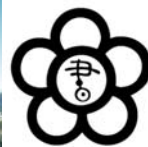




# 国の分野横断統合ポータル 「ジャパンサーチ（仮称）」の取組みについて

～分野横断のためのメタデータフォーマットを中心に～

国立国会図書館 電子情報部  
電子情報流通課標準化推進係  
奥田倫子



## 国立国会図書館について National Diet Library (NDL)

### ● 納本図書館

納本制度に基づく国内出版物の網羅的な収集・保存  
全国書誌の作成

### ● 幅広いサービス対象

国会（議会図書館）、行政・司法、国民

### ● 国の中央図書館

図書館へのサービス、図書館の図書館

# 本スライド資料について

- ① 「ジャパンサーチ」は仮称です、以下（仮称）は省略します。
- ② ジャパンサーチについては、検討段階の内容を多く含みます。
- ③ 本資料の記載内容のうち、意見にわたる部分は、報告者の私見です。

3

## 本日の内容

- 1. 「ジャパンサーチ」構築の背景と現況
- 2. 国立国会図書館における統合ポータル構築小史
- 3. 海外の分野横断統合ポータル（のメタデータ）
- 4. ジャパンサーチにおけるメタデータの集約と変換
- 5. ジャパンサーチの「利活用フォーマット」
- 6. まとめ

4

# 1. 「ジャパンサーチ」構築の背景と現況

## 国におけるデジタルアーカイブ推進の動き

### 知的財産推進計画2015

- ◆重要施策「6. アーカイブの利活用促進に向けた整備の加速化」において、具体的に、**国立国会図書館**が国の統合ポータルサイトの構築に関与することに言及
- ◆**デジタルアーカイブの連携に関する関係省庁等連絡会/実務者協議会**を設置し、具体策に向けた検討を開始

### 知的財産推進計画2016

- ◆「第3. コンテンツの新規展開の推進」「2. アーカイブ利活用の促進」において、**国立国会図書館サーチ**が国の分野横断統合ポータルとして、文化遺産オンラインと連携することに言及
- ◆引き続き、関係省庁等連絡会/実務者協議会において、メタデータのオープン化、サムネイル/プレビューの利用条件等の課題と対応策の検討を実施

## 国におけるデジタルアーカイブ推進の動き

### **知的財産推進計画2017（平成29年5月16日知的財産戦略本部決定）**

- ◆「国の分野横断型の統合ポータル「ジャパンサーチ（仮称）」を構築するためのデータフォーマットなどの課題の解決に取り組み、2020年までにその構築を目指す。（短期・中期 国立国会図書館）」

### **知的財産推進計画2018（平成30年6月12日知的財産戦略本部決定）**

- ◆「今後、2020年までの立ち上げを目指している国の分野横断型の統合ポータル「ジャパンサーチ（仮称）」の構築を進める」

### **経済財政運営と改革の基本方針 2018（平成30年6月15日閣議決定）**

- ◆「文化資源について、各分野のデジタルアーカイブ化を進めるとともに、内外の利用者が活用しやすい統合ポータルの構築を推進する。」

### **未来投資戦略 2018（平成30年6月15日閣議決定）**

- ◆「デジタルアーカイブジャパンの中心となる分野横断型統合ポータル「ジャパンサーチ（仮称）」の本格稼働に向けた取組を推進する。」

## 国におけるデジタルアーカイブ推進の動き

### 「我が国におけるデジタルアーカイブ推進の方向性」

（デジタルアーカイブの連携に関する関係省庁等連絡会・実務者協議会、平成29年4月）

[https://www.kantei.go.jp/jp/singi/titeki2/digitalarchive\\_kyougikai/houkokusho.pdf](https://www.kantei.go.jp/jp/singi/titeki2/digitalarchive_kyougikai/houkokusho.pdf)

### 「デジタルアーカイブの構築・共有・活用ガイドライン」

（デジタルアーカイブの連携に関する関係省庁等連絡会・実務者協議会、平成29年4月）

[http://www.kantei.go.jp/jp/singi/titeki2/digitalarchive\\_kyougikai/guideline\\_gaiyou.pdf](http://www.kantei.go.jp/jp/singi/titeki2/digitalarchive_kyougikai/guideline_gaiyou.pdf)

### 「第一次中間取りまとめ」（実務者検討委員会、平成30年4月）

[https://www.kantei.go.jp/jp/singi/titeki2/digitalarchive\\_suisiniinkai/jitumusya/2017/to\\_rimatome.pdf](https://www.kantei.go.jp/jp/singi/titeki2/digitalarchive_suisiniinkai/jitumusya/2017/to_rimatome.pdf)

※この文書において、ジャパンサーチ上で分野横断で共通で用いられるメタデータフォーマットを「共通メタデータフォーマット」という呼ぶことを定めた。

## 1.背景と現況

# 報告書「我が国におけるデジタルアーカイブ推進の方向性」

### デジタルアーカイブの意義

デジタルアーカイブは、文化の保存・継承・発信に必要であるのみならず、**観光や地方創生、教育研究、ビジネスへの利用など、新たな価値創出、イノベーション推進**にも貢献する取組。そのため、公的機関によるデジタルアーカイブの整備が強く求められる。



### デジタルアーカイブのメリット

○利用者から見ると・・・

- いつでもどこでもコンテンツにアクセスできる
- 教育コンテンツ作成、研究、商品・サービス開発の素材として、また、利用ログ情報をマーケティングデータなどに活用

○アーカイブ機関から見ると・・・

- 収蔵品の管理、貸借等の業務効率化
- 災害時の復興（防災対策）
- 館内展示サービスの充実
- 電子展示会等、ウェブサイトを利用したサービスの充実
- 来館者数・アクセス数の増加

