

# ジャパンサーチ 及び 次世代デジタルライブラリー について

国立国会図書館  
電子情報部

## 1. ジャパンサーチについて

## 1.1 ジャパンサーチの目的と概要

- **さまざまな分野のデジタルアーカイブと連携し、我が国が保有する多様なコンテンツのメタデータ\***をまとめて検索できる**国の分野横断統合ポータル**

\*メタデータとは：  
コンテンツの内容や所在等について記述したデータ。図書館の書誌データ、博物館・美術館の収蔵品の目録データなど。

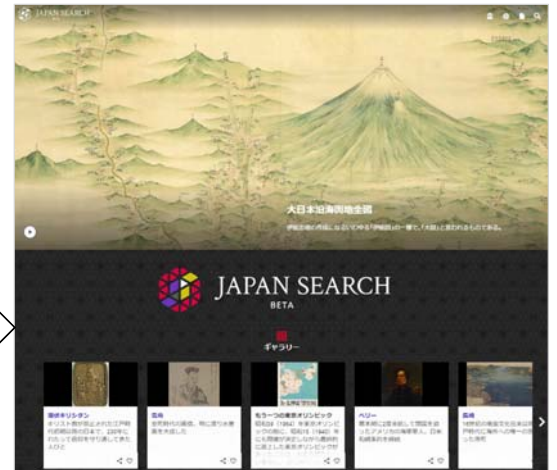
目指す効果

1. コンテンツの所在等の明確化
2. 連携機関へのアクセス促進
3. データの利活用の促進
4. 連携機関への支援
5. 新規ビジネス・サービスの創出等

- **政府の「知的財産推進計画」等に掲げられている国の取組**

運用主体：デジタルアーカイブジャパン推進委員会・実務者検討委員会  
(事務局：内閣府知的財産戦略推進事務局)  
システムの運用担当：国立国会図書館

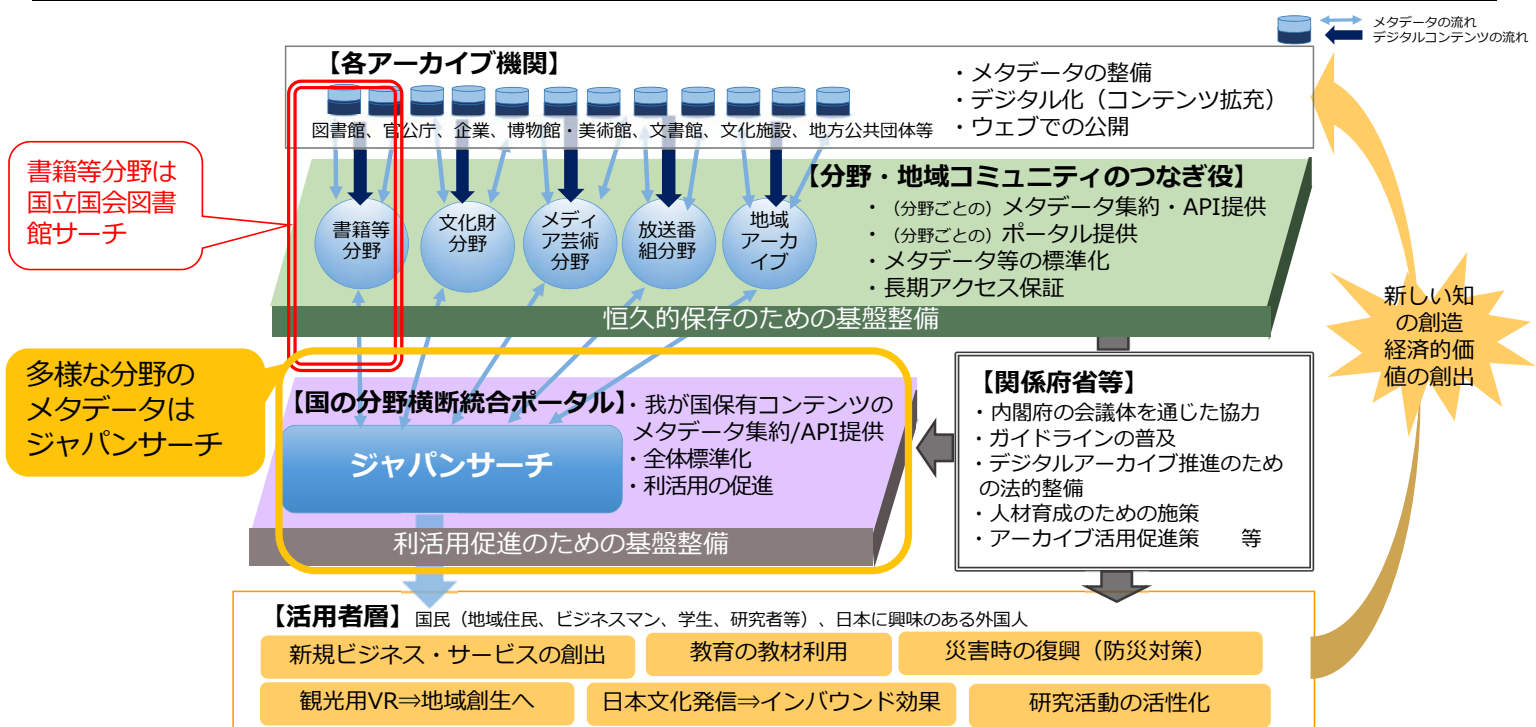
- **2019年2月に試験版を公開** <https://jpsearch.go.jp/>  
2020年の正式版公開に向けて、フィードバックを受付中



