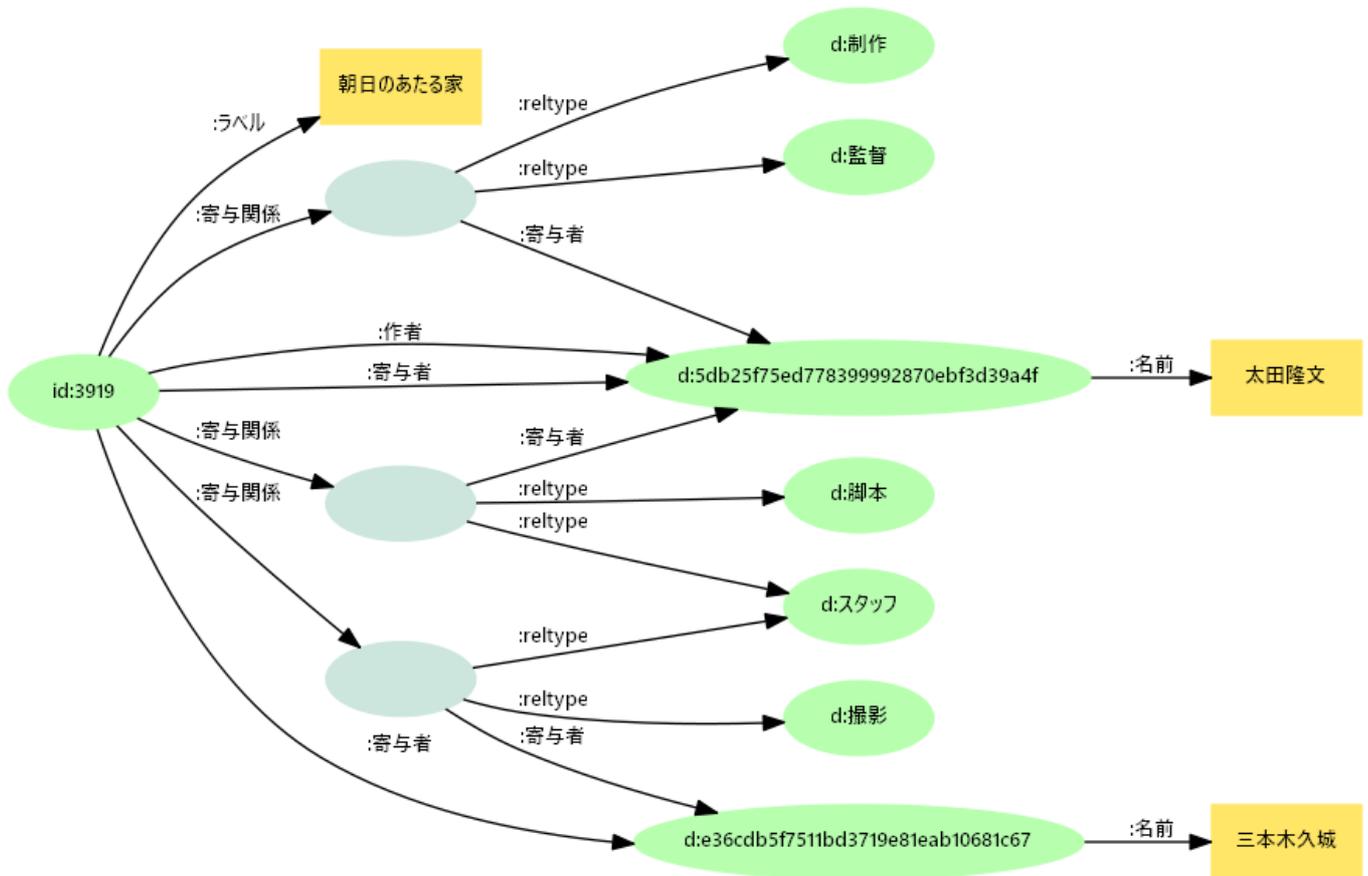


図3: :役割を作品と寄与者実体を直接結ぶプロパティとして用いる

ここで、寄与者は作者などより一段深い位置におかれて利用しにくい可能性や、役割が明示されなければ単に冗長であるという問題にも対応するため、単純プロパティ(ショートカット)も併用する<sup>11</sup>。

- 作者も常に寄与関係の一つとして併記し、役割を「制作」などにする。作者も含め関連する人(実体)を調べたい場合は寄与関係として検索できる。
- さらに作品と寄与者実体を直接結びつける寄与者プロパティを追加する。プロパティが不明でも作品に直接つながる実体を検索すればよい(図3)。

発行者も、「出版」もしくは「発行」という関係において寄与する実体の一つとして捉えることができる。併用する単純プロパティを、一般的な「発行者」とする。

出版に関するイベントモデルは採用しないが、以下に述べるように場所、時間情報についても関係型モデルを用い、その関係を「出版」とすることで、出版に関わる実体やデータを集中させられる。

<sup>10</sup>Schema.org のプロパティ反復形を取り入れている。役割を階層語彙として定義するのも検討に値する。

<sup>11</sup>発見タスクの観点で考えれば、ショートカットは未知の度合いが高い(利用者側の情報が少ない)場合、構造化プロパティは既知の度合いが高い場合に利用できるものと言える。

### 3.2.4 時間関係

作成、公開、発見、内容<sup>12</sup>など対象に関する時間範囲を実体（時間実体）として記述する。またどのような関係でその時間とつながるかを示すための時間関係も併記する（図4）。