自館の  
新たな価値創出

### 我が国の現状

○諸外国（EU、米国、オーストラリア、韓国等）では・・・

・国、地域ごとの統合ポータルを構築し、デジタルコンテンツのメタデータの集約、検索機能の提供、メタデータのAPI提供を実施。メタデータオープン化及びコンテンツの利用条件表示を推進。デジタルコンテンツの充実のための支援のほか、電子展示会や資料セット公開などの活用促進策を展開。



○日本では・・・

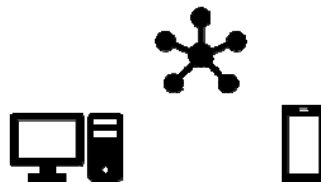
・書籍、公文書等一部アーカイブ構築が進んでいる分野もあるが、デジタルコンテンツの圧倒的不足、利用条件の不備、利用者ニーズに対応できないシステム設計などにより、活用が進まない。メタデータの整備・公開やアーカイブ間連携も不十分。

出典：「国におけるデジタルアーカイブの取り組み」（デジタルアーカイブ産学官フォーラム 資料1、内閣府知的財産戦略推進事務局、平成29年11月14日）  
<https://www.kantei.go.jp/jp/singi/titeki2/forum/2017/gjisisdai.html>

## 1.背景と現況

### 我が国の課題

- ・デジタルアーカイブ業務の位置づけ、評価の仕組みの導入
- ・中小機関及び地方における技術上・法務上の業務支援
- ・メタデータの標準化、オープン化の推進
- ・オーファンワークスの利用等に係る制度の整備



### デジタルアーカイブ社会における保存・共有・活用のサイクル実現を目指して

博物館・美術館等のコンテンツを保有する「**アーカイブ機関**」、分野・地域コミュニティの「**つなぎ役**」（博物館・美術館等の分野では、文化庁の文化遺産オンラインやメディア芸術データベースなど）、そして「**国の分野横断統合ポータル**」、多様なコンテンツを利活用する「**活用者**」、それぞれの取り組みを通じて、我が国のデジタル情報資源が効率的に生み出され、国全体として有効に活用されていくことを目指す。

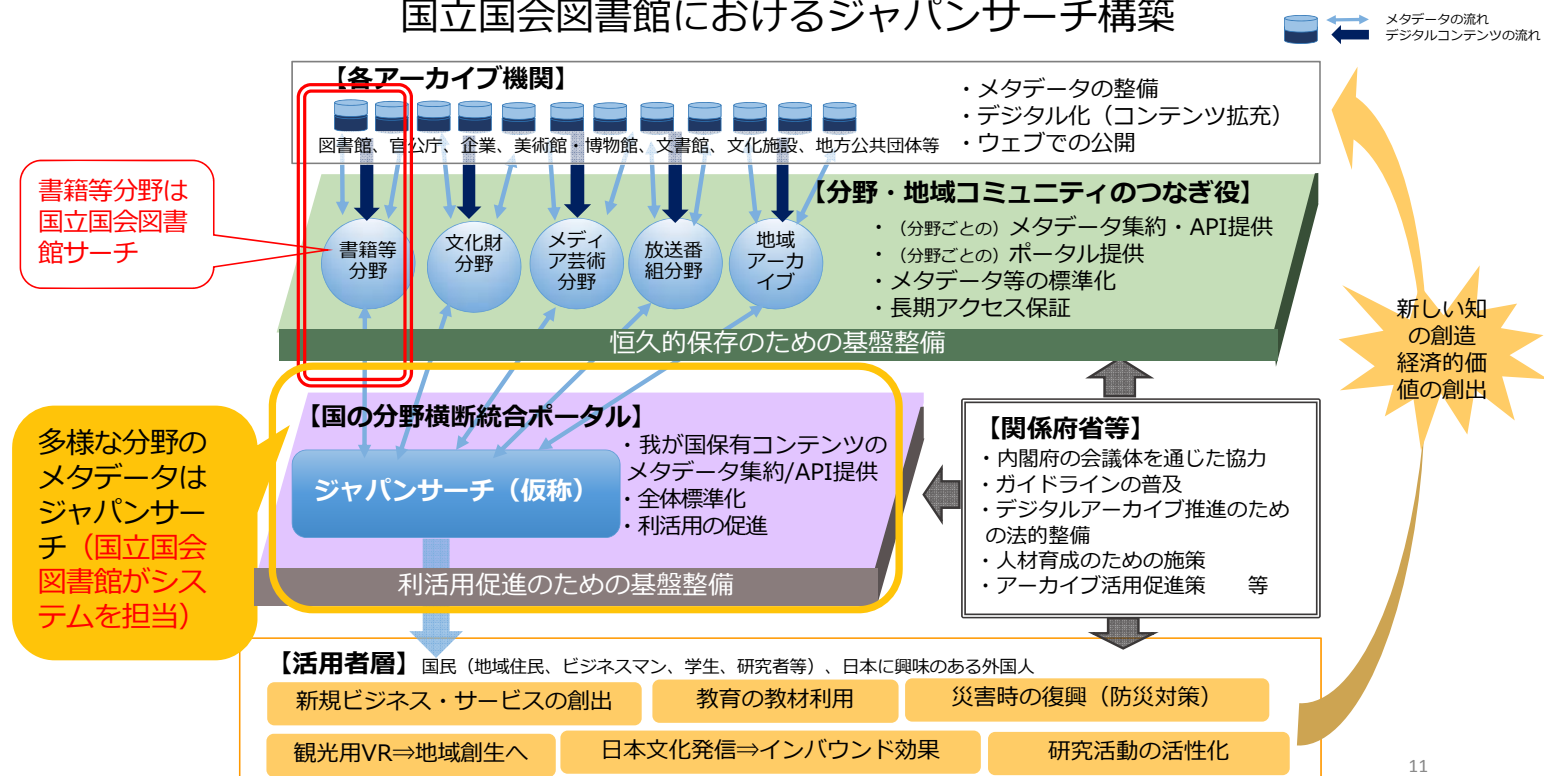
✓「各アーカイブ機関」は、ガイドラインに沿ったメタデータ（多言語化）の整備、オープン化などの取組を推進。技術・法令等の理解、プロデューサー能力・コミュニケーション能力を備えた人材を育成。取り組みを評価する仕組みを設計。