第11回科学技術情報整備審議会

3

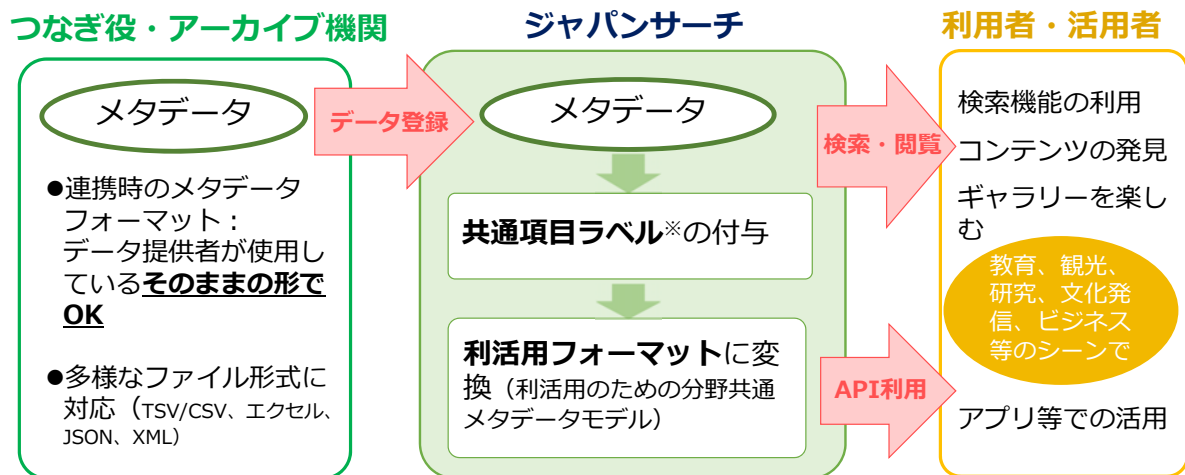
## (参考) ジャパンサーチ構築に向けて：役割の整理



第11回科学技術情報整備審議会

4

## (参考) メタデータ連携の流れ



※名称/タイトル、時間、場所、所蔵機関、URL等の分野横断で共通となりうる項目のみ

### メタデータの二次利用条件

- メタデータは原則CC0 ※国際的流通を意識 (Europeanaと同レベル)
  - ※著作権性のあるもののみ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、サイトポリシーをお願い