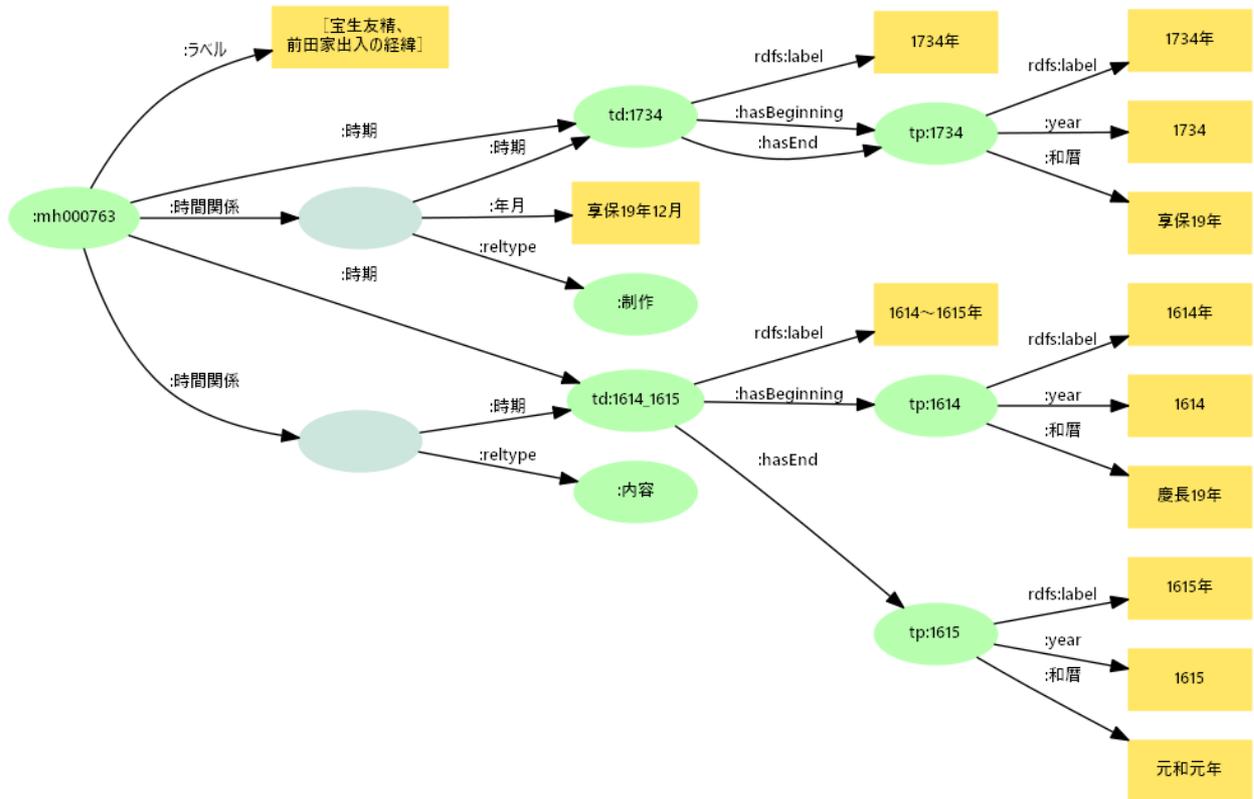


図4: 時間範囲を実体として扱う。時間関係ノードに「作成」「内容」などの型を与えることで区別する。また個別レコードに記述するのは接頭辞 td: のノードまでで、それ以降は共通の情報となる。

- 時間実体は、開始年、終了年（いずれも時間実体）を持ち、範囲検索を可能とする。また 1901 ~ 2000 年の別名実体（sameAs）として「20 世紀」、1603 ~ 1868 年の別名実体として「江戸時代」を定義することで、時代区分も同様に表現可能とする<sup>13</sup>
- 時間関係は、寄与関係と同様に、リソースと時間実体の関係を中間ノードを用いて記述する。また時間実体は年を単位とするので、月日などはこのノードにリテラルで追記する。
- 制作、出版など標準的に用いられるプロパティに対応する場合、時間リテラルをリソースの直接プロパティとして付与する。

時間実体に関する情報（開始/終了年、別名実体など）は、レコードごとに作成するのではなく時間オントロジーとして共有し、煩雑さを回避するとともに、多様な利用を可能とする（図5）。

<sup>12</sup>時間関係による「内容」とは、作品の記述対象である時代など、dct:temporal として記述されるものを念頭に置いている。

<sup>13</sup>範囲ではない特定時期も、開始年と終了年が同一である範囲として扱う。DPLA では dc:date の値を常に構造化し、edm:begin、edm:end によって範囲を示している。

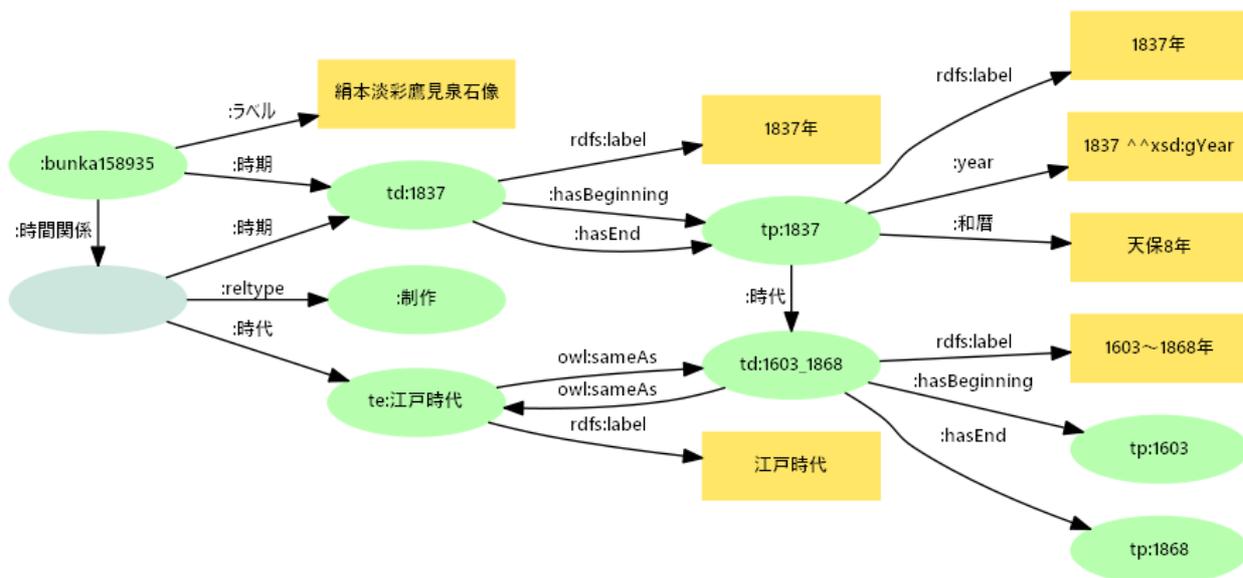


図 5: 時代、和暦、開始/終了年などを時間実体の情報として共有する。

### 3.2.5 場所関係

時間と同様に、作成、公開、発見、内容<sup>14</sup>など対象に関する場所の情報を一括して「場所関係」とする。また場所を示す実体をショートカット結びつける。時間以上に場所を表す項目はさまざまなものが見られるので、プロパティを統一した上で関係を示すモデルは共通利用に有益と思われる（図 6）。