✓「つなぎ役」は、分野内のメタデータ項目の標準化、集約を推進。アーカイブ機関の技術、法務上の課題等に対応できる人材育成をサポート。デジタルアーカイブの評価指標設定。

✓「国」は、分野横断統合ポータルの構築を推進するほか、国・独立行政法人の有する書籍・文化財等文化的資産・メディア芸術・公文書のメタデータの整備やデジタル化を推進。活用促進のためのフォーラムを開催するとともにメタデータフォーマットの在り方や各アーカイブ機関やつなぎ役の取り組みへの支援について検討。

## 1. 背景と現況

### 国立国会図書館におけるジャパンサーチ構築



11

## 1. 背景と現況

### ジャパンサーチ構築の目的

ジャパンサーチ（仮称）との連携促進のための説明会「ジャパンサーチ（仮称）・フェーズゼロ～分野横断統合ポータル構築に向けて」（平成30年5月16日開催）説明資料より

#### 1. コンテンツの所在等の明確化

- ・メタデータの横断的な検索機能を提供し、我が国の様々な分野のコンテンツの所在情報等を明らかにする

#### 2. データ提供機関へのアクセス促進

- ・ジャパンサーチを経由して、個別のデータ提供機関へのアクセスにつなげる

#### 3. データの利活用の促進

- ・ライセンスの明示、APIの提供等により、データ利活用を容易にする。研究・教育、防災、ビジネスでの活用に加え、海外発信機能の強化につなげる

#### 4. データ提供機関への支援

- ・データ提供機関にとって便利な仕組みを提供することで、そのサービスやシステムを支援する

#### 5. 新規ビジネス・サービスの創出等

- ・以上を通じて、データに関する新規ビジネス・サービス創出及びデジタルコンテンツの保存・共有・活用を促進する

12



1.背景と現況

利活用のユースケース「キュレーションページ・電子展示会」

キュレーションページ

- 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、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



テキスト、  
画像、検索  
結果、検索  
画面の組  
合せ

両方に対応  
可能な工  
ディタ機能  
を用意

電子展示会

- コンテンツに解説を付けて一つのまとまりとして公開できるページ
- 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する



1.背景と現況

API・部品機能 ～利活用促進の基盤として

API機能

- 共通項目のデータの他に、データ提供元から収集したそのままの形のメタデータもAPIで提供
- メタデータアナライザの結果（項目特性に関する情報）もAPIで提供
- 検索API以外に、活用者がメタデータの大量取得を可能とするAPIも用意

部品機能

- APIと協調してブラウザ上で動作する部品を提供
- 部品は、ブログパーツのようなイメージでウェブサイトへ貼付可能
- 例えば、連携機関が自館データだけを対象にした（カスタム）検索画面や、自ら作成した電子展示会を自館HPへ簡単に掲載できる
- CSS（ウェブのレイアウト定義）によるデザイン変更も可能。

現在開発予定の埋め込み部品：

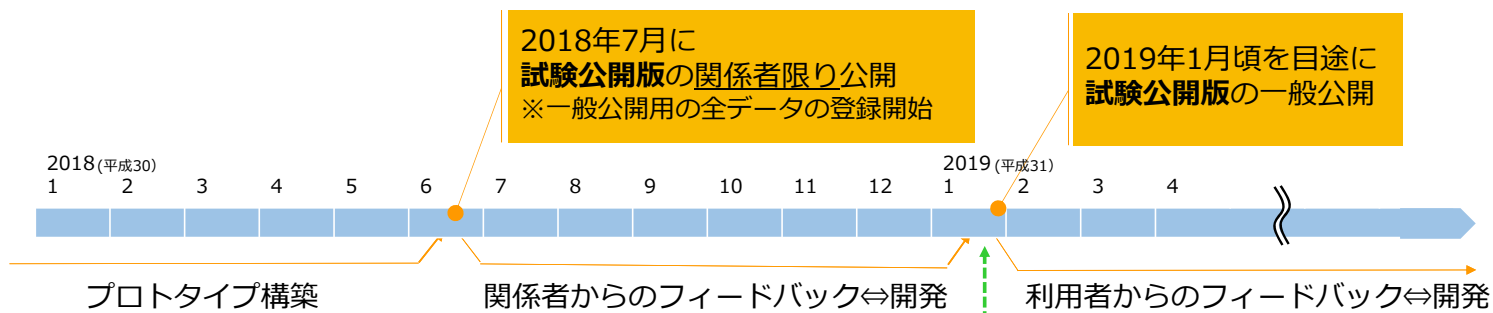
- 検索
- キュレーションページ
- 電子展示会

検索部品を埋め込んだ場合のイメージ



1.背景と現況

ジャパンサーチ開発スケジュール



アーカイブ機関・つなぎ役の皆さまへ お願い

一般公開までに

- メタデータは、原則CC0に※
- サムネイルはCC0/CC BY相当に
- デジタルコンテンツの公開を増やし、可能ならCC BY相当に

※解説文など著作権のある項目のみ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その利用条件が分かるような仕組みを系統的に用意します

