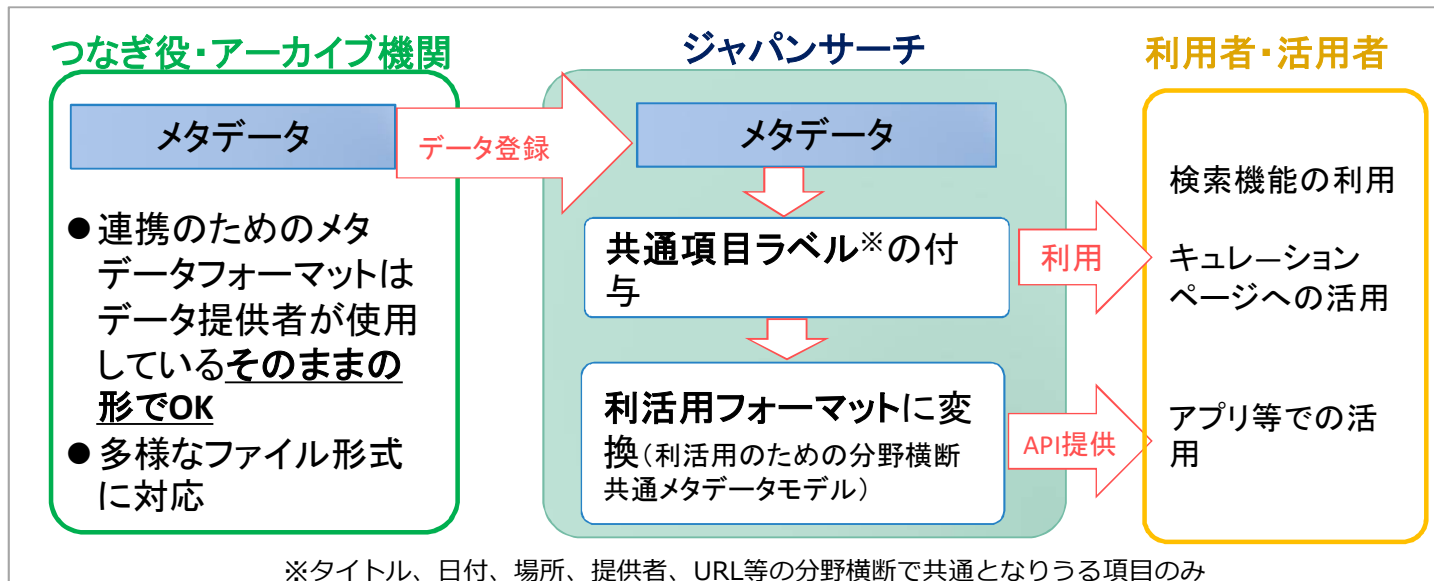


ジャパンサーチのメタデータ連携について

2019年2月

ジャパンサーチ(試験版)の連携方法

◆ データ連携の流れ (データ変遷イメージは参考3参照)