## 1.2 連携状況 (2019年8月9日現在)

**12機関 45データベース メタデータ約1,800万件**

分野	データ提供機関	データベース名
書籍等	国立国会図書館	「国立国会図書館サーチ」から、3件のデータベース
公文書	国立公文書館	「国立公文書館デジタルアーカイブ」
文化財	文化庁	「文化遺産オンライン」から、国指定文化財等データベース
	国立文化財機構	「ColBase 国立博物館所蔵品統合検索システム」
美術	国立美術館	「国立美術館所蔵作品総合目録検索システム」
		「アートコモンズ」
メディア芸術	映像産業振興機構	「Japan Content Catalog」から、2件のデータベース
舞台芸術	早稲田大学坪内博士記念演劇博物館	「演劇情報総合データベース」から、2件のデータベース
自然史・理工学	国立科学博物館	「サイエンスミュージアムネット S-Net」
		「魚類写真資料データベース」
人文学	人間文化研究機構	「人間文化研究機構統合検索システム nihuINT」から、27件のデータベース
	立命館大学アート・リサーチセンター	「ARC浮世絵ポータルデータベース」「ARC古典籍ポータルデータベース」
放送番組	放送番組センター	「放送ライブラリー公開番組データベース」※ドラマのデータ
	日本放送協会	「動画で見るニッポンみちしる」

調整中の連携先 (例)

データセット	内閣官房IT総合戦略室・総務省	データカタログサイト
メディア芸術	文化庁	メディア芸術データベース

## 1.3 ジャパンサーチ（試験版）の機能

### クリックするだけで楽しめる 多様なコンテンツの紹介

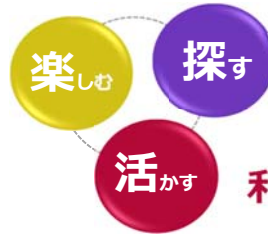


様々なテーマ\*に関する「ギャラリー」を提供

(\*テーマ例：長崎、富士山、刀剣、和食、北斎、伊能忠敬 etc.)

### 分野の特性を活かした 検索機能

### テーマ別検索\*



### 利活用機能

利活用促進の基盤となるAPI・ウェブパーツ機能を提供

さらに、  
ここにも注目！



デジタルコンテンツの二次利用条件表示を整備して、**教育利用 / 非商用利用 / 商用利用の可否がひと目で分かる**ようにしています。



「加藤正清」 広貞/立命館大学ARC所蔵

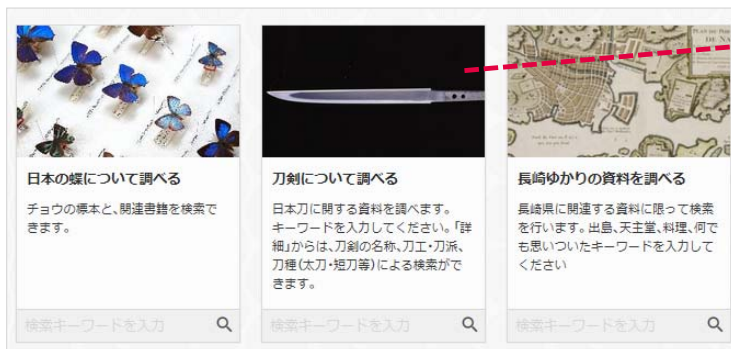
第11回科学技術情報整備審議会

7

## 探す

### (参考) テーマ別検索

- 検索対象とするデータベース・メタデータ項目を独自にマッピングして定義した検索 ⇒ 連携機関のオリジナルデータを活かしながら、特定のテーマに即した検索が可能
- ジャパンサーチのエディタで作成可能（現在は連携機関のみ）



テーマ別検索一覧 <https://jpsearch.go.jp/csearch?from=0>



ユーザからは見えない形で検索対象を絞り込むこと（検索制約）が可能  
【NDL書誌】=NDC756.6のみ 等

「刀剣の名称」の検索フィールド：  
【ColBase】名称、解説、銘文等  
【文化遺産オンライン】名称  
【NDL書誌】タイトル、要約・抄録  
等のDB項目をマッピング