※ジャパンサーチ（仮称）のデータを活用する際は、（CC0であっても）出典等の情報を明記するよう活用者をお願いする予定です

**オープンな利活用が可能なデータの整備・提供へのご協力を！**

1.背景と現況

試験公開版に向けて：メタデータの連携調整中データベース一覧

分野/タイプ	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
公文書	国立公文書館	国立公文書館デジタルアーカイブ アジア歴史資料データベース
文化財	文化庁	文化遺産オンライン（国指定文化財等データベース）
	国立美術館	独立行政法人国立美術館 所蔵作品総合目録検索システム
	国立文化財機構	ColBase 国立博物館所蔵品統合検索システム
メディア芸術 (ゲーム、アニメ、マンガ、メディアアート)	文化庁	メディア芸術データベース
	映像産業振興機構	Japan Content Catalog (JACC)
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
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
放送番組	日本放送協会	動画で見るニッポンみちしる
	放送番組センター	放送ライブラリー番組検索 ※ドラマ

## 2. 国立国会図書館における統合ポータル構築小史

19

### 「電子図書館構想」(1998年)

[http://dl.ndl.go.jp/view/download/digidepo\\_1000791\\_po\\_ndlelc-jpn.pdf?contentNo=1](http://dl.ndl.go.jp/view/download/digidepo_1000791_po_ndlelc-jpn.pdf?contentNo=1)

- 電子出版物の増大や資料の電子的形態での流通
- 印刷物において果たしてきた国立国会図書館としての役割を、電子出版物においても果たす
- 特に重要な電子図書館の枠組みとして
  - 電子図書館による国会サービスの強化
  - 国内で発行された電子出版物の収集と保存
  - 電子出版物の書誌作成と書誌的コントロール
  - 国民への電子出版物の提供と情報源へのナビゲーション
  - 資料の電子化及び電子図書館の構築における国内・国外の関係機関との連携・協力

20

## 「電子図書館中期計画」（2004年）

<http://www.ndl.go.jp/jp/dlib/project/plan2004.html>

### 「電子図書館サービス」の目標

#### 3. デジタル・アーカイブのポータル機能

(ア) 想定する機能

- 統合的な検索  
デジタル・アーカイブやOPACなどの情報資源を効果的に検索可能とするために統合的な検索機能を構築する。一定の範囲で統合した情報検索を可能とし、利用者を的確に情報源に導く。
- サブジェクト・ゲートウェイ  
利用者が主題に沿って系統的に情報資源を発見できるよう案内する機能を提供する。
- 情報探索のオンライン・チュートリアル  
情報の探し方を利用者にオンラインで提供する機能を用意する。

(イ) 「日本のデジタル・アーカイブ・ポータル」（仮称）

当館及び国等の公的機関を対象としたポータル機能を構築する。外部機関のウェブ・アーカイブ構築の動向を見て日本全体のウェブ・アーカイブのポータル機能構築も考慮する。

21

#### ポータル機能の実現方法（「電子図書館中期計画」より）

- **メタデータ収集における共通仕様の確保**  
ポータル構築の基盤として、さまざまな情報提供機関からメタデータを収集し利用する共通の枠組みを確保し、実装・利用の推進を行う。
- **運用体制の確立**  
関係機関との間においては、ポータル機能等の運用に係る協力及び継続した調査研究等の協力を行う。

22

## “DC-NDL”

情報資源の組織化および利用提供のための国内メタデータ標準

2001年	<b>「国立国会図書館メタデータ記述要素」</b> <ul style="list-style-type: none"> <li>「ダブリンコアメタデータ基本記述要素集合」(1999年版)に基づく基本15要素</li> <li>WARPで集めるインターネット上の情報資源を記述するため</li> </ul>
2007年	<b>「国立国会図書館ダブリンコアメタデータ記述要素」</b> <ul style="list-style-type: none"> <li>日本国内の図書館・関係機関等におけるメタデータの相互運用に資するための標準</li> </ul>
2010年	<b>「国立国会図書館ダブリンコアメタデータ記述」</b> <ul style="list-style-type: none"> <li>DCMIメタデータ語彙55要素 + NDL独自定義語彙</li> <li>JAPAN MARCの項目にほぼ対応</li> <li>語彙の記述方法や形式についてのルール</li> <li>機械的な処理を可能にするセマンティックウェブ志向</li> <li>国立国会図書館デジタルコレクション」の情報資源を記述</li> <li>「NDLサーチ」の収集・提供用標準フォーマットで使用</li> </ul>
2011年	<b>「国立国会図書館ダブリンコアメタデータ記述」</b> (現行版)

23

## PORTA (国立国会図書館デジタルアーカイブポータル)

- 様々な機関・団体が提供しているデジタルアーカイブに格納されたデジタル資料をワンストップで検索できるポータルサイト (デジタル資料があるものへのナビゲーション)
- 他のシステムとの間で、相互に検索機能を利用し合える仕組みを提供 (API連携の幕開け)
- DC-NDL 2007年版を拡張した、DC-NDL\_PORTAを標準フォーマットとして連携
- 拡張した要素
  - サムネイル情報
  - 管理情報 (データプロバイダ情報など)
  - エンコーディングスキーム等 (資源タイプなど)