◆ 二次利用条件の設定

➤ メタデータは原則CC0

※著作物性のあるもののみCC BYでも可。民間機関のメタデータなど原則に対応できない場合はデータベースごとに利用条件が分かるよう設定可能

※活王者には、CC0であっても出典等の情報を明記するようお願いする予定

➤ データベースごとにコンテンツの利用条件を設定可能

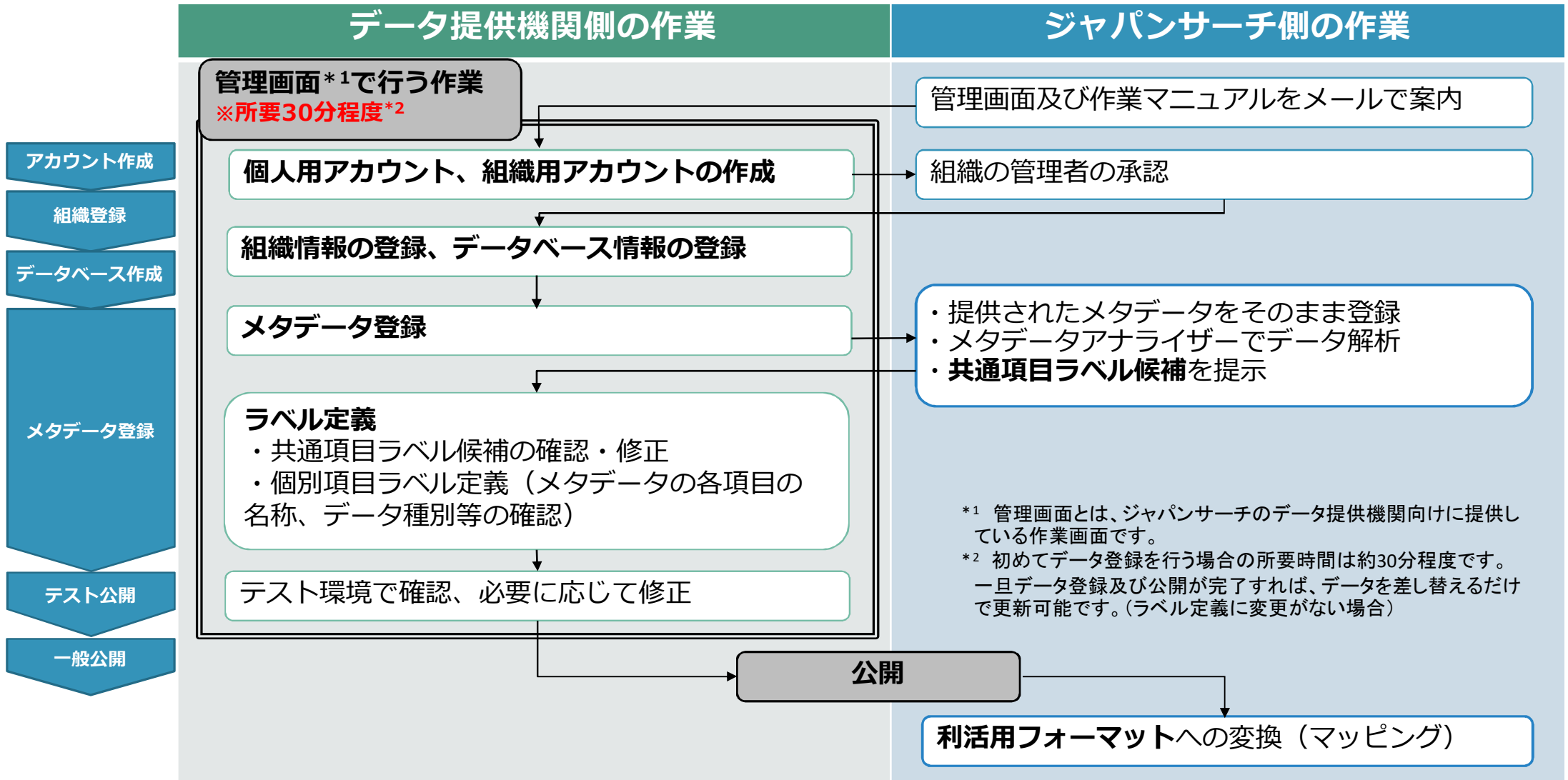
※コンテンツの利用条件がデータベース単位の設定と異なる場合は、メタデータごとにも権利表示を設定可能

国際的流通を意識 (Europeanaと同レベル)

(試験版の管理画面)

The screenshot shows the JAPAN SEARCH management interface. The left sidebar contains a menu with options: アカウント, データベース, DB詳細, 権限管理, データ登録, ラベル定義, テスト公開, 検索定義, 検索テスト, 一般公開, 検索, データ編集, and DBページ. The main content area displays the 'データベースの基本情報' (Basic Information of the Database) for '国立国会図書館デジタルコレクション(PDM)'. It includes fields for ID (dignl), データ登録 (Data Registration), and サブカテゴリ (Subcategory). A blue callout box titled 'データ登録の流れ' (Data Registration Flow) lists the steps: 1. 組織情報登録, 2. データベース定義, 3. メタデータ登録, 4. ラベル定義, 5. テスト公開、検索テスト, 6. 一般公開. The bottom right shows the 'メタデータの権利表示' (Metadatas Rights Display) section with a dropdown menu set to '自由な二次利用可' (Free for secondary use).