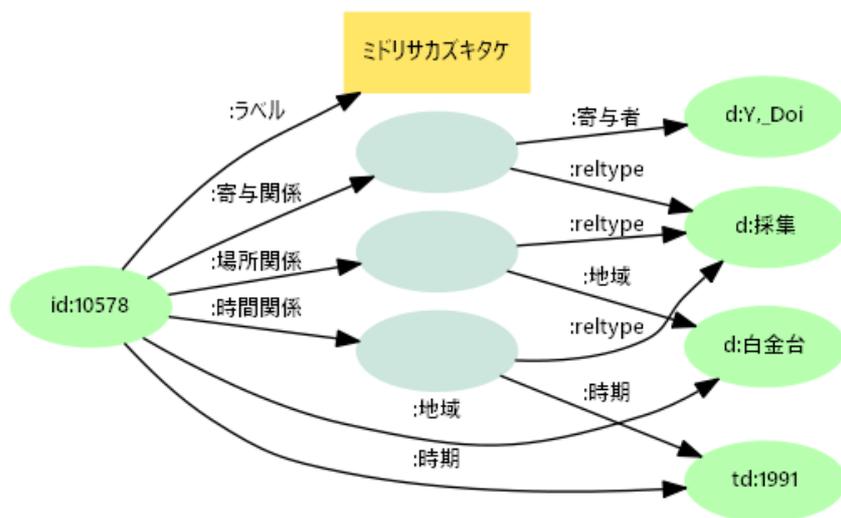


図 6: 場所情報の関係を示すことで、同じ「採集」関係にある寄与者、時間情報が集約できる。

ここで、場所も時間と同様に、国内は都道府県、海外は国単位の統一実体を定義することが、地域情報との連動など利活用の観点では非常に重要になる。ただしサンプルデータの段階でも場所は「日本」から「白金台」まで粒度がさまざまであるため、正規化には一段の工夫が必要。

<sup>14</sup>場所関係による「内容」とは、作品の内容となる対象地域、舞台など、dct:spatial として記述されるもの。

### 3.2.6 主題と区分

資料の主な内容やテーマを表現する件名および分類を、主題とする。NDLSH など統制され URI を持つ値を利用する。キーワード文字列のみの値は、マッピング時の正規化を検討するが、困難であれば特別な名前空間でキーワードを直接 URI 化する<sup>15</sup>。

区分は、汎用体系を用いた主題・分類ではなく、分野における区分けもしくはジャンルを示す。国宝、ニュース、ドキュメンタリーなどが該当する<sup>16</sup>。

### 3.2.7 識別子、言語、画像

資料に与えられる体現形レベル (ISBN 等) の識別子を収める。導入句付き、NDL サーチのようなデータ型付き、あるいは URN など値の持ち方は別途検討する<sup>17</sup>。所蔵館がアイテムに与える個別識別子は、後述の提供情報の一部として扱う。

国際的な利用も踏まえ、内容の言語を URI として収める。複数言語で提供されるものはプロパティを反復する (アブストラクトのみ英語ありといった、部分的な要素は言語情報の対象としない)。

また資料を識別するために、特徴がわかるサムネイル～低解像度画像を用意する。これはアクセス情報として扱う (ある程度高解像度の) 画像とは別に、統合プラットフォームで保持する。

### 3.2.8 記述

他のプロパティで表現できないテキスト情報を、元項目を導入句として加えた形で収める。

資料体の記述 媒体 (メディア) 形態、サイズ、数量など試料の物理的な特徴は、共通のプロパティを用意したうえで、導入句付きの記述として扱う<sup>18</sup>。資料種別は多くの場合いくつかの選択肢による値となっている (カテゴリに近い) ので、URI で表現し、区分に準ずるプロパティで記述する。

内容記述及び注記 概要・要約、注記、備考など物理特徴以外のその他の情報は「記述」として扱う<sup>19</sup>。

項目 (プロパティ) を細分化しないのは、その方が検索利用しやすいこと、識別・選択のためには導入句があれば用が足りること、分野 / 提供者ごとに異なる多様な項目をアーカイブで反映するのは困難なこと、詳しくはソース情報によって確認できること、による。

なお情報の表示にあたっては、導入句を値から切り離し、項目名に付加して「記述 (注記)」のような表示項目として用いることで、より確認しやすくなると考えられる。

### 3.2.9 上位資料

ソースデータ内に上位資料を識別できる ID がある場合に設定する。このとき上位資料は：

<sup>15</sup>たとえば CiNii は論文のキーワードをそのまま URI 化して foaf:topic の値としている。同音異義語の衝突可能性を了解した上で利用すれば、十分有益だと思われる。

<sup>16</sup>日本十進分類のようなあらゆる領域をカバーする汎用的で精緻な分類ではなく、その分野固有の大きな分類と考えることができる。提供情報によっては、タイプと同じになってしまうことがあるので、その扱いは一考を要する。

<sup>17</sup>導入句を加えるとプロパティ値が識別子自身と違ってしまふ。人間が確認するという意味での利用者タスクにとっては問題ないが、項目名ではなく値については機械利用という観点も考える必要がある。

<sup>18</sup>サイズ、数量などは形式ばかりでなく単位のあり方もばらばらなので、元の項目名を保っても機械利用は難しい。現状では、人間利用者の識別、選択タスク用である。

<sup>19</sup>概要・要約を独立させることも検討したが、提供される割合が低く、逆に検索漏れを生じる可能性が高くなることから、記述と一体化する。

1. ソースが上位資料を独立レコードとして含んでいれば、それに対応する共通情報リソース
2. 上位 ID はあるが、その ID に対応するレコードとしては記録されない場合は、(ラベルがあればそれのみを持つ) 接続情報としてのノード
3. 上位の名称のみがある場合は、名前に基づく仮 URI を生成して接続ノードを作る<sup>20</sup>。

