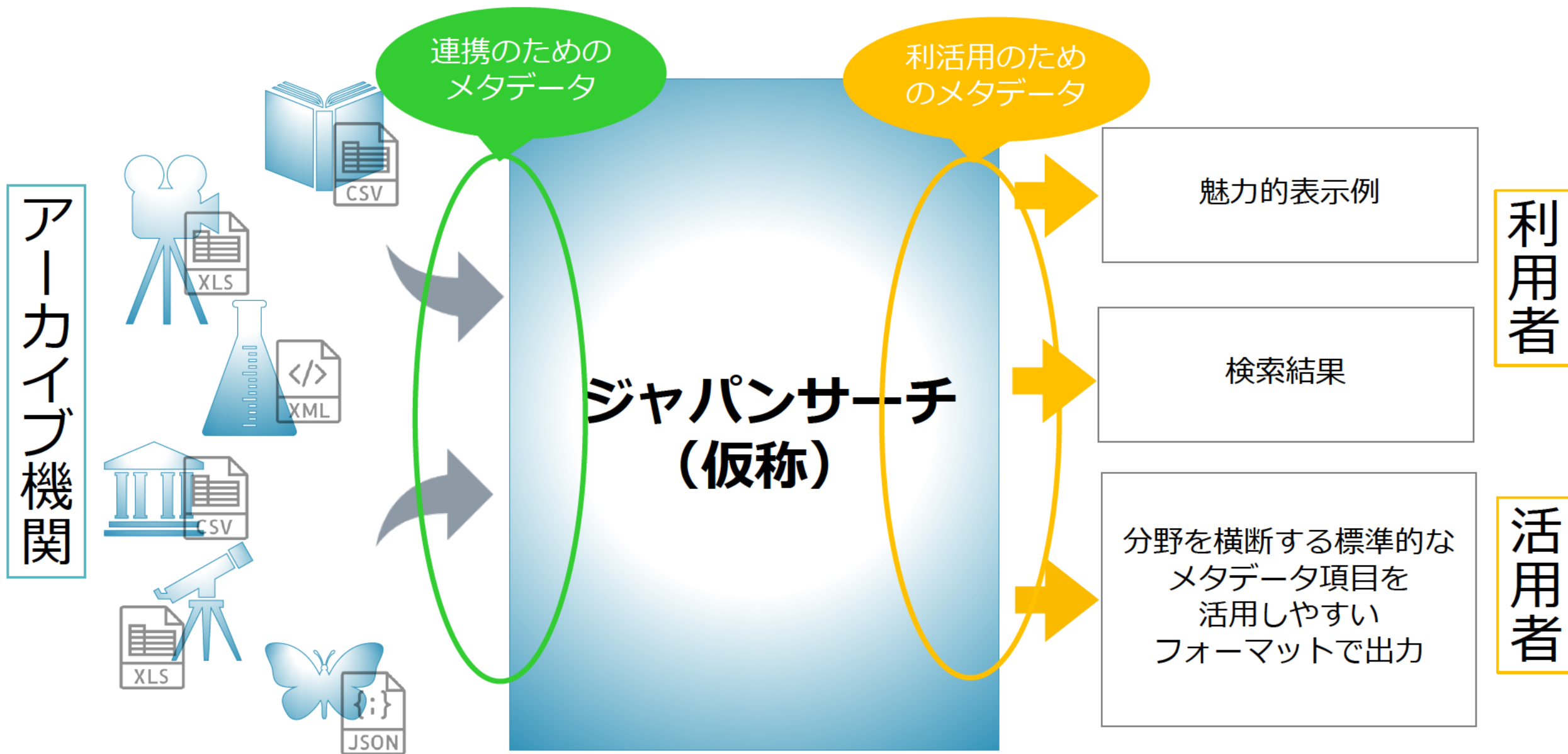


# 標準メタデータフォーマットの検討について

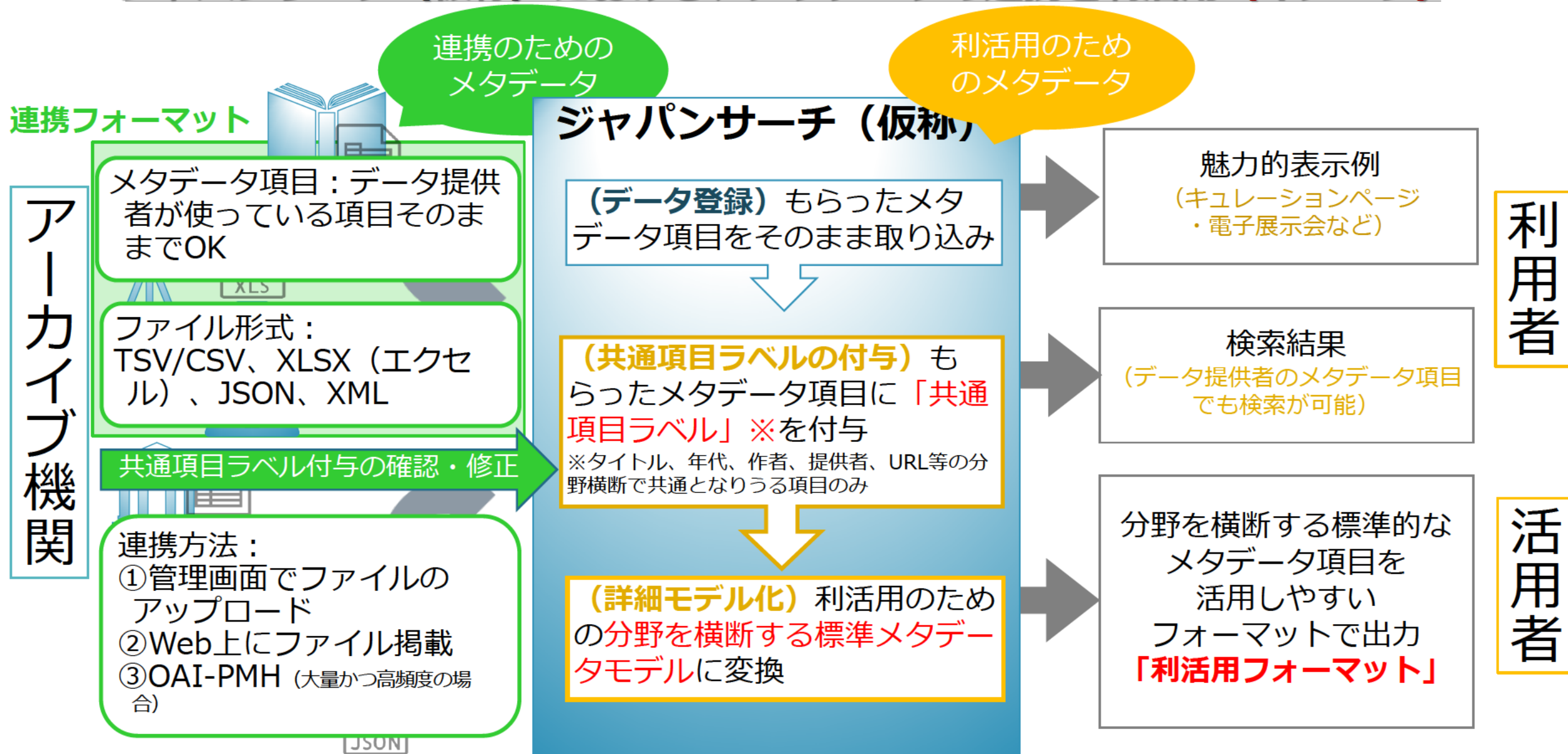
---

国立国会図書館 電子情報部

## ジャパンサーチ（仮称）におけるメタデータの連携と利活用のために



# ジャパンサーチ（仮称）における、メタデータの連携と利活用（イメージ）



# 標準メタデータフォーマットの検討状況と進め方

## ◆連携のためのメタデータフォーマット（連携フォーマット）

⇒国立国会図書館から現時点の想定を報告します

## ◆利活用のためのメタデータフォーマット（利活用フォーマット）

⇒神崎正英 ゼノン・リミテッド・パートナーズ代表からこれまでの検討内容をご発表いただきます

### 今後のスケジュール

- 2月～3月：関係機関等へのヒアリング
- 3月末頃：見直した内容の連携フォーマット及び利活用フォーマットを提示
- 平成30年度第1回（4月頃）の実務者検討委員会に、ヒアリングの結果を踏まえた最終案を提示
- 平成30年4～5月「第一次中間取りまとめ」の資料に含めて紹介

# 連携フォーマット（案）について

---

# 基本的な考え方

## ◆連携フォーマット

- 連携のための単一のメタデータモデルは定義をせず、データを提供する連携機関のメタデータモデルをそのままの形（オリジナルのモデル）で受け入れる。
- 共通項目ラベルを付与することで、分野横断的な表示・検索を可能にする。

## ◆連携仕様

- 原則、ファイルベースでの連携を行う（OAI-PMHを必須としない。）
- 多様なファイル形式に対応する。

## ◆利活用フォーマット

- メタデータの利活用促進のため、分野を横断する標準的なメタデータモデル（利活用フォーマット）を定義する。
- メタデータは、オリジナルモデルと利活用フォーマットの両方で保持する。