検索結果で表示する項目やその  
順番、ファセットも設定可能

テーマ別検索「刀剣について調べる」  
<https://jpsearch.go.jp/csearch/ndl-46KaOzRRnZz>

第11回科学技術情報整備審議会

8

# (参考) 利活用のユースケース「ギャラリー」

- ・ ジャパンサーチの入口として、「検索機能」を使用しない場合でも見て楽しめるページ  
⇒特定の主題ごとに、画像、解説文、予め定義した検索結果等を表示し、コンテンツへの理解を促進する
  - ・ ジャパンサーチのエディタで作成可能（現在は連携機能のみ）
  - ・ 画像はIIIF（画像の相互運用のための規格）に変換して掲載
  - ・ SNSで簡便に拡散可能な仕組みも用意



ギャラリー「長崎」  
<https://jpsearch.go.jp/curation/ndl-i6b49EaX1nFwB9>



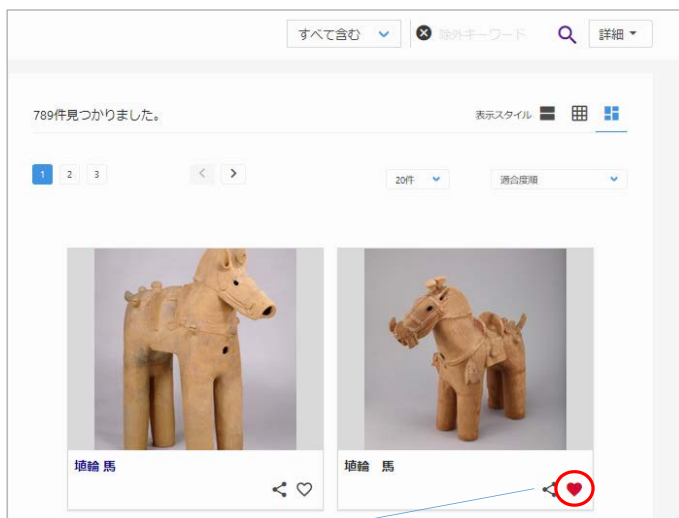
ギャラリー「このことごと〜筆と筆曲」  
<https://jpsearch.go.jp/curation/ndl-DZvPpbMP0nczD0>