例 1：国立公文書館のデータ 資料群、簿冊、件名/細目がそれぞれ別レコードとして提供され、かつ「親メタデータ ID」によって上位と関連付けられている(図 7)。

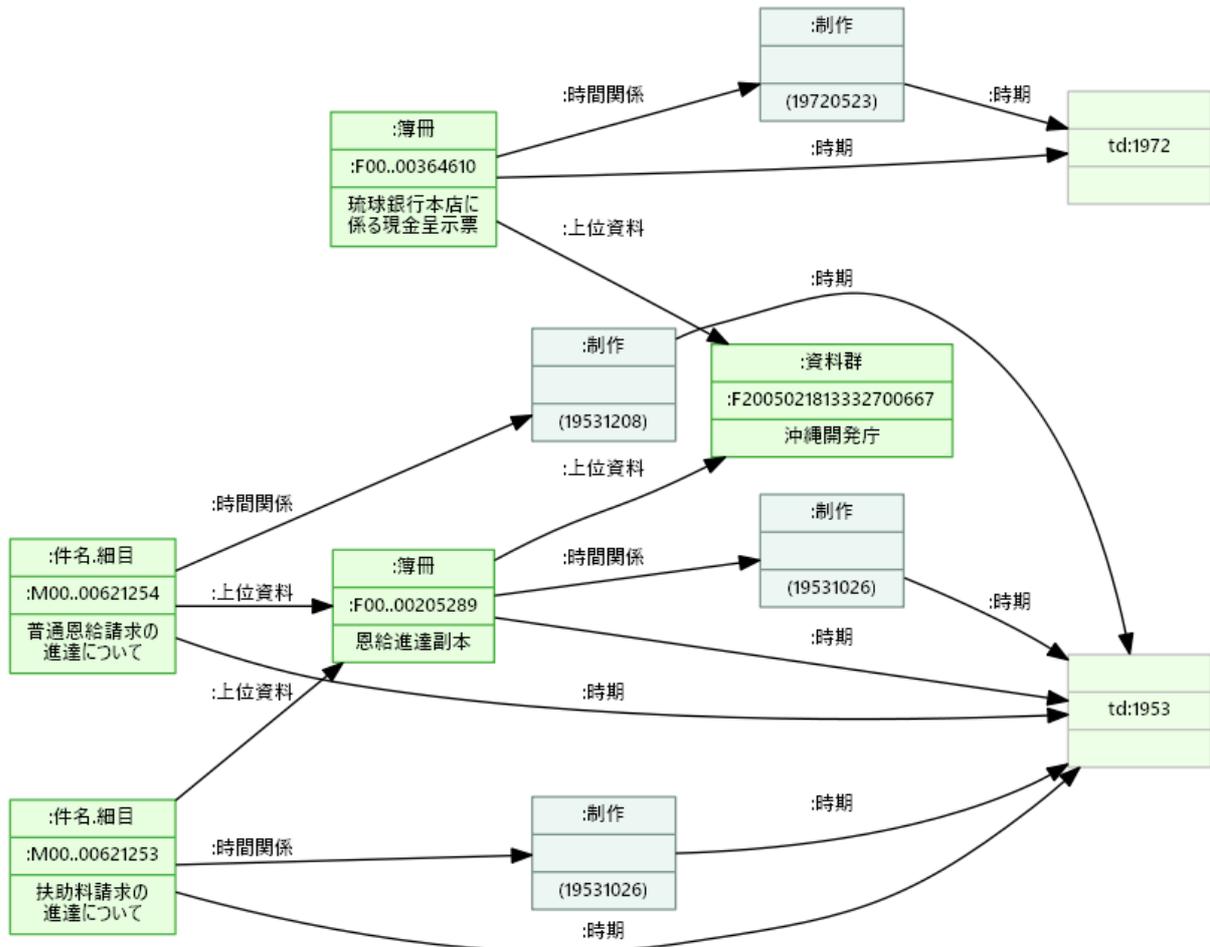


図 7: 上位 ID がレコードとして含まれる場合

この関係をたどり、公文書などで重要とされる「最上階層レコードとの結びつき」が確保できる。

例 2：ADEAC のデータ 各資料が含まれる「刊本 ID」「刊本名」を持つが、刊本 ID に対応する独立したレコードはソースデータには存在しない(図 8)。

この場合、上位資料についての情報は共通アーカイブからは得られないが、同じ上位資料を共有するレコードのつながりを見出すことができる。また上位資料の URI が参照解決可能(リンクしている)であれば、リンクを辿ってその情報を取得することが期待できる。

<sup>20</sup> 名前からの URI 生成は、データの正確性に依存する。表記の揺れや句読点違いなどがあると成立しない。

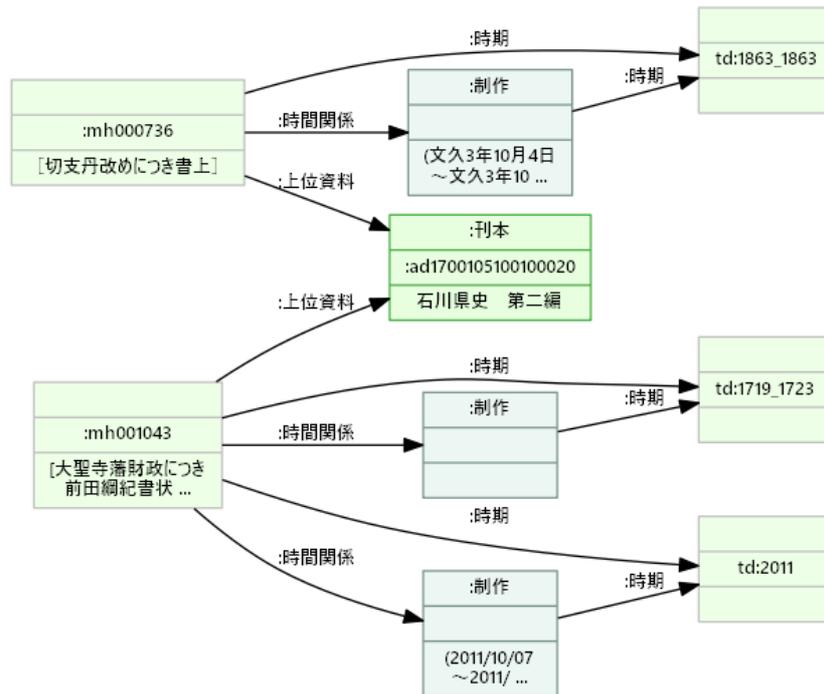


図 8: 上位 ID はあるがレコードに含まれない場合

例 3 : NDL サーチのデータ 書籍のシリーズ、雑誌の特集などは `dcndl:seriesTitle` にリテラル値があるのみなので、リテラル値のハッシュなどから仮 URI を生成して同じシリーズを結びつける<sup>21</sup> (図 9)。

