



## ジャパンサーチのLOD

デジタルアーカイブが「つながる」ために

令和4年6月26日 第2回DAフォーラム

国立国会図書館 電子情報部電子情報流通課  
町屋大地, 伊藤実, 高橋美知子

# ジャパンサーチ

# ジャパンサーチとは

## ● デジタルアーカイブの検索・閲覧・活用基盤

さまざまな分野のデジタルアーカイブと連携し、我が国の多様なコンテンツの**メタデータ\***をまとめて検索・閲覧・活用

## ● 国全体の取組

◆ 政府の「知的財産推進計画」に掲げられている国全体の取組

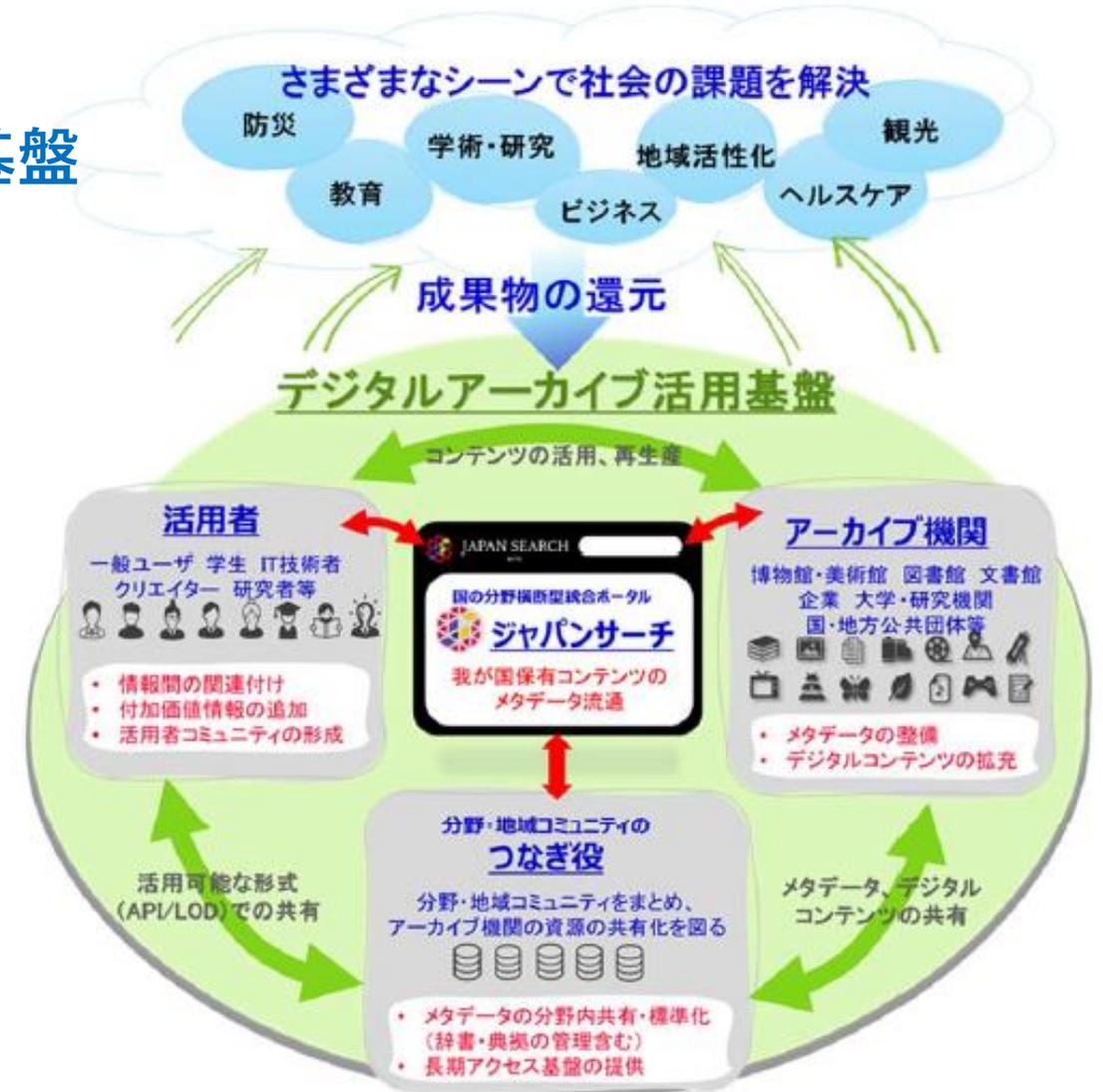
◆ 運営主体：デジタルアーカイブジャパン推進委員会・実務者検討委員会

(事務局：内閣府知的財産戦略推進事務局)

システム運用・連携協力担当：国立国会図書館

### \*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。



「3か年総括報告書 我が国が目指すデジタルアーカイブ社会の実現に向けて」  
(令和2年8月19日 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)  
[https://www.kantei.go.jp/jp/singi/titeki2/digitalarchive\\_suisiniinkai/pdf/r0208\\_3kanen\\_houkoku\\_honbun.pdf](https://www.kantei.go.jp/jp/singi/titeki2/digitalarchive_suisiniinkai/pdf/r0208_3kanen_houkoku_honbun.pdf)



