

# ジャパンサーチ利活用フォーマット

1. 基本データモデルの考え方
2. ソース情報分離モデル
3. 共通情報の主要項目
4. 扱いやすさと適切さのバランス
5. 寄与者関係の共通情報表示例
6. 時間関係情報
7. 時間関係の共通情報表示例
8. 場所関係情報
9. 場所関係の共通情報表示例
10. 提供=アクセスのための情報
11. 提供/ソース関連の共通情報表示例
12. ライセンス情報の明示
13. 共通情報整備の課題

DA実務者検討委員会  
2018-01-22 神崎正英

## 1 基本データモデルの考え方

### ■ ソース情報の分離と利用者のタスクに基づく共通情報

- 提供元からのソースデータは来歴情報明記の上でそのまま保持
- 利用者の4つのタスクの観点で共通情報を生成
  - ▶ **Find**：発見=タイトル、キーワードなどによる検索（標目、アクセスポイント）
  - ▶ **Identify**：識別=示されている対象が何なのか、求めている（既知の）コンテンツかどうか判断できる情報
  - ▶ **Select**：選択=（検索結果などの）選択肢のうち、対象が求めている（未知の）コンテンツかどうかを判断できる情報。識別より広い意味での記述
  - ▶ **Obtain**：取得=示されている（選択した）対象を取得する手段（の情報）の提供

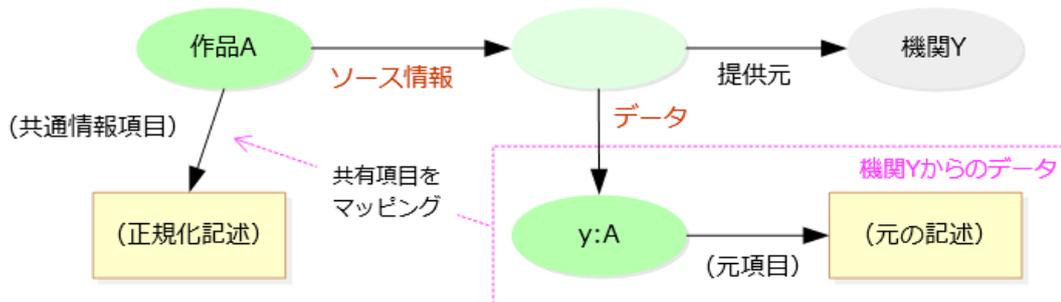
### ■ 共通情報とタスク

- いつ、どこで、だれが、何を（→特に発見タスク）を基本にシンプルな項目設定
- 単純プロパティと構造的プロパティの併置（識別、選択タスク）
- 提供=アクセス情報の充実（取得タスク）
- アプリケーションによる活用も踏まえた情報形態（Linked Data）

## 2 ソース情報分離モデル

### ■ 共通情報とソースデータ/情報

- 利用者タスクに基づき、共通利用や付加価値を持つ再利用のための項目をマッピング
- 共通情報項目の1つとして「ソース情報」を持ち、ソースデータをそのまま保持
  - ▶ ソースデータはJSONに変換して保持し、単純検索、全文検索などでも利用



### ■ Europeana、DPLAとの関係

- Europeana、DPLAともに収集データとそれを整理した共通データを分離して扱っている
  - ▶ Europeanaは提供元に一定の標準モデル準拠を求める。DPLAは複数形式を許容
- Europeanaは値の正規化、Linke Data化を重視。ジャパンサーチはより多様な活用が可能なモデルを工夫