メタデータ連携の流れ



試験版の管理画面

JAPAN SEARCH

管理 / データベース / 全国書誌

全国書誌

ステータス 基本情報

データ件数	5724340
管理組織	国立国会図書館
公開状態	公開

現在公開中のデータ

名前	収集日	収集方法	フォーマット	状態
定期収集-20190316	2019/03/16 01:00:05	OAI-PMH	XML	一般公開終了しました

詳細

現在テスト公開中のデータ

名前	収集日	収集方法	フォーマット	状態
定期収集の即時実行-20190304	2019/03/04 16:50:16	OAI-PMH	XML	一般公開終了しました

詳細

データベース管理メニュー

DBトップ

データ登録

ラベル定義

テスト公開

検索定義

検索テスト

一般公開

検索

選択非公開

DBページ

ギャラリー

テーマ別検索

連携機関

連携データベース

使い方

管理画面でできる操作

- 組織情報の登録
- 組織メンバー管理・権限管理
- データベース登録
- カスタム検索の作成 等

1. データベース定義


JAPAN SEARCH
BETA


国立国会図書館デジタルコレクション(PDM)

ステータス 基本情報

ID
 dignl

データベース名
 国 表 名
 国立国会図書館 デジタルコレクション(PDM)

データベースの説明
 国 表 名
 著作権保護期間満了資料

カテゴリ
 書籍学 × +

サブカテゴリ
 +

メタデータの権利表示
 (空欄)

サムネイル画像の権利表示
 (空欄)

コンテンツの権利区分
 自由な二次利用可 ▼

コンテンツの権利表示
 (空欄)

コンテンツ公開状況
 (空欄) ▼

データベースのURL
 (空欄)

代表画像

◆ データベースの基本情報・権利表示の設定

必須	ID	データベースのID	各機関で付与。ジャパンサーチ上のデータベース紹介ページのURLに使用
必須	データベース名（日/英）	データベースの名称（日本語/英語）	
必須	データベース名（ヨミ）	データベースの名称のヨミ（全角カナ）	DB一覧の表示順に使用
	データベースの説明（日/英）	データベースの説明（日本語/英語）	100字以上は折り畳み表示
必須	カテゴリ	データベースが扱うコンテンツの種別。選択式	データベース当たり一つが推奨だが、複数選択可
	サブカテゴリ	データベースが扱うコンテンツの種別。選択式のカテゴリを補うための情報。テキストによる自由記入。	
	メタデータの権利表示	メタデータの権利情報/二次利用条件について、例外の場合のみ記載。原則はCC0（著作権のある項目のみCC BYでも可だが、その場合も項目名とCC BYの記載が必要。）	自由記入。クリエイティブ・コモンズライセンス、政府標準利用規約等。権利情報について記述されている外部資源へのリンク（URL）でも可
	サムネイル画像の権利表示	対象のサムネイル画像がある場合のその権利情報/二次利用条件	
	コンテンツの権利表示	対象のデジタルデータがある場合の権利情報/二次利用条件。例外がある場合は（メタデータの）共通項目で定義することが可能。	
	コンテンツの権利区分	コンテンツの権利情報/二次利用条件だが、検索・絞込み用に選択式になっている	
	データベースのURL	データベースのURL	
	提供機関（データベースの管理者）	データベースの管理者（オーナー）が自動で入る	
	コンテンツ公開状況	デジタルコンテンツのアクセス範囲を選択式（「ウェブ公開」「限定公開」「デジタルコンテンツなし」）で設定。アイテムにより公開状況が異なる場合は（メタデータの）共通項目で定義することが可能。	
	代表画像	データベースの代表画像	