# ジャパンサーチの分野別連携状況（2022年6月10日現在）

**33**連携機関(つなぎ役) **171**データベース 約**2,560万件**のメタデータ

分野	主なデータベース名	メタデータ件数
書籍等	「国立国会図書館サーチ（NDLサーチ）」から「国立国会図書館デジタルコレクション」、「青森県立図書館デジタルアーカイブ」、「デジタルアーカイブ福井」、「佐賀県立図書館データベース」等	11,946,943
自然・理工学	「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」、「南方熊楠邸資料：蔵書、標本、文書」、「蚕織錦絵コレクション」等	5,431,704
人文学	「人間文化研究機構統合検索システム nihulNT」から54件のデータベース、「ARC古典籍ポータルデータベース」、「渋沢栄一関連会社名・団体名変遷図」、「東京大学学術資産等アーカイブズポータル」、「京都大学研究資源アーカイブ」から3件のデータベース等	5,045,478
公文書	「国立公文書館デジタルアーカイブ」、「和歌山県歴史資料アーカイブ」等	4,099,609
地域	「三重の歴史・文化デジタルアーカイブ」から19件のデータベース、「にいがた 地域映像アーカイブ データベース」、「信州デジタルcommons」から3件のデータベース、とっとりデジタルコレクション、なんじょうデジタルアーカイブ、Tokyo Museum Collection、広島平和記念資料館 平和データベース、北海道デジタルミュージアム等	660,694
美術	「国立美術館所蔵作品総合目録検索システム」、「写真原板データベース」、「愛知県美術館コレクション」、「大阪市立東洋陶磁美術館収蔵品画像オープンデータ」等	596,313
文化財	「文化遺産オンライン」から国指定文化財等データベース、「文化財デジタルコンテンツダウンロード機能」「ColBase 国立博物館所蔵品統合検索システム」、「Keio Object Hub」等	520,291
アニメ・ゲーム・メディアアート	「日本アニメーション映画クラシックス」、「Japan Content Catalog」から「JACCクリエイター検索」、「メディア芸術データベース」等	190,506
舞台芸術	「演劇情報総合データベース」から4件のデータベース等	84,741
映画	「国立映画アーカイブ所蔵映画フィルム検索システム」、「映像でみる明治の日本」等	82,707
放送番組	「放送ライブラリー公開番組データベース（ドラマ）」、「動画で見るニッポンみちしる」等	82,332
データセット	「データカタログサイト」等	24,439
地図	「古地図コレクション」	3,349

## 「デジタルアーカイブを日常にする」

### ミッション

新しい情報技術とアーカイブ連携を通じて、日本の文化的・学術的コンテンツの発見可能性を高め、それらを活用しやすい基盤を提供することで、デジタルアーカイブが日常に溶け込んだ豊かな創造的社會を実現します