PORTA (国立国会図書館デジタルアーカイブポータル) の検索画面のスクリーンショット。検索欄、ナビゲーションメニュー、検索オプション、お知らせ、資料ピックアップなどの要素が確認できる。

24

## 2.ポータル構築小史

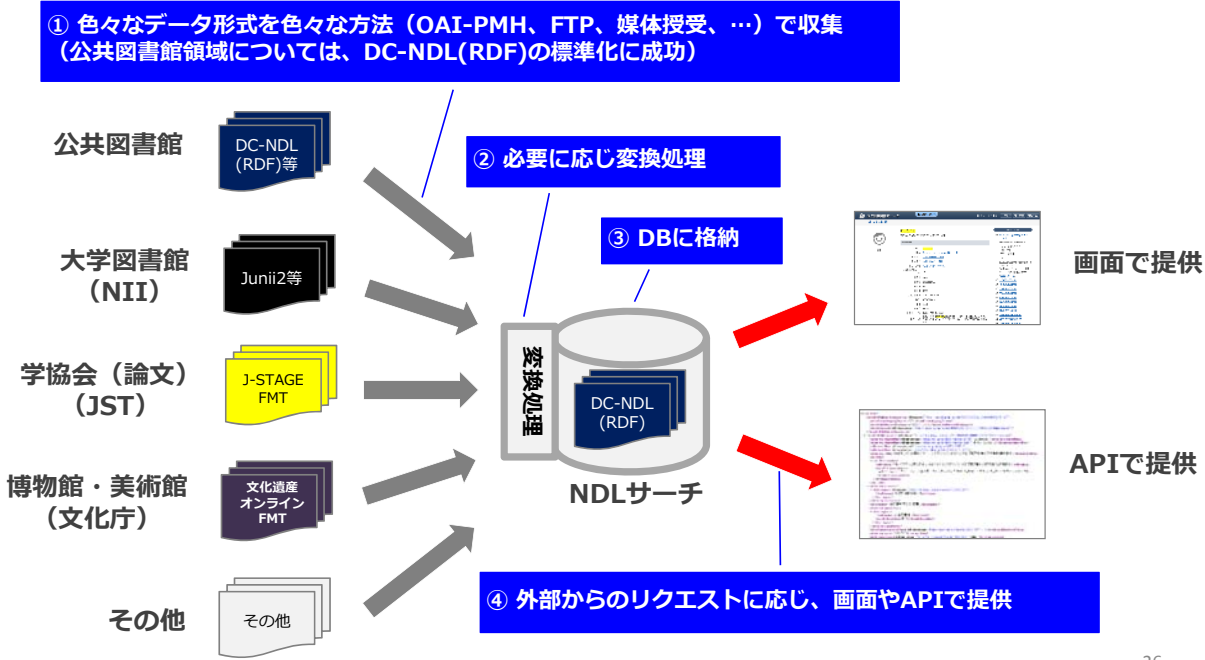
### NDLサーチ（国立国会図書館サーチ）



- PORTAや3つの総合目録を統合、検索対象拡大
  - 国立国会図書館所蔵の全ての資料（一次情報、二次情報）
  - 都道府県や政令都市立の図書館の所蔵資料
  - PORTAで検索可能になっていたデジタル情報をベースに順次連携対象を拡張
- 現在、「DC-NDL2011年版」をRDF/XML形式で表現する「DC-NDL (RDF)」により連携、「書籍分野」に最適化した統合検索サービスが実現
- 今年7月にJPOとの連携により、近刊情報の提供が実現

## 2.ポータル構築小史

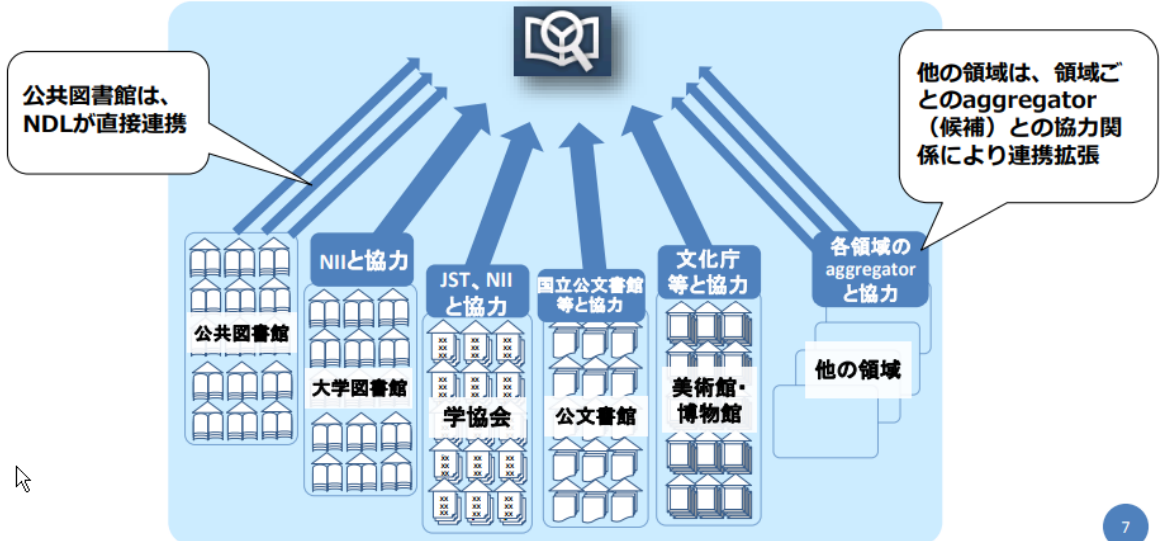
### 国立国会図書館サーチにおけるメタデータの連携と提供（現在）



2.ポータル構築小史

「国立国会図書館サーチ連携拡張にかかると実施計画」（2015年3月）

NDLサーチの統合検索サービス提供における連携イメージ



出典 : [http://iss.ndl.go.jp/information/wp-content/uploads/2015/02/20150122symposium\\_ozawa1.pdf](http://iss.ndl.go.jp/information/wp-content/uploads/2015/02/20150122symposium_ozawa1.pdf) 27