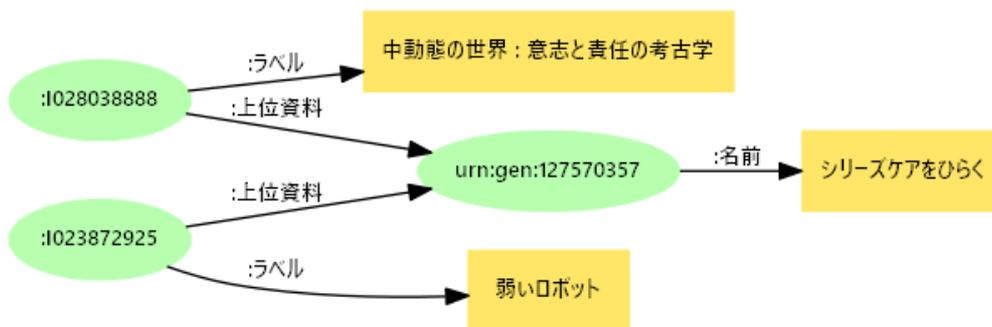


図 9: 上位については名称のみがある場合

### 3.3 資料のアクセスとソース情報に関するメタデータ項目

#### 3.3.1 資料アクセスの提供情報

資料オブジェクト自身、そのデジタル化や複製物、および関連情報へのアクセスを提供するための情報を構造化して記述する。EDM の Aggregation に近いが、直接収集したものに限らず、URL によって

<sup>21</sup>NDL サーチのウェブページ詳細画面では検索リンクが提供されているが、データとしては関連付けられていない。ハッシュに基づく URI であれば、同じタイトルは同一の URI として表現し、結びつけることができる。

提供元にアクセスできるもの<sup>22</sup>や、現地に行って閲覧できるものも含む。以下の項目を持つ。

- 提供者：資料（の複製物）の提供者を識別する URL。保管者が別であればその URL も<sup>23</sup>。資料自身のほかに解説や複製物などが複数箇所から提供される場合、それらは次のリンクで示し、提供者は資料自身に最も近いものを提供する博物館などを示す。なお、ソース情報がアグリゲータ経由で提供される場合、その元となる一次データの作者がこの提供者となる<sup>24</sup>。
- リンク：資料の紹介ページやアクセス情報が記載されたページの URL。画像資料の IIIF マニフェストなど特別な URL は、クラスを付与して判別できるようにする。
- オブジェクト：資料のデジタル画像もしくは音声動画などの URL。共通項目の識別用画像とは別の、提供元が保持するもので、複数の場合もあり得る（画像閲覧システムの該当ページは「リンク」で）
- 権利情報：提供者が権利情報を示していれば、そのテキスト。画像などオブジェクトのライセンスがある場合は、オブジェクト URI のプロパティとしてライセンスの URI 記述するか、提供オブジェクト一括のライセンスを記述する（§3.3.4 も参照）。
- 識別子：提供者 / 所有者が管理するアイテムとしての識別子（請求記号など）。
- 記述：その他、資料へのアクセスに関連する情報があれば導入句付きの説明記述。

ここでの権利情報は、提供される資料（の複製物）に関するものを扱う。メタデータ（共通情報および次項の公開用ソースデータ）は原則として CC0 レベルのライセンスとし、全体に共通する情報として共通アーカイブの説明ページなどに明記する。

### 3.3.2 ソース情報

資料に関する 元メタデータ（収集したソースデータ） についての情報を記述する。§2.3 のソース分離モデルにおけるソース情報に相当する。

ソースデータの提供者のほか、更新日などその管理メタデータを示し、アグリゲータに目録等の情報ページがあればリンクとしてその URL も加える<sup>25</sup>。さらにソースデータをそのまま<sup>26</sup>手を加えずデータとして提供する<sup>27</sup>。

アグリゲータを介する場合は、アグリゲータがこのソースデータ提供者という位置づけになる。資料へのアクセスを提供する主体とソースデータ提供者が同一という場合もあり得る。

### 3.3.3 アクセス提供情報とソース情報の関係

アクセス提供情報とソース情報の提供者が同一の場合 一次データを直接収集する場合は、提供情報とソース情報の提供者が同一となる。たとえば民博ビデオテーク・データベースはソース情報を提供するとともに資料へのアクセスも提供する（図 10）。

<sup>22</sup>EDM においては、これは提供元側の Aggregation として別のグループを構成する。

<sup>23</sup>提供者、保管者の扱いは、ソース情報で公開されるものに準ずる

<sup>24</sup>EDM での dataProvider に相当する。

<sup>25</sup>デジタル画像などがアグリゲータ経由で提供されていても、これらはアクセス提供情報の“オブジェクト”として扱う。これは §2.3 で提供情報を Aggregation と位置付けたとおり、資料に関連するリソースへのアクセスは提供情報に集約するため、アグリゲータの目録は、共通情報の元であるソースデータをウェブページで確認するという目的のためにソース情報の URL とする。こちらでもアクセス提供情報の一部にしたほうが分かりやすいという議論はあるかもしれない。

<sup>26</sup>データ形式は JSON 等の共通フォーマットに変換するが、項目名（たとえば JSON キー）、値は元のままにしておく。

<sup>27</sup>ソースデータ提供時に公開不可項目はあらかじめ提供側が取り除き、共通情報と同じく CC0 で提供する。ライセンスは全体で共通して表示する。

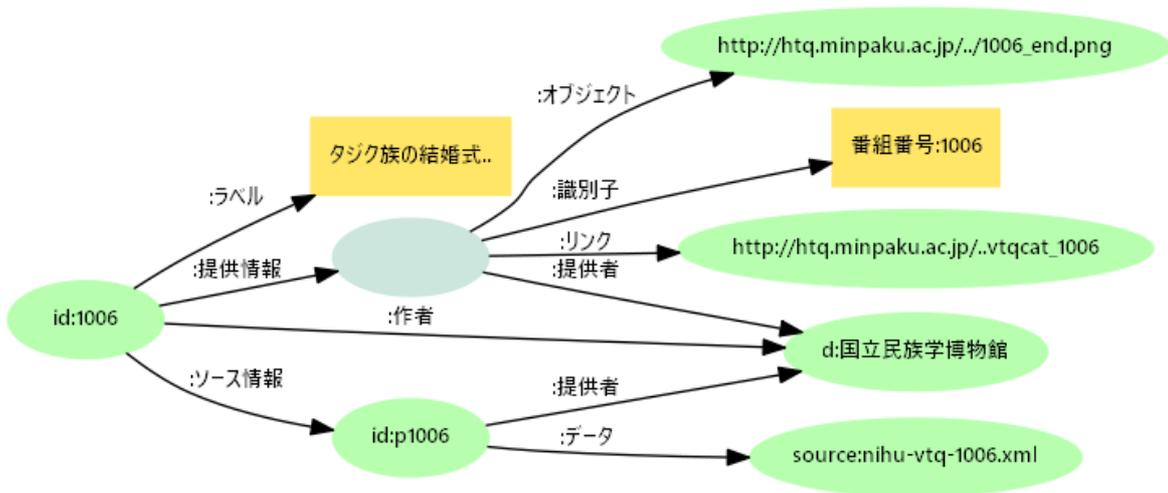


図 10: アクセス情報とソース情報の提供者が同一となるグラフ。ここでは提供者が作者でもある

アクセス提供情報とソース情報の提供者が異なる場合 アグリゲータとは別に資料の提供者（保管者）がある場合は、2つの情報は基本的には分離される（図 11）<sup>28</sup>。