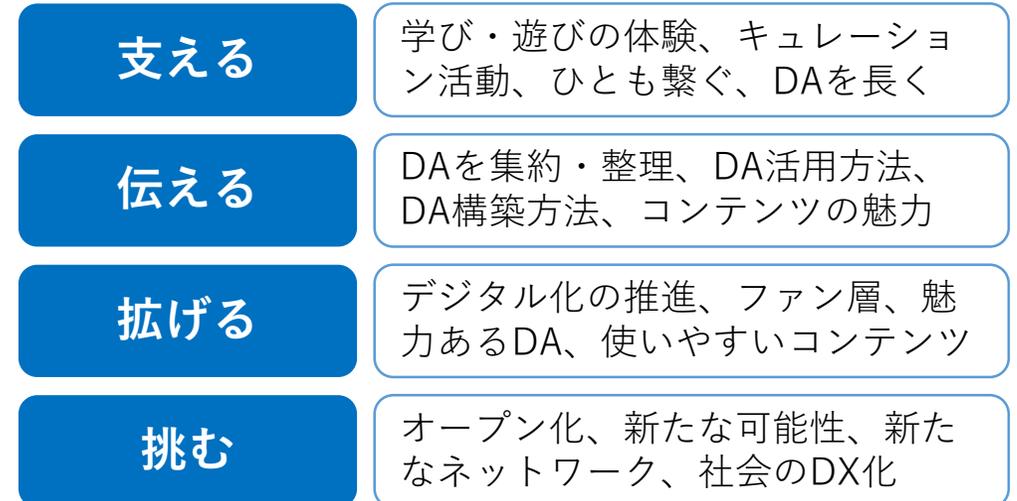
### 3つの価値：

デジタルアーカイブの大切な役割



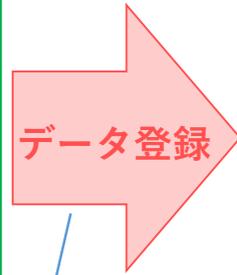
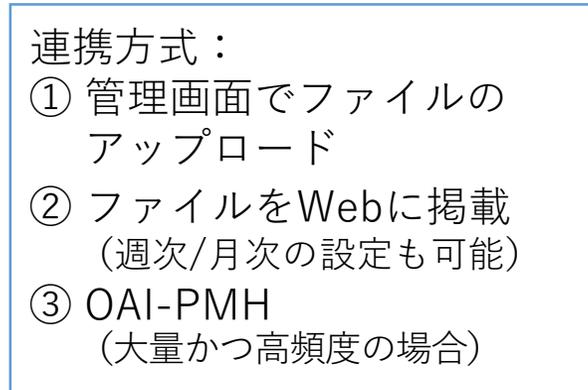
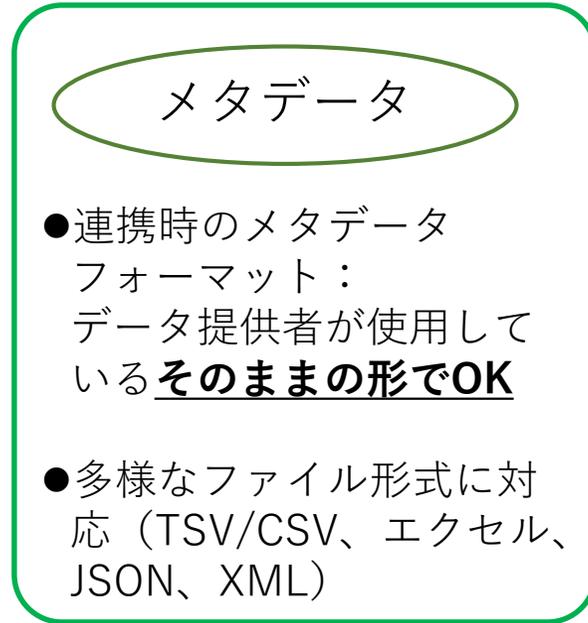
### 4つのアクション：

ジャパンサーチを使った活動の柱

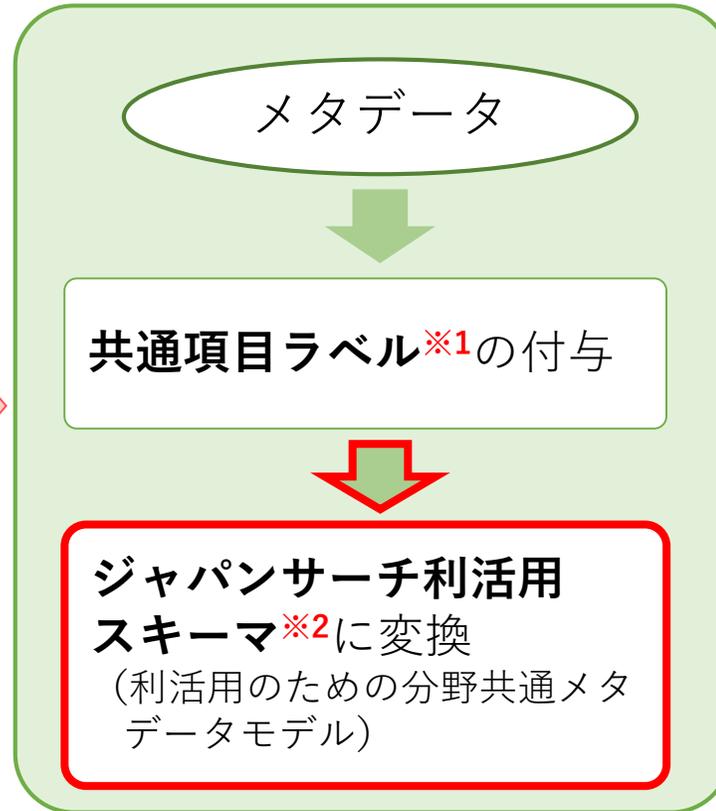


# メタデータ連携と活用までの流れ

## つなぎ役・アーカイブ機関



## ジャパンサーチ



## 利用者・活王者



※1：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

※2：分野横断して利活用可能とするジャパンサーチにおけるメタデータの標準形式。  
Europeana等とも連携可能

# ジャパンサーチでのLODの提供



# Linked Open Data とは

●**Linked Open Data(LOD)** = リンクされたオープンなデータ

●**リンクされたデータ (Linked Data)**

- ウェブ上に存在する他のデータと「リンク」されているデータ
- Linked Data の4原則 (Tim Berners-Lee) <https://www.w3.org/DesignIssues/LinkedData.html>
  - 事物の名前としてURIを用いる
  - これらの名前を参照できるようにHTTP URIを用いる
  - URIが参照されたときRDF\*やSPARQL\*のような標準技術で有用な情報を提供する
  - さらに多くの事物を発見できるように、他のURIへのリンクを含む