多様なデータ形式に対応、連携機関の作業負担の軽減を図りつつ、充分利活用できる仕組みを用意する

# データ連携の流れ（イメージ）

## アーカイブ機関側

※赤字がアーカイブ機関の行う作業

### ①データベース定義

データベースの基本情報の**提供**/管理画面から入力

### ②データ登録

管理画面からファイルを**アップロード**/ファイルを**Webに掲載**

### ③ラベル定義

- ・ 共通項目ラベル候補の**確認・修正**
- ・ 個別項目ラベルを**定義**（オリジナルのメタデータの各項目の名称、データ形式、説明等の情報提供）

テスト環境で確認、必要に応じて修正

## ジャパンサーチ（仮称）

（情報をもらってこちら側で登録も可能）

- ・ 提供されたメタデータをそのまま登録
- ・ メタデータアナライザーでデータ解析
- ・ **共通項目ラベル候補**を提示

**共通項目ラベル**：提示された候補の確認のみ  
**個別項目ラベル**：入力画面でExcel等を受け付けて画面で入力する必要を無くす、元データから簡単にコピペできるようにする等の省力化を検討中

公開

- ・ **利活用フォーマット**への変換（マッピング）

## ◆データベース定義が必要な項目

|      |            |                        |
|------|------------|------------------------|
| 必須項目 | ID         | データベースのID（ジャパンサーチ側で付与） |
|      | 名称（日/英）    | データベースの名称              |
|      | タイプ/カテゴリ   | データベースが扱うコンテンツの種別。選択式  |
|      | メタデータの権利表示 | メタデータの権利情報             |
|      | コンテンツの権利表示 | 対象のデジタルデータの権利情報        |
|      | メタデータの権利区分 | 選択式のメタデータの権利情報         |
|      | コンテンツの権利区分 | 選択式のコンテンツの権利情報         |
| 任意項目 | 組織名（日/英）   | データベースの所有者の名称          |
|      | 説明（日/英）    | データベースの説明              |
|      | URL        | データベースのURL             |
|      | 組織URL      | データベースの所有者のURL         |

### 要検討★権利表示

○メタデータの権利表示は必要か？（CC0以外の場合のみ記載？）

○権利表示での検索（絞込み）を可能にするには、「権利区分」に選択式のデータ項目が別途必要と考えるが、どのような選択式の区分が望ましいか？

### 要検討★タイプ/カテゴリの区分

検索結果の範囲指定（絞込み）等への活用を想定。どのような区分が望ましいか？

- **(分野)種別**：書籍、文化財、放送番組、データセット…
- **主題**：歴史/地方誌、美術、政治/法律/行政、自然科学/科学技術…

## ◆メタデータ項目

メタデータ項目は自由だが、最低限の必須項目（ID、タイトル等）がある（⇒詳細は「③ラベル定義」参照）

## ◆連携方式 ※OAI-PMHは、大量かつ高頻度の場合に想定

|             |          |                                                           |
|-------------|----------|-----------------------------------------------------------|
| ファイルのアップロード |          | 管理画面から、手動でファイルをアップロードする方法                                 |
| ファイルをWebに掲載 | ファイル取得   | Web上にファイルを掲載し、管理画面でそのURLを指定すると、ジャパンサーチ（仮称）側がファイルを取得しに行く方法 |
|             | ファイル定期取得 | 上記のファイル取得を、指定した日時に自動実行し、データを自動更新する方法                      |

## ◆対応ファイル形式

|         |                                     |
|---------|-------------------------------------|
| TSV・CSV | 文字コードはUTF-8。同じデータ項目であれば複数ファイルも可。    |
| XLSX    | 複数シート・結合セルには対応できない。XLS非対応。          |
| JSON    | 1行1レコードのJSON Lines形式を推奨。他の構造にも対応予定。 |
| XML     | 1行1レコード形式を推奨。他の構造にも対応予定。            |

## ◆ 共通項目ラベルを付与する項目

|          |            |                        |
|----------|------------|------------------------|
| 必須項目     | ID         | オリジナルデータ内での一意のID（管理番号） |
|          | 名称/タイトル    | レコードの名称。検索結果表示に使われる    |
| あれば必須の項目 | 名称/タイトルヨミ  | レコードの名称の読み             |
|          | 名称/タイトル英語  | レコードの英語名称又はローマ字        |
|          | 最終更新日      | データの最終更新日（日付型）         |
|          | URL        | レコードのリンク先のURL          |
|          | サムネイル画像URL | サムネイル画像のURL            |
|          | 提供者        | オリジナルのコンテンツの提供者        |
| 任意項目（推奨） | コンテンツの権利表示 | データベース定義の情報と異なる場合のみ    |
|          | 寄与者（人物）    | 作者、発行者、出演者など           |
|          | 時間（日付）     | 制作年、対象時期など             |
|          | 場所         | 発行地、制作地など              |

※なお、個別項目ラベルの定義では、オリジナル（ソース）データの各項目の名称や説明、データ形式、検索対象とするかどうかといった情報提供のみ。

# (参考) ジャパンサーチ (仮称) における、データ変遷 (イメージ)

## アーカイブ機関

## ジャパンサーチ (仮称)