## 3 共通情報の主要項目

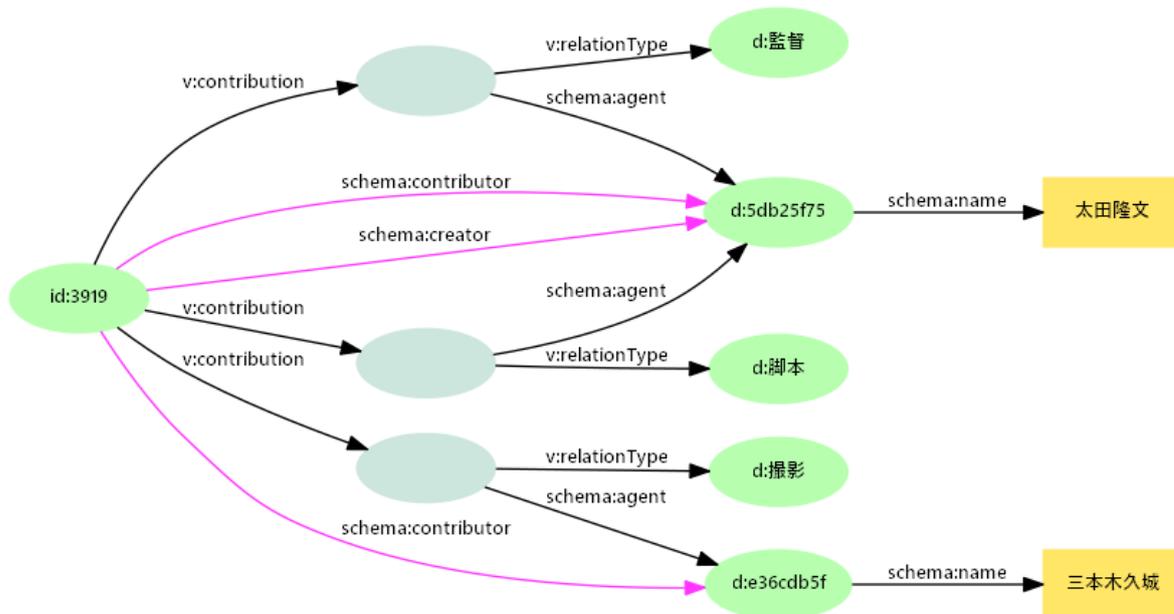
### ■ 共通情報の基本項目

| 基本項目    | 内容                                  |
|---------|-------------------------------------|
| タイプ     | コンテンツの基本区分（図書、絵画、標本など）              |
| 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   | コンテンツに寄与した人／組織（作者、発行者、出演者など）        |
| 場所関係    | 場所に関する情報（発行地、制作地など）                 |
| 時間関係    | 時間に関する情報（制作年、対象時期など）                |
| 主題／区分   | 主題・分類／カテゴリー                         |
| 識別子、言語  | ISBNなどの識別子、コンテンツの記述言語               |
| サムネイル画像 | コンテンツの特徴を確認するための画像（提供元とは別にサムネイルを保持） |
| 記述      | コンテンツの物理的特徴・素材等の記述、個別項目に収録できない情報    |
| 上位コンテンツ | 現在のコンテンツがその一部である上位コンテンツ（公文書などの資料階層） |
| 提供情報    | コンテンツの提供・アクセス、ライセンスに関する情報           |
| ソース情報   | ソースデータおよびその提供者に関する情報                |

# 4 扱いやすさと適切さのバランス

## ■ 単純プロパティと構造化プロパティの併用

- 監督、脚本=太田隆文；撮影=三本木久城というメタデータがある時
- **単純プロパティ**：作者は共通のプロパティ creator で調べたい。関与した人は一括して contributor で調べたい（発見） → 原則として Schema.org を使用
- **構造化プロパティ**：監督、照明といった元データの項目情報も生かしたい（識別、選択） → 必要に応じて独自語彙