●**オープンなデータ (Open Data)**

- オープンなライセンス (CC0、CC BY等) で公開されているデータ
- データの二次利用が可能になる



# RDFとSPARQL

## ●RDF = Resource Description Framework

- セマンティックウェブ（情報資源に意味の明確なデータを付与し、機械的な意味処理を目指す構想）におけるメタデータの標準的な表現方法
- 主語（Subject）・述語（Predicate）・目的語（Object）の3つ組（トリプル）からなる文（ステートメント）の組み合わせによってメタデータを表現する
- 主語・述語・目的語にURIを用いることで表現の曖昧さを回避



## ●SPARQL = RDFを検索するための言語



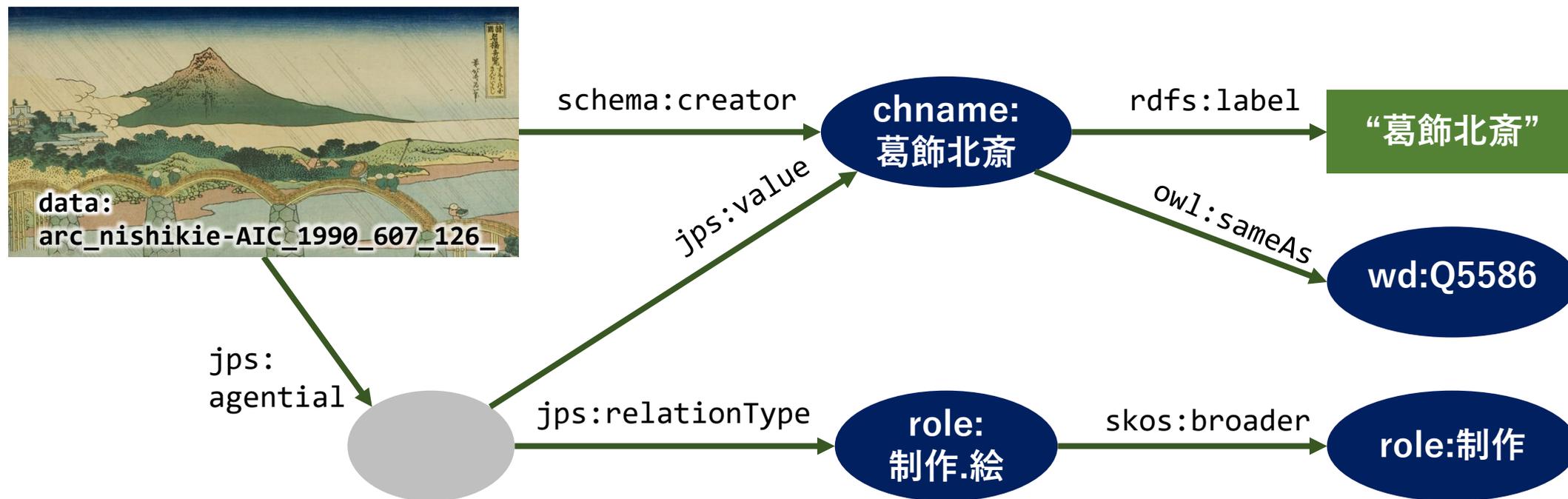
## ジャパンサーチで提供するLOD

- ジャパンサーチで検索可能なデータベースのうち、提供元がAPIでの提供を認めているデータベースを順次LODに変換
- LOD変換・提供：142データベース（約2,200万件）（2022年6月10日時点）  
※ ジャパンサーチ連携：171データベース（約2,600万件）
- メタデータは「ジャパンサーチ利活用スキーマ」に基づくRDFへ変換  
<https://jpsearch.go.jp/api/introduction/>
- ジャパンサーチのSPARQLエンドポイントから利用可能  
<https://jpsearch.go.jp/rdf/sparql/>
- 利用を支援するブラウザ向けインターフェースEasySPARQLを提供  
<https://jpsearch.go.jp/rdf/sparql/easy/>



# ジャパンサーチ利活用スキーマ

- メタデータの利活用のため扱いやすさを重視
  - 多様なデータベース由来のメタデータを共通の構造に変換
  - シンプルな基本記述プロパティと構造化記述プロパティを併用
  - プロパティの語彙をschema.orgと独自語彙に絞り込み
  - 検索に重要な値を正規化しURIを付与





# ジャパンサーチにおけるデータ変遷（イメージ）

## 連携機関

## ジャパンサーチ

**(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
時間/時代	サイズ：68.1弱
人物/団体	和暦：長禄2
提供者	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

マッピング

## 詳細モデル化

種別：刀剣	名称：刀
コンテンツ記述：68.1弱	サイズ：68.1弱
時間：1459	和暦：長禄2
主題：工芸	分野：工芸
作者：長光	作者名：長光
提供者：X博物館	所蔵：X博物館
アグリゲーター：Y DB	<b>ソースデータ</b>
タイプ：文化財	

**共通情報**

- 共通項目ラベルは自動で付与後、連携機関で確認・修正が可能
- 最小限度の項目のみ実施

**魅力的表示**

**検索機能**

- 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

**検索機能の精緻・高度化**

**「利活用スキーマ」で出力**

ジャパンサーチでの正規化

# ジャパンサーチでの正規化



## メタデータ中の実体の正規化

- メタデータ中で表現される実体（時間・場所・人物・組織等）は、多くのデータベースでは文字列として記録されており、表記ゆれ等が生じている
- 単純なキーワード検索では網羅的な検索ができない場合がある
- この課題を解消するため実体をURIの形で正規化した