2.ポータル構築小史

ひなぎく（国立国会図書館東日本大震災アーカイブ）

総務省とNDLが構築（-2013年3月）

国として、東日本大震災に関するあらゆる記録を次世代へ伝えることを目的

DC-NDL 2010年版を拡張した、国立国会図書館東日本大震災アーカイブメタデータスキーマ（2013年3月公開、2016年10月改訂）

- 重要となった要素
- 場所に関する情報
  - 日時に関する情報
  - 利用条件情報
  - 資料種別
  - 技術情報



### ひなぎく（国立国会図書館東日本大震災アーカイブ）

- NDLにおいては「『知識インフラ』構築の先行的事業の一つ」という位置付け。閲覧だけでなく、震災や復興の記憶・教訓が活用（二次利用）されることを目的にした
- 図書や雑誌記事などの文書資料のほか、動画、音声、写真、Webページ等を包括的に収集、ファクトデータ（生データ）も対象にした（平成29年度末の数字は、メタデータ376万件、連携アーカイブは47、コンテンツは52機関から収集）
- メタデータ作成者は非専門職（ボランティア含む）を想定
- ユーザに提供する機能としては以下を重視
  1. どの機関・アーカイブがどのようなコンテンツを持っているのかを提示でき、コンテンツにたどり着けること（ポータル機能）
  2. 場所や日時で絞り込めること ※現在は「カテゴリ検索」内で実現

29

### ひなぎくのメタデータスキーマで拡張または特に配慮した要素

- 場所に関する情報
  - 「コンテンツの主題としての場所」  
dcterms:spatialを構造化。緯度経度情報をGeo Vocabularyの語彙、地名・住所情報をvCard Ontologyの語彙で格納。緯度経度の値は世界測地系の形式に正規化
  - 「コンテンツが作成された場所」  
緯度経度情報をOntology for Media Resources 1.0 の語彙、地名・住所情報をvCard Ontologyの語彙で格納
- 日時に関する情報
  - 「コンテンツが公開された日時」をdcterms:created, 「コンテンツが公開された日時」をdcterms:issued, 「コンテンツの主題としての日時」をdcterms:temporalに格納、W3CDTFで正規化
  - ※課題として、時間的範囲を表す標準的で普及した語彙がないこと

30

### ひなぎくのメタデータスキーマで拡張または特に配慮した要素

- 利用条件情報  
「活用」のために重要な情報。dcterms:accessRightsにインターネット公開、NDL館内閲覧限定かのコード値を、クリエイティブコモンズのライセンスが表示されている場合は、dcterms:licenceに該当ライセンスのURIを、詳細な文言による利用条件の記述はdcterms:rightsに格納。
- 資料種別  
DC-NDLのNDLタイプ語彙を使用
- 技術情報  
「フォーマットの名称」をpremis:formatNameを、「フォーマットのバージョン」をpremis:formatVersionを採用

### ひなぎく = 「知識インフラ」の先行例の構築において見えてきたこと

- 地名・住所情報の記述の標準化が困難（地方公共団体コード？→改訂！市町村合併！）
- 時間的範囲を表す標準的で普及した語彙がない
- 利用条件の標準的な表示方法が確立していない
- NDLのタイプ語彙はContent Type, Carrier Type, Media Typeを整理できていない。項目が多く複雑
- 保存のためのメタデータについての知識とリソースの不足
  
- 「非専門職または多様な分野の専門職が作成するコンテンツやメタデータ」と「ユーザフレンドリーなGUI」「活用しやすいデータ提供」の間の、インフラ側の負担・・・

なお、ひなぎくのメタデータ連携における基本的な考え方・・・

- 提供元のアーカイブで付与したメタデータを可能な限り欠落させない
- 各メタデータ項目について、利用シーンを考慮して、マッピングの考え方を分ける  
(例) GUIやWeb APIで詳細検索やファセットに使用する項目については、マッピングの際に値を正規化、検索に使用しない項目については元の値をそのまま格納など

### 3. 海外の分野横断統合ポータル

(平成28年度「ジャパンサーチ（仮）のメタデータ要件に係る  
検討整理のための簡易調査」より)

「ジャパンサーチ（仮）のメタデータ要件に係る検討整理のための簡易調査」

- 「ひなぎく」構築において見えてきた課題
- 「国の分野横断統合ポータル」構築において期待されること



平成28年度に職員による「簡易調査」

- 想定されていた器（NDLサーチ）を点検してみる
- 海外のポータル構築例を見してみる

NDLサーチ（DC-NDL連携方式）との課題（簡易調査より）

- これまで、書籍等の複製資料が主な対象であったため、美術館・博物館における単一の文化財や、階層体系で管理される公文書類を十分に扱えるメタデータ設計となっていない。
- DC-NDLは、ダブリンコアをベースとしているが、全てのプロパティを含むわけではない。またNDLサーチはDC-NDLで採用しているプロパティを全て使用しているわけではない。
- アグリゲータの個別の参加機関（データ提供元の機関）を適切に記録する語彙がないなど、アグリゲータとの連携を前提とした体系になっていない。