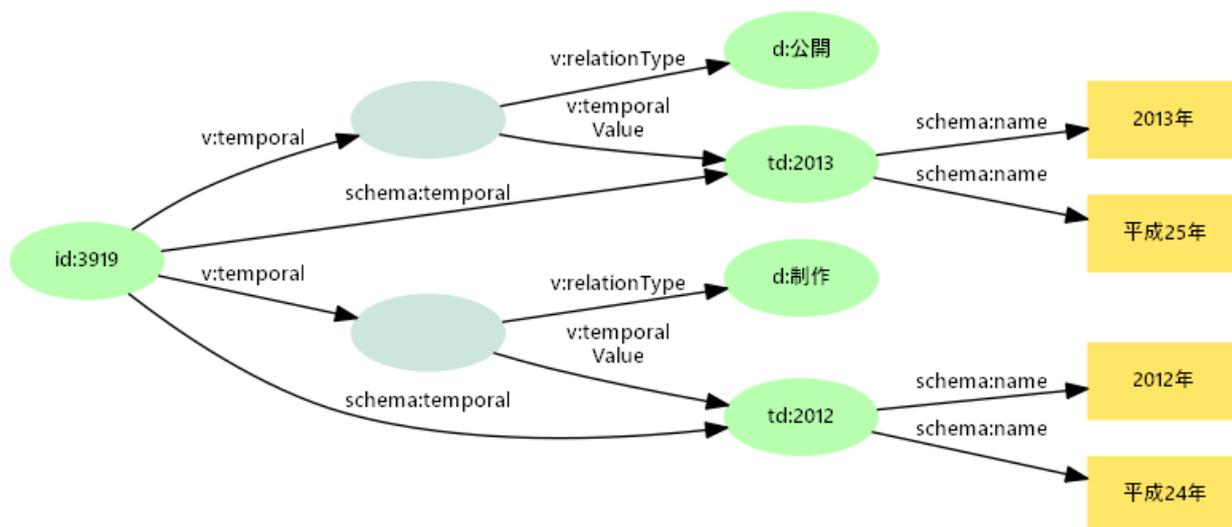
# 5 寄与者関係の共通情報表示例

|                |                 |            |                |                                 |
|----------------|-----------------|------------|----------------|---------------------------------|
| v:contribution | schema:roleName | 「撮影」       | schema:agent   | ndlent:00444960 (☞ James Joyce) |
|                | schema:agent    | d:今村互 ☞    | v:relationType | d:制作 ☞                          |
|                | v:relationType  | d:staffs ☞ |                |                                 |
| v:contribution | schema:roleName | 「編集」       | schema:agent   | ndlent:00040168 (☞ 丸谷オー)        |
|                | schema:agent    | d:太田隆文 ☞   | v:relationType | d:翻訳 ☞                          |
|                | v:relationType  | d:staffs ☞ |                |                                 |
| v:contribution | schema:roleName | 「平田俊夫」     | schema:agent   | ndlent:00049772 (☞ 永川玲二)        |
|                | schema:agent    | d:並樹史朗 ☞   | v:relationType | d:翻訳 ☞                          |
|                | v:relationType  | d:casts ☞  | schema:agent   | ndlent:00078566 (☞ 高松雄一)        |
| v:contribution |                 |            | v:relationType | d:翻訳 ☞                          |
|                |                 |            | schema:agent   | d:日本障害者リハビリテーション協会 ☞            |
|                |                 |            | v:relationType | d:出版 ☞                          |

## 6 時間関係情報

### ■ 「いつ」に関する情報を集約する

- 制作、収集、内容（対象時代）を区別せずに時間関係として扱う（発見）
- 構造化プロパティで関係の違いを示す（識別、選択）
  - ▶ 値は年を単位に正規化し、元記述が月日まで含むなどの場合は構造化情報に保持する



- 時間を文字列ではなくURIで表し、西暦と和暦、時代区分の対応などを「時間オントロジー」としてまとめる
  - ▶ 時間範囲も `td:1868_1913` のようなURIを介して開始年、終了年を構造化することで、単一年の値と同様に処理できる（DPLAでも時間範囲は構造化）