データベース A	北斎
データベース B	葛飾北斎筆
データベース C	前北斎為一筆
データベース D	葛飾前北斎為一老人画



<https://jpsearch.go.jp/entity/chname/葛飾北斎>



## 正規化URIを使用した検索（SPARQL）

- 正規化URIを用いたクエリをジャパンサーチのSPARQLエンドポイントに送信することで原DBの記述に依らず網羅的な検索を行うことが可能
- 葛飾北斎が制作した作品を検索するクエリの例

```
SELECT ?s ?label WHERE {  
  ?s rdfs:label ?label .  
  ?s jps:agential [  
    jps:relationType/skos:broader* role:制作 ;  
    jps:value cname:葛飾北斎 ] .  
} LIMIT 200
```

※実際のリクエスト時は名前空間を宣言するPREFIX句が必要だが本スライドでは原則として省略  
ジャパンサーチのEasySPARQLでは利活用スキーマで使用する名前空間のPREFIXが補完される



## 正規化URIと外部のURIとのリンク

- 正規化URIはWeb NDL Authorities、DBpedia、Wikidata、VIAF等の外部のURIと同定しリンクしている
- これにより、外部のURIを検索条件として検索することも可能である

<https://jpsearch.go.jp/entity/chname/葛飾北斎>



owl:sameAs

<http://collection.britishmuseum.org/id/person-institution/1820>  
<http://data.bnf.fr/ark:/12148/cb124954814#about>  
<http://dbpedia.org/resource/Hokusai>  
<http://edan.si.edu/saam/id/person-institution/2267>  
<http://id.ndl.go.jp/auth/entity/00270331>  
<http://ja.dbpedia.org/resource/葛飾北斎>  
<http://lod.ac/id/1626>  
<http://viaf.org/viaf/69033717>  
<http://www.wikidata.org/entity/Q5586>



## 外部のURIを使用した検索

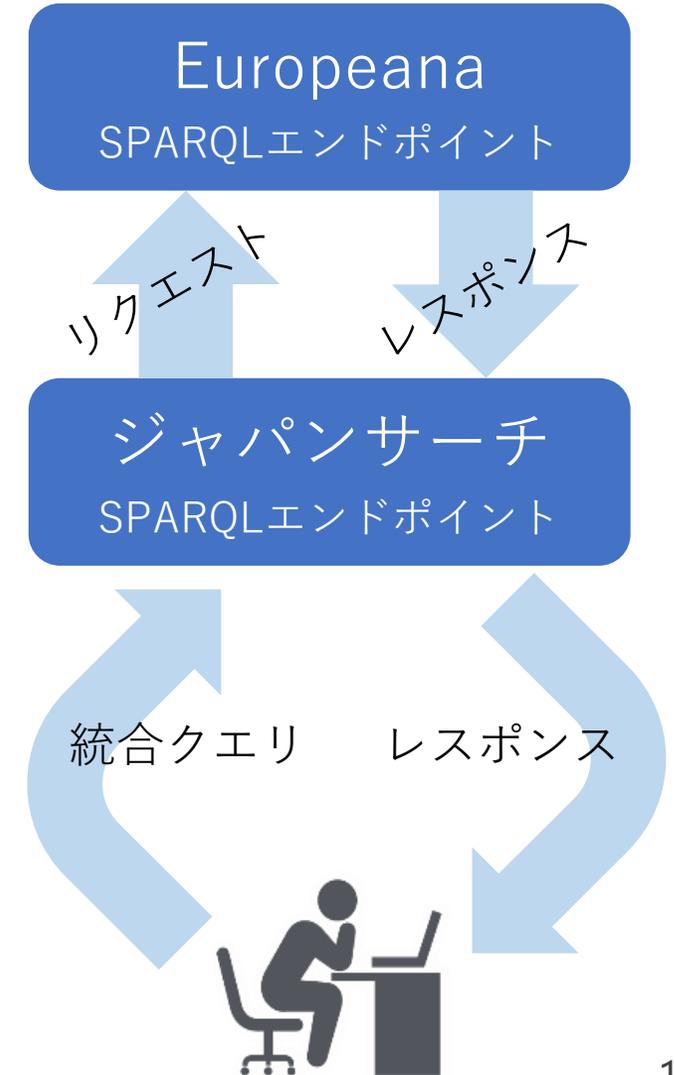
- 正規化URIを用いたクエリをジャパンサーチのSPARQLエンドポイントに送信することで原DBの記述に依らず網羅的な検索を行うことが可能
- DBpediaのURIを用いて葛飾北斎が制作した作品を検索するクエリの例

```
PREFIX dbr: <http://dbpedia.org/resource/>
SELECT ?s ?label WHERE {
    ?s rdfs:label ?label .
    ?s jps:agential [
        jps:relationType/skos:broader* role:制作 ;
        jps:value/owl:sameAs+ dbr:Hokusai ] .
} LIMIT 200
```