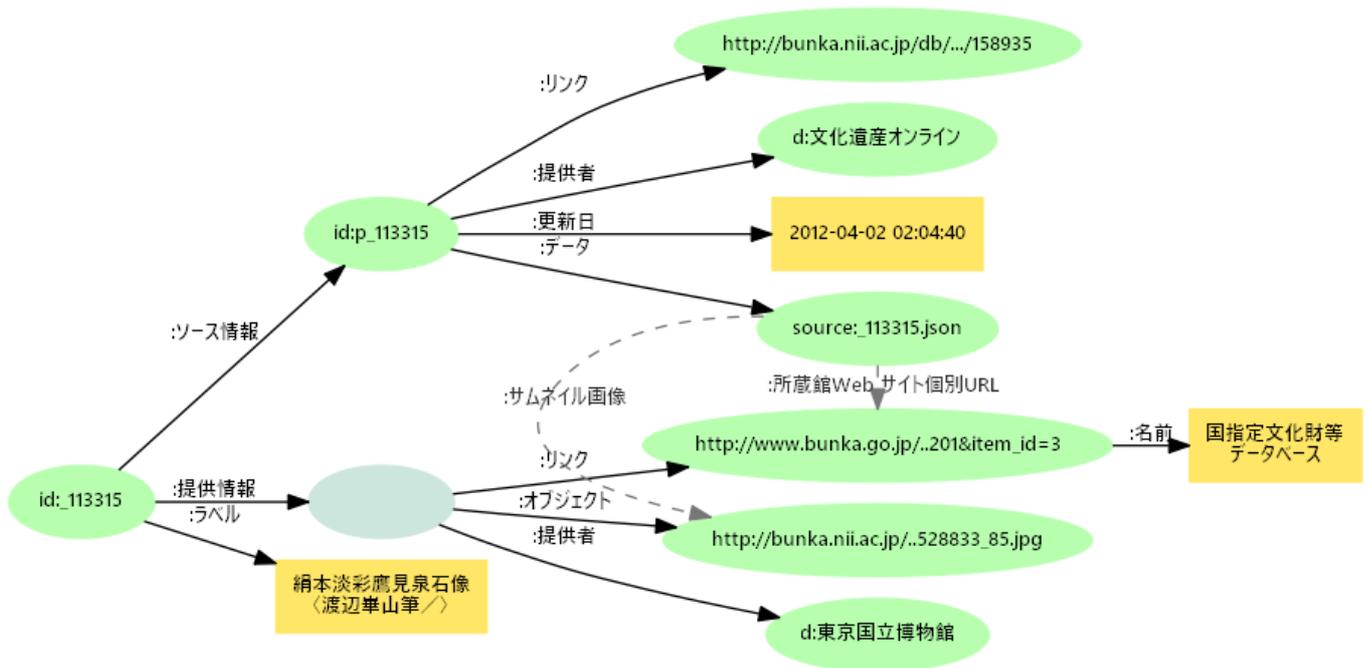


図 11: 提供情報とソース情報の提供者が異なるグラフ。画像はアグリゲータ経由であっても提供情報としている

<sup>28</sup>文化遺産オンラインの場合は、提供者リンクにさらに「国宝」があるが、ここでは省略した。ソース情報のデータ内には資料提供に関する情報が含まれるので、これが RDF であれば、図の破線のようにつながりができることになる。

### 3.3.4 アクセス情報リンク先リソースのプロパティ

権利の記述 アクセス情報として提供する画像リソースについて、ソースデータにライセンスのプロパティがあればそれを用いればよいが、そうでない場合に統合プラットフォームでCC0などのライセンスを記述するのは、メタデータの管理においてやや疑問がある。

Europeanaでは、Aggregationとしての一括権利情報(edm:rights)は必須となっているが、全ての画像にライセンスが記述されているわけではない。これに倣えば、アクセス提供情報単位でライセンスを付与するプロパティを用意する方法もある。ただし画像によってライセンスが異なるという複雑な場合は、これでは対応できない。

もう一つの方法としては、プロパティの使い分けによる区別も考えられる。

- 「オブジェクト」プロパティを用いた場合は自動的にライセンスがCC0とみなせる、などの規約を設ける
- これに該当しないリソースは、アクセス提供情報での記述に別のプロパティを用いる

この方法はモデルとしてはシンプルになるが、ライセンスに関係なく画像を調べるといった検索時に複数プロパティを調べなければならないというデメリットも伴う。

リソースタイプの記述 リンクあるいはオブジェクトでアクセス提供情報に関連付けるリソースに、型情報を与えることが考えられる。§3.3.1でも示したように、IIIF マニフェストなど特別な情報を提供するリソースについては、そのためのプロパティを毎回定義するよりもリソースの型によって判別可能にするほうが利用しやすい。

このとき、特定のリソースだけが型を持つよりは、全てのリンク先リソースが型を持つほうが一貫性があり、利用者にとっても分かりやすい。またオブジェクトについては、画像以外のメディアの提供が今後増加する可能性がある。これらを考えると、最初から全てのリソースに型情報を付与しておくのが賢明と言える。

可能であればリンクおよびオブジェクトには型を付与することにしたいが、提供されるソース情報の項目から曖昧さなく型を推定できるかどうか、検討する必要があるかもしれない。

## 4 語彙の検討

### 4.1 基本記述語彙の選定

共通アーカイブ情報およびソース情報を記述する語彙については、次の点を考慮して検討する。

1. モデルの狙いが的確に表現でき、各項目（トリプル）の意味に整合性があること
2. 少なくとも基本的な部分については、広く理解される既存語彙を再利用できることが望ましい
3. 多数の語彙を複雑に組み合わせるよりは、シンプルな語彙（名前空間）構成のほうが、データを扱う人間にとって把握しやすい

ユーザのタスク（§2.1）に照らしてみると、特に「発見」に用いられるプロパティには広く理解される語彙を用いることが望ましく、「識別」「選択」については的確な表現を重視する必要があるといえるだろう。

この観点から、Dublin Core、Schema.org、BIBFRAME を候補として、基本データモデルをどの程度記述できるかを比較してみる。そのうえで、3. を踏まえて基本記述語彙を一つ選び、不足する部分は原則として独自定義（もしくは DC-NDL を拡張）として、統合プラットフォーム情報の記述語彙を選定する。