連携フォーマット ②メタデータ登録（1）

 JAPAN SEARCH
BETA

データの登録・収集

登録・収集されたデータ 単発登録・収集 定期収集

登録方法

☒ アップロード ☐ HTTP ☐ OAI-PMH

ファイルの名前

ファイルの選択

ファイルをドラッグ&ドロップするか、クリックして選択してください。

ファイルフォーマット

☐ CSV ☐ TSV ☒ XLSX ☐ JSON ☐ XML ☐ RDF

ヘッダの扱い

ラベルに利用

圧縮の有無

無し

◆ 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）がある
（⇒詳細は「③ラベル定義」参照）

◆ メタデータ登録方法

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。 ※アップロード可能なファイルサイズは最大1GB。 これを超える場合は、ファイル取得を利用するか、ファイルを圧縮してアップロードする（多重圧縮は不可）。
ファイルをWebに掲載	ファイル取得（HTTP）	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

※OAI-PMHは、大量かつ高頻度の場合に想定

連携フォーマット ②メタデータ登録(2)

◆ メタデータ項目として対応可能な値

文字列	文字列としての検索が可能。改行については、「¥n」にエスケープされ、UIに表示される際にはBRタグに変換される。項目名の最後に「_s」が追加される。
HTML	文字列のヴァリエーションで、UIで表示される時にHTMLとして表示される。XSSを防ぐため、タグ・属性は「H、P、BR、DIV、SPAN、TABLE、TR、TH、TD、attr:style」しか利用できない（データ登録時に消去される）。項目名の最後に「_h」が追加される。
URL	リンクとして表示されるURL。項目名の最後に「_u」が追加される。
日付	日付として変換される（UnixTimeとして保持する）。変換に失敗した場合、nullになる。項目名の最後に「_d」が追加される。
真偽値	真偽値。JSONのtrue/falseに対応。「true」「TRUE」「1」がtrueとして、それ以外の全ての値はfalseとして解釈される。項目名の最後に、「_b」が追加される。
数字	数字（小数点含む）。JSONのnumberに対応。数字に変換できないエラーデータが入っている場合、0が入る。項目名の最後に、「_n」が追加される。


◆ 対応ファイル形式と制約

※XLSX以外のファイルでは、文字コードはUTF-8の使用を推奨

TSV・CSV	CSVの場合、RFC4180準拠であること。(")で囲まれた範囲で改行を含んでもよい。TSVについても、エスケープ等はCSVに準拠する。
XLSX	複数シート・結合セルには対応できない。XLS非対応。セルの表示形式は「標準」を推奨。
JSON	1行1レコードのJSON Lines形式を推奨。ただし、ルートを実体として、1レコード1オブジェクトとする形式、ルートを実体として、key:ValueのValueを実体とする形式にも対応可能とする。
XML	1行1レコード形式を推奨。ただし、XPath等で、ルートの下に複数のレコードを続けていく方式にも対応可能。