※**dbr:Hokusai** <http://dbpedia.org/resource/Hokusai>

# ジャパンサーチとEuropeanaの横断検索

- ジャパンサーチでは正規化した人物のURIをDBpediaのURIと紐づけている
- Europeanaは一部の人物をDBpediaのURIで正規化しているため、統合クエリによりジャパンサーチとEuropeanaの横断検索が可能になる
- 例えば、葛飾北斎を示すDBpediaのURI `<http://dbpedia.org/resource/Hokusai>` をもとに、ジャパンサーチ・Europeana双方で葛飾北斎の作品を検索することが可能となる





## ジャパンサーチとEuropeanaの統合検索のクエリ例

- 葛飾北斎が制作した作品を横断検索する統合クエリ

```
PREFIX dbr: <http://dbpedia.org/resource/>
```

```
PREFIX dc: <http://purl.org/dc/elements/1.1/>
```

```
PREFIX ore: <http://www.openarchives.org/ore/terms/>
```

```
SELECT ?s ?label WHERE{
```

```
  {SERVICE <http://sparql.europeana.eu/> {
```

```
    ?proxy dc:creator dbr:Hokusai ; ore:proxyFor ?cho .
```

```
    ?s dc:title ?label ; ore:proxyFor ?cho; edm:europeanaProxy "false".
```

```
  }} UNION
```

```
  {SERVICE <https://jpsearch.go.jp/rdf/sparql> {
```

```
    ?s rdfs:label ?label;
```

```
    schema:creator/owl:sameAs+ dbr:Hokusai .
```

```
  }}
```

```
} LIMIT 200
```



# Wikidataを利用した検索の可能性

- Wikidata = ユーザーの共同編集による構造化データベース
    - Wikipediaと同じくウィキメディア財団がサービスを提供
    - 分野を限定せず多様なデータをCC0で提供・編集可能
  - ジャパンサーチで保持していない多様な識別子での検索が可能
    - 人物用の識別子だけで約2,500のプロパティが存在
      - ※Wikidata Query Serviceで以下のクエリにより確認
- ```
SELECT DISTINCT ?item WHERE { ?item wdt:P31/wdt:P279* wd:Q19595382. }
```
- 多言語に対応しているため日本語話者でないユーザーにとっても検索しやすくなることが期待される
  - 葛飾北斎であれば100種以上の識別子と150言語以上での名称が記録されている



## (参考) Wikidataへのクエリ例

- 葛飾北斎(Q5586)に登録されている識別子の一覧 <https://w.wiki/5JHG>

```
SELECT ?prop ?propLabel (SAMPLE(?identifiers) AS ?identifier) WHERE {  
  wd:Q5586 ?directProp ?identifiers .  
  ?prop wikibase:directClaim ?directProp .  
  ?prop wdt:P31/wdt:P279* wd:Q18614948 .  
  SERVICE wikibase:label { bd:serviceParam wikibase:language "ja, en". }  
} GROUP BY ?prop ?propLabel
```

- 葛飾北斎(Q5586)に登録されている各言語の名称の一覧 <https://w.wiki/5J37>

```
SELECT ?language ?label WHERE {  
  wd:Q5586 rdfs:label ?label .  
  BIND(LANG(?label) AS ?language)  
}
```

※ スライド中の <https://w.wiki/> から始まるURLはWikidata Query Serviceへのリンクです

ジャパンサーチを「つなげる」



## ジャパンサーチへの外部データセットの取り込み

- ジャパンサーチでは、連携対象となるデジタルアーカイブのメタデータに加えて、データセットを取り込むことで効果的な検索を可能にしている
- NDC-LD = 日本十進分類法（NDC）のRDF形式のオープンデータ
  - 日本図書館協会と国立国会図書館の共同研究の成果として公開
  - SPARQLエンドポイントが提供されていないため利活用のハードルが高かった
  - ジャパンサーチに取り込みNDCが含まれるデータと紐づけた
- 学習指導要領LOD = 学習指導要領コードや関連情報をLOD化したもの
  - 学習指導要領小学6年社会科の「我が国の歴史上の主な事象に関する内容」の各項目のURIとジャパンサーチの正規化URIとを紐づけた
  - 学習指導要領をもとにデジタルアーカイブが探索可能になる可能性