なお独自語彙を定義する場合、ツールでの扱いやすさや他語彙との親和性、国際的な利用も考慮し、用語は英単語（の組合せ）での命名を原則とする。

#### 4.1.1 基本プロパティの比較

3 つの語彙で基本項目に該当するプロパティを表 3 に示す（\*印は定義とずれるもの / 提案中）。

**BIBFRAME** 作者も含め寄与者のモデルのみで記述するため、作者に相当する短縮プロパティがない。また出版事項にイベントモデルを採用していることから、発行者の短縮プロパティがないほか、発行に関する時期、地域も異なるモデルとなるなど、基本情報については該当なしの項目がやや多い。

一方、基本モデル案では記述としてまとめている資料体情報は `extent`、`dimensions`、`baseMaterial`、`appliedMaterial`、`bookFormat` など詳細に記述するプロパティが用意されている。

**Dublin Core** 検索に用いる基本項目は、区分を除きカバーできている。画像を示すプロパティはない。

名称に対応するプロパティは `title` となるが、これは作品などの（作者が与えた）名称という受け止めが多いかも知れず、標本などの名称としては違和感が残る可能性がある。また `spatial`、`temporal` は `coverage` のサブプロパティであるため、制作に関する時間、場所の記述に用いるのは不正確（モデルの的確で整合性のある表現ができない）といえる。

**Schema.org** 基本項目は画像も含めカバーしている。対象とする分野も広い。

`name` は、リソース一般の名称としては DC の `title` より違和感が少ないと思われる。`spatial`、`temporal` は現在のところ Dataset の対象地域・時期を示すためのものとして定義されているが、CreativeWork 一般に適用できるよう提案中である。

また資料体記述については、W3C の Schema Archetypes コミュニティ・グループと協力して `physicalDescription`（もしくは `materialExtent`）を提案中である。

表 3: 基本プロパティの比較

基本項目	Dublin Core	schema	BIBFRAME	F	I	S	O
タイプ	rdf:type	rdf:type	rdf:type		-		-
ラベル	rdfs:label	rdfs:label	rdfs:label	-			-
名称	title	name	title				-
寄与者 / 関係			contribution (+role)				-
	contributor	contributor					-
作者	creator	creator					-
発行者	publisher	publisher	(provisionActivity + agent)				-
場所 / 関係							-
	spatial*	location*	place				-
時間 / 関係							-
	temporal*	temporal*					-
	created	dateCreated	creationDate				-
主題	subject	about	subject				-
区分		genre	genreForm				-
識別子	identifier	identifier	identifiedBy			-	
言語	language	inLanguage	language				-
画像		image		-			-
資料体記述	extent	physicalDescription*					-
記述	description	description					-
上位資料	isPartOf	isPartOf	partOf				-
提供情報				-	-	-	-
提供者		provider		-			
リンク	related	url	relatedTo	-	-	-	
オブジェクト				-	-		
権利情報	rights	license		-	-		
個別識別子	identifier	identifier	identifiedBy			-	
ソース情報				-	-	-	-
提供者		provider		-	-	-	
データ		distribution		-	-	-	
リンク	related	url	relatedTo	-	-	-	
更新日	modified	dateModified	changeDate	-	-		

#### 4.1.2 基本記述に用いる語彙

以上の比較をもとに、基本プロパティの記述には Schema.org を採用することを提案する。

Schema.org は W3C の Schema Bib Extend コミュニティ・グループにより書誌関連の拡張が導入され、VIAF や WorldCat の LD 記述に採用されている。さらに前述の Schema Architypes コミュニティ・グループが、アーカイブ全般の記述についての拡張提案を準備している。また、商品、イベント、レシピなど多様な分野の記述が可能で広い範囲で利用されているため、領域を超えたデータの活用に大きく資することが期待できる。

## 4.2 関係モデルの構造化プロパティ

関係モデル(寄与者/関係、場所/関係、時間/関係)のうち、寄与者については既存語彙も記述方法を提供している。たとえば Schema.org はプロパティを反復する Role モデルを示している。BIBFRAME では contribution + role、agent によってそのまま寄与者関係が記述できる。Schema.org の Role モデルは、時間/場所関係にも適用できなくもない。

ただし Schema.org は基本語彙(ショートカット)として用いることにしており、BIBFRAME では時間/場所関係を記述できない。そのため、関係モデルには独自の語彙を用意し(接頭辞 v: で示す)、可能な部分は Schema.org の用語(接頭辞 schema: で示す)で記述することにする(表 4)。

表 4: 関係モデルに用いるプロパティ案

項目	内容	プロパティ
寄与者/関係	制作に関与した人/組織の関係	v:contribution
	寄与のタイプ。領域を超えて一般化できる制作、編集、翻訳など	v:relationType
	寄与した人物・団体	schema:agent
	領域固有の役割名、キャストの配役など	schema:roleName
	関連リンク	schema:url
時間/関係	作成、公開、発見、主題など対象に関する時間	v:temporal
	時間関係のタイプ。制作、主題など	v:relationType
	時間関係における時間(範囲)を示す	v:temporalValue
	" 時代を示す	v:era
	月日ほか補足情報	schema:description
場所/関係	作成、公開、主題など対象に関係する場所・地域	v:spatial
	場所関係のタイプ。制作、主題など	v:relationType
	場所関係における場所を示す	v:spatialValue
	補足記述	schema:description
上位資料/関係	上位資料との関係(掲載誌のページ)などを記述する	v:partOf
	上位資料との関係のタイプ。掲載など	v:relationType
	上位資料関係における上位資料そのものを示す	v:source
	上位資料内の特定部分を示す	v:selector
	補足記述	schema:description

寄与関係には Schema.org の Role のプロパティを利用した。ただし Schema.org では関係の種類を rdf:type で表現するのに対し、ここでは他の関係と合わせるために独自の relationType を用いる。

たとえば映画において監督の `relationType` を「制作」とすることで制作に関する日付や場所とも関連付ける。このとき `roleName` に「監督」を記述し、領域特有の役割表現を示す<sup>29</sup>。