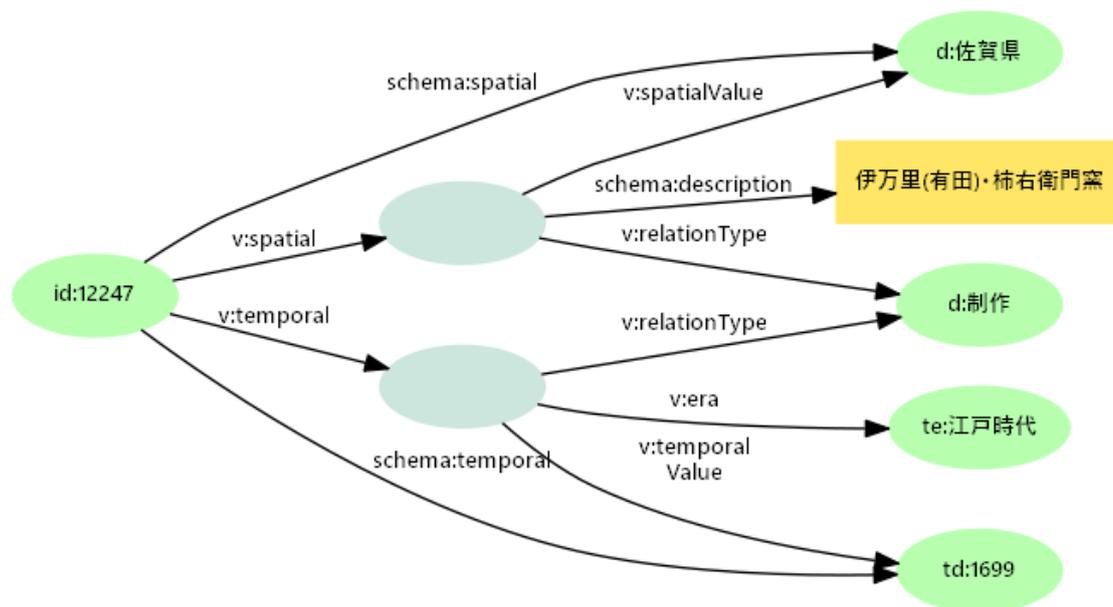
## 7 時間関係の共通情報表示例

|                    |                 |                     |
|--------------------|-----------------|---------------------|
| v:temporal (2) ☒   | v:relationType  | d:公開 ☒              |
|                    | v:temporalValue | td:2013 ☒ (⇨ 平成25年) |
| v:temporal         | v:relationType  | d:制作 ☒              |
|                    | v:temporalValue | td:2012 ☒ (⇨ 平成24年) |
| v:temporal         | v:era           |                     |
|                    | v:temporalValue | td:1903 ☒ (⇨ 明治36年) |
| te:明治時代 ☒          |                 |                     |
| schema:name (2) ☒  |                 | "1868~1913年"        |
| schema:name        |                 | "明治時代"              |
| owl:sameAs         |                 | td:1868_1913 ☒      |
| schema:description |                 | 放送開始日:2013/12/29 ☒  |
| v:temporal         |                 |                     |
| v:relationType     |                 | d:放送 ☒              |
| v:temporalValue    |                 | td:2013 ☒ (⇨ 平成25年) |

## 8 場所関係情報

### ■ 「どこで」に関する情報を集約する

- 時間軸と同じく、空間（場所）関係を一括して扱う（発見）
- 関係の違いは構造化プロパティで示す（識別）
- 同じ関係（たとえば「制作」）により、同じ関係を持つ時間、場所、関与者などが結びつく
- 場所を都道府県単位（国単位）で正規化→地域活性化のためのデータ利用



## 9 場所関係の共通情報表示例

|           |                         |                           |        |
|-----------|-------------------------|---------------------------|--------|
| v:spatial | v:relationType          | d:制作                      |        |
|           | v:spatialValue          | d:伊万里(有田)・柿右衛門窯           |        |
| v:spatial | v:relationType          | d:出土                      |        |
|           | v:spatialValue          | d:佐味田宝塚古墳 (奈良県北葛城郡河合町) 出土 |        |
| v:spatial | schema: description (2) | 採集場所：都道府県 (日本語) :静岡県      | d:関連地名 |
|           | schema: description     | 採集場所：詳細 (日本語) :梨本 南伊豆     | d:加賀   |
|           | v:relationType          | d:採集                      | d:関連地名 |
|           | v:spatialValue          | d:日本                      | d:越中   |
| v:spatial | v:relationType          | d:関連地名                    |        |
|           | v:spatialValue          | d:能登                      |        |