## ジャパンサーチでのNDC-LDの活用例

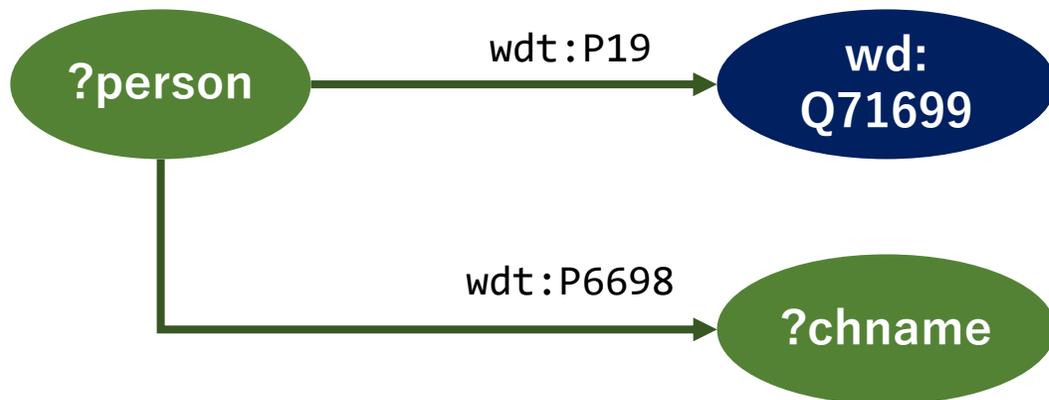
- NDCが49（医学・薬学）から始まるメタデータをNDC別に集計する

```
SELECT ?ndc (count(?cho) as ?count) ?ndcLabel WHERE {  
  ?cho schema:about/skos:closeMatch ?ndc .  
  ?ndc skos:broader+ <http://jla.or.jp/data/ndc9#49> ;  
      rdfs:label ?ndcLabel .  
  FILTER NOT EXISTS { ?cho jps:sourceInfo/schema:provider cname:全国書誌 . }  
} ORDER BY ?ndc
```

| ndc                                                                                 | count | ndcLabel                                  |
|-------------------------------------------------------------------------------------|-------|-------------------------------------------|
| <a href="http://jla.or.jp/data/ndc9#493.157">http://jla.or.jp/data/ndc9#493.157</a> | 1     | "内科学--中毒症--食中毒"                           |
| <a href="http://jla.or.jp/data/ndc9#493.16">http://jla.or.jp/data/ndc9#493.16</a>   | 34    | "内科学--全身病. 一般的疾患--寄生虫病. 地方病 [風土病]"        |
| <a href="http://jla.or.jp/data/ndc9#493.17">http://jla.or.jp/data/ndc9#493.17</a>   | 14    | "内科学--全身病. 一般的疾患--血液疾患"                   |
| <a href="http://jla.or.jp/data/ndc9#493.18">http://jla.or.jp/data/ndc9#493.18</a>   | 32    | "内科学--全身病. 一般的疾患--成人病. 成人医学"              |
| <a href="http://jla.or.jp/data/ndc9#493.2">http://jla.or.jp/data/ndc9#493.2</a>     | 110   | "内科学--循環器疾患"                              |
| <a href="http://jla.or.jp/data/ndc9#493.23">http://jla.or.jp/data/ndc9#493.23</a>   | 1     | "内科学--循環器疾患--心臓：心筋梗塞症, 心臓神経症, 心臓弁膜症, 心膜炎" |
| <a href="http://jla.or.jp/data/ndc9#493.3">http://jla.or.jp/data/ndc9#493.3</a>     | 53    | "内科学--呼吸器疾患"                              |
| <a href="http://jla.or.jp/data/ndc9#493.38">http://jla.or.jp/data/ndc9#493.38</a>   | 1     | "内科学--呼吸器疾患--肺：肺炎, 肺化膿症, 肺気腫, 肺塞栓症"       |

# WikidataへのジャパンサーチURIの登録

- Wikidataにはジャパンサーチで正規化したURIのためのプロパティ (P6698) が設定されており、未登録のURIをWikidataへ登録した
- Wikidataは識別子のハブであると同時に、分野を問わない大規模な知識データベースであるため、より多様な活用が期待できる
- 例えば、Wikidataは人に関するプロパティとして出生地 (P19)、死亡地 (P20)、性別 (P21) などを持つため、地元にはゆかりのある人物のジャパンサーチの正規化URIの一覧を抽出し検索に使用することができる





# Wikidata上の多様な情報を活用した検索(1)

- 青森県出身女性のジャパンサーチ正規化URIを検索するWikidataクエリの例 <https://w.wiki/5KF5>

```
SELECT ?person ?personLabel ?chname WHERE {
    ?person wdt:P31 wd:Q5 ;
    wdt:P19/wdt:P131* wd:Q71699 ;
    wdt:P21 wd:Q6581072 ;
    wdt:P6698 ?chnameID .
    SERVICE wikibase:label { bd:serviceParam wikibase:language "ja, en". }
    BIND (IRI (CONCAT ("https://jpsearch.go.jp/entity/chname/", ?chnameID) )
AS ?chname)
}
```



## Wikidata上の多様な情報を活用した検索(2)

- 青森県出身の女性が制作した作品を検索するクエリの例 (統合クエリ)

```
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
```

```
PREFIX wd: <http://www.wikidata.org/entity/>
```

```
SELECT ?cho ?chname ?label WHERE {
```

```
  ?cho rdfs:label ?label;
```

```
  schema:creator/owl:sameAs* ?chname.
```

```
{ SELECT ?chname WHERE {
```

```
  SERVICE <https://query.wikidata.org/sparql> {
```

```
    ?creator wdt:P31 wd:Q5 ;
```

```
    wdt:P19/wdt:P131* wd:Q71699 ;
```

```
    wdt:P21 wd:Q6581072 ;
```

```
    wdt:P6698 ?chnameID .
```

```
    BIND (IRI (CONCAT ("https://jpsearch.go.jp/entity/chname/", ?chnameID) )
```

```
  AS ?chname) }
```

```
  } ORDER BY ?chname LIMIT 10 }
```

```
}
```