連携フォーマット ③ラベル定義

◆ 共通項目ラベルを付与する項目



JAPAN SEARCH
BETA
共通項目ラベル

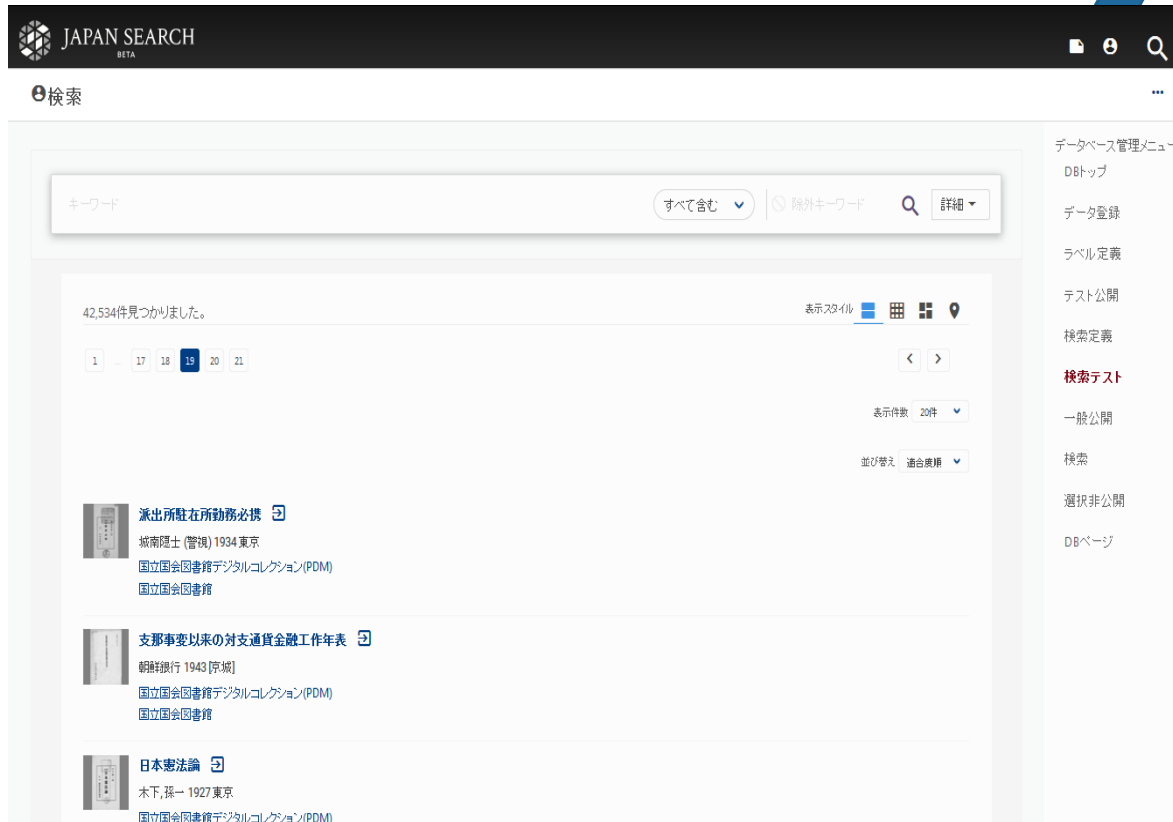
ID	必須	MARC番号(リテラル)
名称/タイトル	必須	タイトル
名称/タイトル英語	あれば必須	
名称/タイトルヨミ	あれば必須	
最終更新日	あれば必須	
URL	あれば必須	URL(デジタル)
サムネイル画像URL	あれば必須	URL(サムネイル)
IIIFマニフェストURL	あれば必須	
提供者	あれば必須	
提供者URL	任意	
コンテンツの権利	あれば必須	
コンテンツの権利区分	あれば必須	

必須	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	名称/タイトルヨミ	レコードの名称の読み（カタカナ・平仮名問わない）
	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（つなぎ役と提供元の両方共）
	サムネイル画像URL	サムネイル画像のURL
	IIIFマニフェストURL	IIIFマニフェストのURL
	提供者	オリジナルのコンテンツの提供者
	提供者URL	オリジナルのコンテンツの提供者の紹介URL
	コンテンツの権利区分	データベース定義の情報と異なる場合のみ。選択式
任意	コンテンツ公開状況	データベース定義の情報と異なる場合のみ。選択式
	座標	座標
	場所	発行地、制作地など（複数可）
	時間/時代	制作年、対象時期など（複数可）
	人物/団体	作者、発行者、出演者など（複数可）