上位資料に関しては、§3.2.9 の例 3 に示した関係の記述のために、部分指定のプロパティを導入する。

#### 4.3 アクセス提供情報とソース情報

アクセス提供情報及びソース情報に用いるプロパティは、表 5 のように定義する<sup>30</sup>。

表 5: アクセス提供情報及びソース情報に用いるプロパティ案

項目	内容	プロパティ
提供情報	資料の提供に関する情報	<code>w:accessInfo</code>
	資料（に関する情報）の提供者	<code>schema:provider</code>
	資料の保管者（提供者と別の場合）	<code>v:contentHolder</code>
	資料の紹介ページやアクセス情報が記載されたページの URL	<code>schema:url</code>
	資料のデジタル画像、音声動画などの URI	<code>v:digitalObject</code>
	資料のデジタル複製物利用のライセンス URI（画像などに対するプロパティ）	<code>schema:license</code>
	テキストによる権利記述	<code>v:contentRights</code>
	提供元が付与する識別子	<code>v:contentId</code>
	補足記述	<code>schema:description</code>
ソース情報	収集したデータ及びその提供元に関する情報	<code>w:sourceInfo</code>
	ソース情報の提供者（アグリゲータ）	<code>schema:provider</code>
	集約元情報の公開ページ URL	<code>schema:url</code>
	プラットフォームが保持・提供するソースデータ	<code>schema:distribution</code>
	ソース情報が Linked Data である場合	<code>rdf:seeAlso</code>
	補足記述	<code>schema:description</code>
	収集元データの更新日、なければアグリゲート日	<code>schema:dateModified</code>

アクセス提供情報は Schema.org の `Service` にやや近いと考えられるが、サービスで提供するもの（画像など）も記述するため、その部分は独自のプロパティを用いる<sup>31</sup>。プロパティ名は、この点についての誤解を与えないよう `content` などを前置したものとす。§3.3.4 の一括権利情報を採用する場合は、`v:aggregatedRights` などが考えられる。

ソース情報は Schema.org の `Dataset` にほぼ準ずるものとしてそのプロパティを用いている<sup>32</sup>が、`distribution` は厳密には独自に定義すべきかもしれない。なおソース情報が Linked Data として提供される場合は、元のデータを直接 `rdfs:seeAlso` で参照してリンクする<sup>33</sup>。

<sup>29</sup>異なる領域の役割を一般化して `relationType` とすることが望ましいが、実際の運用では提供元項目名をそのまま用いるしかなく、`roleName` との使い分けは難しい可能性もある。§??マッピングレベルを参照。

<sup>30</sup>接頭辞 `w:` はデモシステムでの表示順のために使い分けたもので、`v:` と同じ独自語彙を表す。

<sup>31</sup>ここで例えば `schema:image` を用いると、提供情報 / サービスを表す画像を意味してしまうので、独自のプロパティを定義する。

<sup>32</sup>提供データはデータセットではなくその中の 1 レコードであり、`schema:Dataset` そのものではない。また §2.3 で述べたように、ソース情報は EDM の `Proxy` に相当するものとして「アーカイブ情報」クラスを与える。

<sup>33</sup>共通情報の抽出や基本検索のため、Linked Data であってもソースデータは収集し、内部的に保持する。

なお、それぞれのリンク情報に `schema:url` を充てているが、これは各情報リソース自身に対応する URL を意味することになるので、もう少し検討が必要かもしれない<sup>34</sup>。

## 4.4 クラスと概念体系

### 4.4.1 クラス定義

統合プラットフォームにおいては、収集者の視点によるメタデータレコードを表現するクラス（§ 2.3 参照）としてアーカイブ情報を定義する。また共通アーカイブ情報の各レコードにはそのサブクラスである共通アーカイブ情報クラスを用いる。

また資料の基本区分として、各提供データの分類などに基づくクラス（文化財など）をそれぞれ用いる。これらは（可能であれば）マッピング生成時に調整し、全体として整合性のあるものを定義する。

ただし § 3.2.1 でも述べたとおり、「文化財」でありかつ「アーカイブ情報」でもあるというリソースは予備知識なしで利用する場合に混乱を招く恐れがあるので、基本情報に付与する「共通アーカイブ情報」については、`schema:additionalType` を用いて、`rdf:type` とは異なる位置付けとして扱う。

### 4.4.2 関係概念

関係モデルで用いる `relationType` の値は、クラスとして定義せずに `skos:Concept` として扱う。これらをクラスとして扱おうと、たとえば寄与者関係ノードと時間関係ノードを表すクラスはそれぞれ別の基底クラス（寄与者関係クラス、時間関係クラス）のサブクラスとしなければならず、おなじ「制作」関係を通じてノードを結びつけることができなくなるからである。

時間に関しては、§ 3.2.4 で示した時間オントロジーを定義する。

### 4.4.3 主題とキーワード

主題は、可能であれば NDLSH を用いるが、各提供データが独自の主題体系（コード化されるなど URI 化可能なもの）を用いているときは、原則としてそれをそのまま共通情報にも用いる。

文字列のみのキーワードの場合は、§ 3.2.6 で示したように、特別な名前空間を用いて文字列を直接 URI 化する。

なお、場所関係における場所を示す URI については、国内は都道府県、海外は国（もしくは日本十進分類に示されるレベル）を念頭に集約・正規化して URI を付与することを目指す、困難である場合はキーワードと同様に扱う。

## 4.5 Linked Data としての共通アーカイブ情報

統合プラットフォームの共通アーカイブ情報を広く活用できるようにするためには、Linked Data の標準に従ったデータモデル、語彙、URI を採用するだけでなく、LD クラウドのハブとなるようなデータセットとの結びつきが必要である。

出発点としては、主題や著者の記述に Web NDLA を可能な限り用いることで、ここを経由して VIAF、DBpedia などと間接的にリンクしていく。主題などに各機関のシソーラスなどを用いる場合

<sup>34</sup>提供情報リソースの `schema:url` として、資料アクセスの方法が示されたページの URL は適切といえるが、資料提供元の目録ページなどは適当ではないかもしれない。最小名前空間の方針からは外れるが、オブジェクトも含め、EDM の `isShownAt` などの `WebResource` を示すプロパティを利用する方法もある。

も、この視点からできるだけ NDLA や DBpedia との関連付けができるよう、調整を進めることが望まれる。

一方、個々の作品（レコード）については、極めて有名なもの以外は外部にリンクできるリソースがない場合が大半と考えられる。これらはむしろ、日本（関連）の文化資源について言及する場合に、統合プラットフォームの URI が標準的に用いられるようにすることで、日本文化情報のリンクのハブになることを目指す。統合プラットフォームは、単独でメタデータを提供するだけでなく、こうした具体的記述情報と結びついてこそ、その価値を発揮することができる。

#### 4.6 標準語彙との関係

独自定義したプロパティは、RDF スキーマを整備する段階で、Dublin Core など標準的な語彙とのサブプロパティ関係を定義する。