# まとめ



# ジャパンサーチ戦略方針2021-2025（再掲）

## 「デジタルアーカイブを日常にする」

### ミッション

新しい情報技術とアーカイブ連携を通じて、日本の文化的・学術的コンテンツの発見可能性を高め、それらを活用しやすい基盤を提供することで、デジタルアーカイブが日常に溶け込んだ豊かな創造的社會を実現します

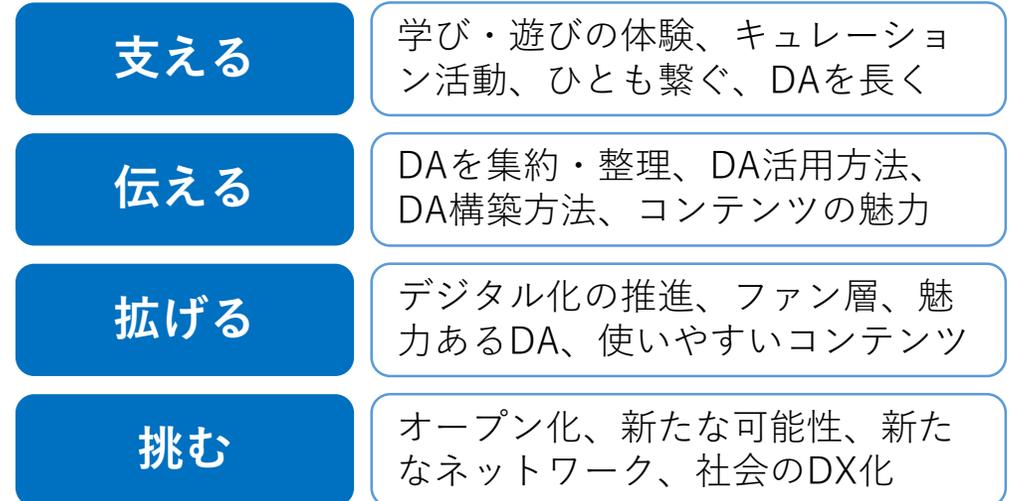
### 3つの価値：

デジタルアーカイブの大切な役割



### 4つのアクション：

ジャパンサーチを使った活動の柱





## まとめ

- ジャパンサーチでは連携データベースのメタデータをLOD化して提供
- データの正規化や外部のURIとの関連付けによって利活用をサポート
- 他のLODやデータセットとの連携によって多様な利活用の可能性



- ジャパンサーチでのLODの提供は戦略方針に示す「日本の文化的・学術的コンテンツの発見可能性を高め、それらを活用しやすい基盤を提供する」というミッションの重要な達成手段
- 提供LODの拡大と同時に、フィードバックを受けた改善を継続

**LODの利用・フィードバックをお待ちしています**



## 参考資料

- 神崎正英. ジャパンサーチ利活用スキーマの設計と応用. デジタルアーカイブ学会誌. 2020, 4(4), p.342-347.  
[https://doi.org/10.24506/jsda.4.4\\_342](https://doi.org/10.24506/jsda.4.4_342)
- 神崎正英. RDFとSPARQLによる多様なデータの活用. 情報の科学と技術. 2020, vol. 70, no. 8, pp.399-405.  
[https://doi.org/10.18919/jkg.70.8\\_399](https://doi.org/10.18919/jkg.70.8_399)
- 大向一輝. 識別子としてのWikidata. 情報の科学と技術. 2020, vol. 70, no. 11, pp.559-562.  
[https://doi.org/10.18919/jkg.70.11\\_559](https://doi.org/10.18919/jkg.70.11_559)
- 日本図書館協会. NDC データ (NDC8 版および 9 版) .  
<https://www.jla.or.jp/committees/bunrui/tabid/789/Default.aspx>
- 教育データプラス研究会. 学習指導要領LOD. <https://w3id.org/jp-cos/>
- ジャパンサーチ利活用スキーマ概説. <https://jpsearch.go.jp/api/introduction/>
- ジャパンサーチのSPARQLエンドポイント <https://jpsearch.go.jp/api/sparql-explain/>
- ジャパンサーチ戦略方針2021-2025. <https://jpsearch.go.jp/about/strategy2021-2025>

# ご清聴ありがとうございました

フィードバックはお気軽に以下からお寄せください

<https://jpsearch.go.jp/contact>



JAPAN SEARCH 公式  
@jpsearch\_go



JAPAN SEARCH 公式 YouTubeチャンネル  
<https://www.youtube.com/channel/UCSK72iSEQMzEcu8BWLEFzPQ>



ジャパンサーチ / Japan Search Facebook  
<https://www.facebook.com/jpsearchgo/>