テキスト、画像、  
 検索結果、検索画面  
 を組合せて表示

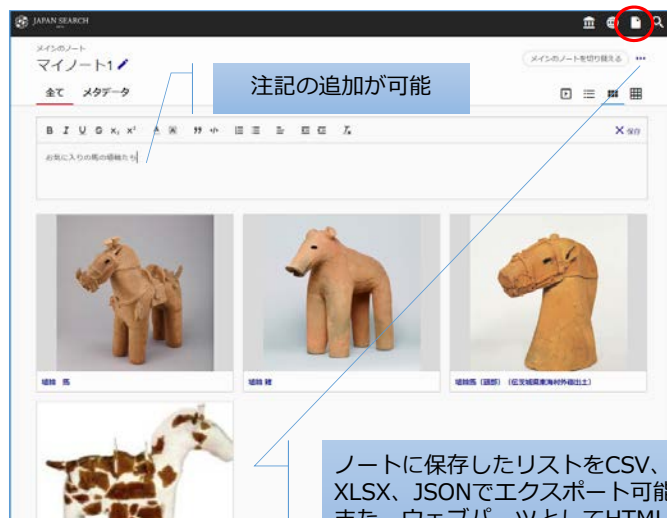
資料と解説文を  
 組み合わせた  
 電子展覧会

# (参考) 利活用のユースケース／利活用機能「マイノート」

- ・ お気に入りのギャラリー、メタデータをマイノートに登録できる
- ・ マイノートに登録したデータは、注記を加えたり、CSVなどの形式での出力も可能



検索結果の♡アイコンを  
 押すことで、お気に入りの  
 ギャラリー、メタデー  
 タをマイノートに登録

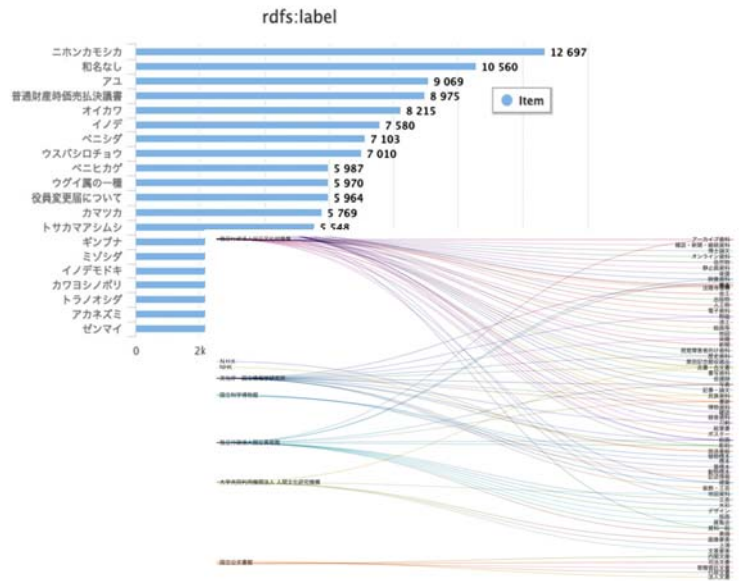


ノートに保存したリストをCSV、  
 XLSX、JSONでエクスポート可能  
 また、ウェブパーツとしてHTMLへの  
 貼り付けも可能

## データの分析・可視化

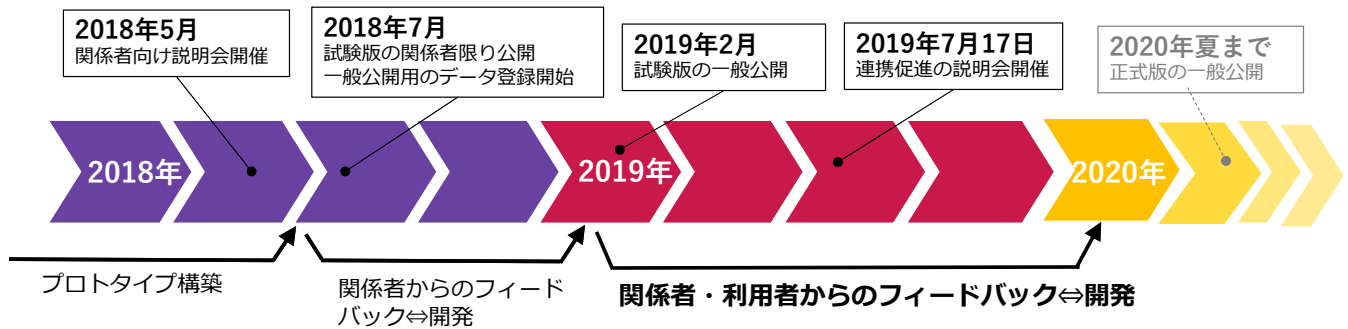
### ・ ジャパンサーチにどのようなコンテンツが含まれているか

- ・ [データベース別のアイテム数](#)
- ・ [連携機関別のアイテム数](#)
- ・ [各プロパティの値トップ20](#)
- ・ [年代別のアイテム数](#)
- ・ [資料種別の階層関係](#)
- ・ [連携機関と資料種別の関係性](#)



中村覚「ジャパンサーチのメタデータ活用事例『ジャパンサーチを使う』」  
 (2019年7月17日「ジャパンサーチ発進!～連携拡大に向けて」発表資料)  
<https://www.ndl.go.jp/ip/event/events/201907ips.html>