一方、DC-NDLの大幅改訂は、既存ユーザ（主に図書館）に影響を与える。

#### ジャパンサーチ構築にあたり求められる要件（簡易調査より）

- アグリゲータを介した連携においても、アグリゲータ傘下の各機関の名称を検索、表示、提供時に出力できるようにする（参加機関のインセンティブ向上のため）
- デジタル化コンテンツの有無や、ライセンス情報の区分で絞り込みができる
- 従来図書館では扱ってこなかった各種文化財についても、適切な形で検索、表示、提供できるよう、必要な情報を記録できることが望ましい。例えば、材質、構造、技法などのモノに関する情報、その由来情報、地理的な情報、著者以外の関係者にかかる情報など。一方、博物館等における標準的なメタデータのうち、専ら内部の資料管理のための項目はジャパンサーチではなくてもよいのでは。

#### Europeana Data Model (EDM)

- 発足当時はESE（Europeana Semantic Elements）、提供されたデータの損失を生じさせないためEDMへ
- GUI画面でのデータ提供のほか、各種API（REST, OAI-PMH, SPARQL）を介してEDM形式のデータ提供（RDF, JSON, JSON-LD）
- ベースになった既存語彙
  - OAI ORE オブジェクトのメタデータとデジタル表現の関係づけ
  - Dublin Core 記述メタデータ
  - SKOS 概念語彙の表現
  - CIDOC-CRM イベント、オブジェクト間関係づけ

## Europeana Data Model (EDM)

	要件	実現した方法
1	提供された文化遺産機関の資料（以下「オブジェクト」）とそのデジタル表現の分離	クラス [edm: ProvidedCHO]と[edm: WebResource]
2	オブジェクトとそのレコードの分離	クラス [edm: ProvidedCHO]と 記述的プロパティ群
3	同一のオブジェクトに対する複数のレコードの許容	クラス [ore: Aggregation] と[ore: Proxy]
4	他のオブジェクトから成るオブジェクトのサポート	階層関係や順序を表現するプロパティ群
5	抽象レベルの異なる記述との互換性と維持	オブジェクトとイベントに焦点を当てたモデル
6	詳細化も可能な標準的メタデータ形式	既存の語彙の利用と既存の標準とのマッピング
7	統制語彙由来の概念など文脈的リソースのサポート	セマンティックウェブ対応とツール

参考 : Europeana Data Model Premier.

[https://pro.europeana.eu/files/Europeana\\_Professional/Share\\_your\\_data/Technical\\_requirements/EDM\\_Documentation/EDM\\_Primer\\_130714.pdf](https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf)

福山樹里. Europeanaのメタデータ : デジタルアーカイブの連携の基盤. 情報の科学と技術. Vol.7. 2017. no.2

39

## Europeana Data Model (EDM)の鍵となる概念（実体？）

**アグリゲーション (ore: Aggregation)**

- 単一のリソースとして扱えるように1つに集められた、関連リソースの集合
- EDMにおいては、一つのデータプロバイダから提供された、一つのオブジェクトに関連するリソースの集合として扱うための実体。

**プロキシ (ore: Proxy)**

- ある特定の集合体（Aggregation）の中にある特定のリソースを表す。
- EDMにおいては、オブジェクトの代理となつて、特定のデータプロバイダから見た特定のオブジェクトについての記述の対象になる。
- 同一のオブジェクトに対して複数のプロバイダのメタデータレコードを許容するため、また元データを維持しつつ、提供されたメタデータに文脈を付加したEuropeana独自のデータを追加的に持つために必要な実体。

40

EDMのアグリゲーション (ore: Aggregation)

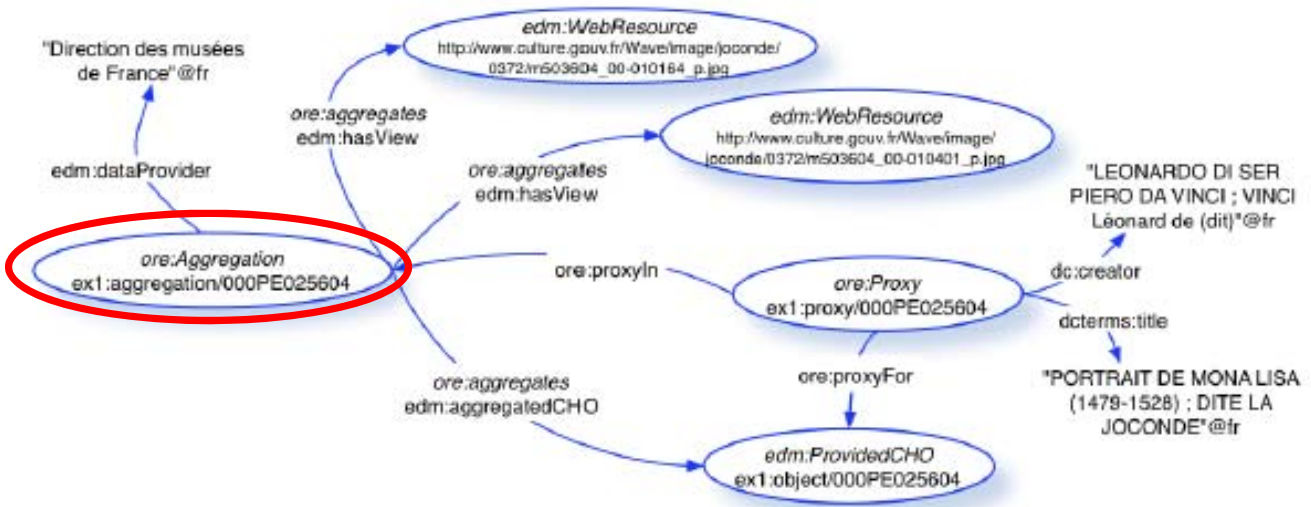


Fig. 14. Provider's aggregation, provided object and proxy—simple case with only one provider for the object