# 10 提供 = アクセスのための情報

## ■ コンテンツにアクセスするための情報

- 提供者（保管者）の情報
- 提供者のコンテンツ概要ページなどへのリンク
  - ▶ アクセス情報とソース情報の提供者は、一致する場合もあれば、異なる場合（つなぎ役がソース情報提供者）もある
- 提供者においてコンテンツを確認するための識別子（請求記号など）

## ■ デジタルコンテンツにアクセスするための情報

- 画像や動画など、コンテンツをデジタル化したもののURL
  - ▶ デジタルコンテンツURLは複数となる（さらに提供者も複数）場合もある
- デジタルコンテンツ利用のための技術情報（フォーマットなど）
- デジタルコンテンツ利用のためのライセンス情報

# 11 提供/ソース関連の共通情報表示例

|                     |                           |   |
|---------------------|---------------------------|---|
| w:accessInfo        | schema:url                | < <a href="http://www.narahaku.go.jp/collection/225-0.html">http://www.narahaku.go.jp/collection/225-0.html</a> >   |
|                     | schema:description        | "機関管理番号:225-0"  |
|                     | v:contentHolder           | d: <a href="#">奈良国立博物館</a>  |
|                     | v:digitalObject           | < <a href="https://colbase.nich.go.jp/uploads/collection_item_images/thumbnail/feb2ffccff8c589f119dd60ee8c5ebcf.jpg">https://colbase.nich.go.jp/uploads/collection_item_images/thumbnail/feb2ffccff8c589f119dd60ee8c5ebcf.jpg</a> > |
| w:sourceInfo        | id:p14813                 |   |
|                     | rdf:type                  | c: <a href="#">アーカイブ情報</a>  |
|                     | schema:provider           | < <a href="https://colbase.nich.go.jp/">https://colbase.nich.go.jp/</a> > ( <a href="#">ColBase:国立博...</a> )  |
|                     | schema:url                | < <a href="https://colbase.nich.go.jp/collectionItems/view/5b7114ef7684b844c73356fa61c11915/225-0">https://colbase.nich.go.jp/collectionItems/view/5b7114ef7684b844c73356fa61c11915/225-0</a> >                                     |
|                     | schema:dateModified       | "1970/01/01"  |
| schema:distribution | source:colbase-14813.json |   |

---

# 12 ライセンス情報の明示

---

## ■ コンテンツ自身のライセンス

- メディア芸術におけるライセンス記述の標準化と一括提供

## ■ デジタル化したコンテンツのライセンス（案）

- 提供元毎などに標準ライセンスを決めて一括記述
  - ▶ 統合プラットフォームで情報が得られるものは原則としてCC0にするなど
- 特別なライセンスが必要な例外のみ、オブジェクトの情報として記述
  - ▶ ブラウザなどでの表示の際には、標準ライセンスの場合もアイコンなどで明示する

## ■ メタデータに関するライセンス

- 共通情報はすべてCC0
- ソースデータも統合プラットフォームで提供できるものはCC0

---

# 13 共通情報整備の課題

---

## ■ データの品質と正規化

- 提供元データの記述方法のばらつき、表記の揺れについて、どこまで厳密に正規化するか  
→ 発見タスク、特に「いつ」「どこで」「だれが」については可能な限り正規化し、URIとして扱う。Linked Dataも念頭に
- 識別、選択情報は原則として正規化せず、ソースデータを生かす
  - ▶ プロパティも無理にマッピングせず、汎用プロパティに導入句付きの値で
- 単一フィールドに複数值が入ると正確な処理が困難
  - ▶ 区切り文字による複数值、（）を用いた読み、URLと表示名の併記など

## ■ 利活用のための提供データの調整と補完

- 名前（ラベル）は必須。提供されない場合はIDなどから生成する
- コンテンツの基本区分である**タイプ**は、選択のために必須。あらかじめ定義したタイプが提供される方向へ調整したい
- ライセンスは、提供元ごとに基本（デフォルト）ライセンスを設定したい
- 場所情報はサンプルデータでは提供例が少ないが、利活用に有益。提供機関の所在地を適用することも