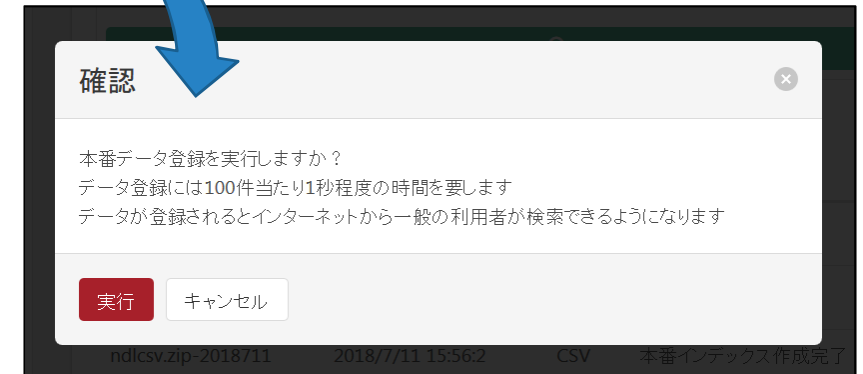
◆ 個別項目ラベルの定義

（オリジナル）ソースデータの項目の名称、データ形式等の定義の確認を行うのみ。必要に応じて項目の説明の追加が可能。

連携フォーマット ④検索テスト、一般公開



検索テスト画面



必要に応じて修正後、公開

(参考)データ登録作業イメージのデモ動画
<http://www.ndl.go.jp/jp/event/events/201805jps.html>

2018年5月開催の説明会「ジャパンサーチ・フェーズゼロ～分野横断統合ポータル構築に向けて」でご紹介したもの

利活用フォーマット ①概要

◆ 目的

- ✓ ジャパンサーチに登録された多種多様なコンテンツのメタデータを共通の形式に変換して提供することにより、より精緻な検索を可能にすること。また、分野横断的な利活用（付加価値をもたらす二次利用）を促進すること。

※検討に当たっては、Europeana 等の海外事例も踏まえつつ、シンプルで使いやすいものであると同時に必要十分な情報を提供するためにはどのようなモデルがよいかといった視点に留意した

◆ データモデル

- ✓ 提供元からのソースデータを来歴情報明記の上でそのまま保持
- ✓ ソースデータのうち、利用者の4つのタスク（発見・識別・選択・取得）に特に有益な項目を共通の形式に変換した共通情報を新たに生成し保持

◆ 提供

- ✓ 必要十分な情報を、平易な構造で、汎用性のある形式で ファイル出力
- ✓ 詳細画面で共通情報とソースデータの両方を 表示

利活用フォーマット ②共通情報の基本項目

「いつ」、「どこで」、「だれが」、「何を」を基本に項目設定。

基本項目	内容
タイプ	コンテンツの基本区分（書籍、文化財など大きく情報を区分する枠）
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／各分野のキーワード的共通認識のある区分（国宝、ドキュメンタリーなど）
識別子	コンテンツを特定するための識別子（ISBNなど）
言語	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する場合）
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	当該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）

基本項目	内容
提供情報	コンテンツにアクセスするための情報。下記サブ項目からなる
提供者	コンテンツ（に関する情報）の提供者を識別するURI。保管者が別であればURIも。
リンク	コンテンツ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	コンテンツのデジタル画像や音声・動画のURI
権利情報	コンテンツの利用に関するライセンス及び権利に関する情報（ライセンスのURIなど）
個別識別子	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）
ソース情報	ソースデータ（ジャパンサーチ（仮称）が連携フォーマットで受け取ったデータ）とその提供者に関する情報。
提供者	ソースデータの提供者（つなぎ役）
データ	プラットフォームが保持・提供するソースデータ
リンク	つなぎ役におけるソースデータの掲載ページのURL
更新日	収集元データの更新日又はつなぎ役による収集日

ジャパンサーチにおける、データ変遷 (イメージ)

つなぎ役／ アーカイブ機関

(オリジナル) ソースデータ

名称: 刀
サイズ: 68.1弱
和暦: 長禄2
分野: 工芸
作者名: 長光
所蔵: X博物館



登録・共通項目ラベル付与

共通項目ラベルの付与

共通項目ラベル名

名称/タイトル

名称: 刀

サイズ: 68.1弱

和暦: 長禄2

分野: 工芸

作者名: 長光

所蔵: X博物館

DB単位
の情報

データベース: Y DB

タイプ: 文化財

魅力的表示



検索機能



- 共通項目ラベルは自動で付与後、データ提供館で確認・修正が可能
- 最小限度の項目のみ実施

ジャパンサーチ

詳細モデル化

ラベル: 刀剣 (<https://id.ndl.go.jp/auth/ndlsh/00573180>)

コンテンツ記述: 68.1弱

時間: 1459

主題: 工芸

作者: 長光

提供者: X博物館

アグリゲーター: Y DB

タイプ: 文化財

共通情報

名称: 刀

サイズ: 68.1弱

和暦: 長禄2

分野: 工芸

作者名: 長光

所蔵: X博物館

ソースデータ

マッピング

検索機能の
精緻・高度化「利活用フォーマット」
で出力

- 一部項目を抽出・コピー/変換 (正規化)
- 値はできるだけURI化
- ソースデータも保持