## 1.4 ジャパンサーチの今後の予定



### ● 2020年正式版公開に向けて

#### □ 機能拡張

- ・ UXデザインの改善
- ・ 利活用のための情報共有のページの作成
- ・ 連携機関向けページの作成
- ・ テーマ別検索及びギャラリーのエディタ機能の改善 など

#### □ 連携の拡大

- ・ 国の機関、大学等研究機関に加えて、地方公共団体等の地域アーカイブとの連携へ

#### □ 正式版の運用体制の構築

## 2. 次世代デジタルライブラリーについて

### 2.1 次世代デジタルライブラリーの目的と概要

- 電子情報企画課次世代システム開発研究室がこれまでの研究成果をもとに開発
- 機械学習(AI)の技術を活用したデジタルライブラリーの新しい機能を検討するための実験システム
- 2019年3月末に公開



<https://lab.ndl.go.jp/dl/>

#### 対象資料

まずはNDLデジタルコレクションのうち、1800~1945年に出版された  
**NDC6類(産業)のパブリックドメイン資料 約2万5000点を搭載** (順次拡大中)

#### 次世代デジタルライブラリーのねらい

- OCRを活用した全文検索機能 & IIIF APIの実証実験
- 機械学習技術を活用した新しい切り口でのサービス提供 & デジタル化資料の利活用促進

## 2.2 次世代デジタルライブラリーの機能①



### 機能①

#### 光学文字認識(OCR)による全文検索機能

読み取りソフトウェアによる認識結果をそのまま全文検索に使うことで低コストに提供

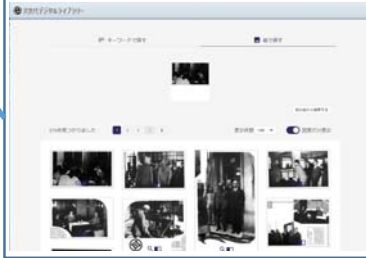
検索にヒットした箇所の**ハイライト**  
&  
ヒットした箇所の前後のスニペット表示

### 機能②

#### 図版の自動抽出及び画像検索機能

資料に含まれる挿絵や写真の位置を、機械学習により自動検出・抽出し、一覧表示や類似の画像を検索する機能を提供

#### 画像検索のUI



自動抽出された図版の一覧

画像の表示には  
**IIIF(International Image Interoperability Framework) API**を利用

第11回科学技術情報整備審議会

15

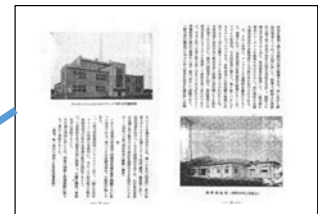
## 2.3 次世代デジタルライブラリーの機能②



### 機能③

#### 変色を自動的に除去して提供する白色化機能

加工された資料画像と元の資料画像の対応関係を学習させることで、紙の変色を除去し、可読性を高める加工を自動で行う機能を提供。



### 機能④

#### 縦長ディスプレイを想定した表示の自動加工機能



第11回科学技術情報整備審議会

16



## 2.4 次世代デジタルライブラリーに関連する今後の予定

---

### ● 本文検索サービスの拡充に向けて

令和元年度：本文検索サービスの対象資料の選定基準の策定

デジタル化資料のテキスト化の精度向上に向けた実験（継続）

令和2年度：「国立国会図書館デジタルコレクション」におけるデジタル化資料の本文検索サービスの開始

### ● データセットの活用に向けて

令和元年度：データセットを用いたイベント等の実施（ハッカソン（9月実施）、技術コンペ）

解析目的でのデータセット提供の枠組の整備

GitHub上でのデータセット及びプログラムの提供開始

データセットを活用して、書誌データの自動生成、レファレンスの自動応答サービスの研究（継続）

令和2年度：データセットの利活用事例共有のための発信強化の取組

引き続き「次世代デジタルライブラリー」の機能をさらに発展させ、より利用しやすい電子図書館サービスの実現へ