1 データプロバイダにつき、1 アグリゲーション、1 プロキシ

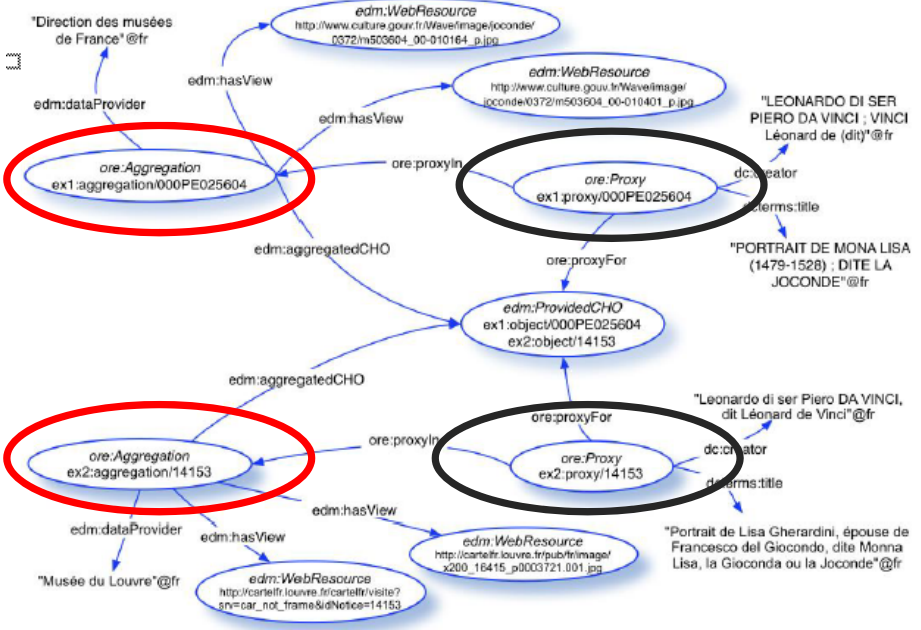


Fig. 15. Providers' aggregations, provided object and proxies—complex case with two providers for the object



3. 海外の統合ポータル

Europeana内で変換され、文脈情報を付加されたメタデータ



Fig. 16- Europeana aggregation—simple case with only one provider for the object

3. 海外の統合ポータル

イベントモデル

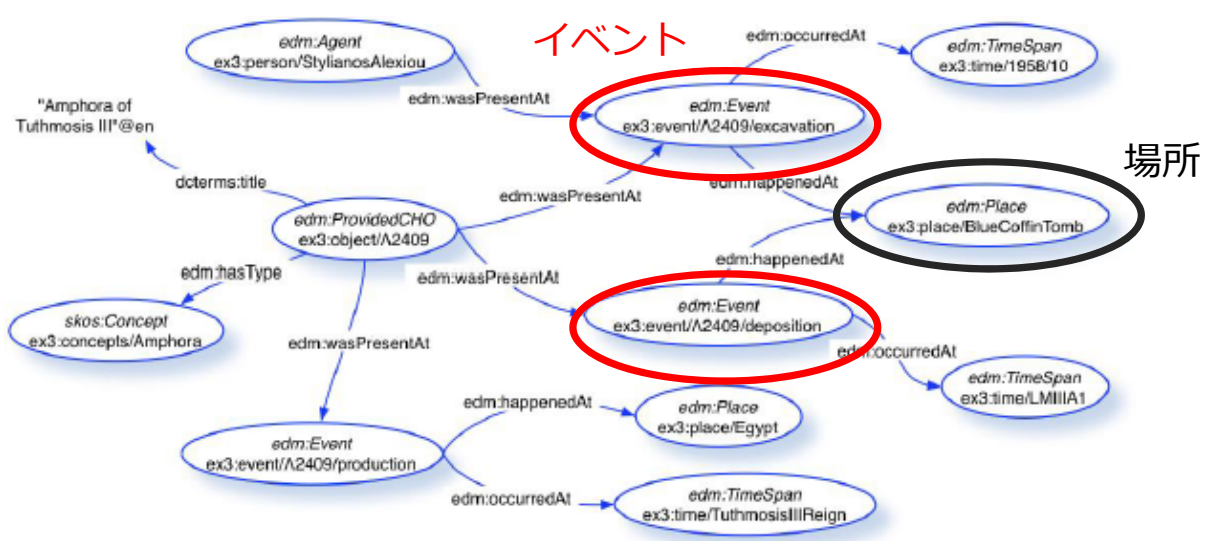


Fig. 11. Amphora of Tuthmosis III – an event-centric description (without aggregation and related digital representation entities)

## EDMでの必須項目

1. タイトル
2. コンテンツの言語（テキスト資料の場合）
3. デジタルコンテンツの種別
4. コンテンツの文脈や詳細についての情報（主題、分野、場所、空間的・時間的特性）
5. ユーザ生成コンテンツかどうか
6. データ提供機関の情報
7. アグリゲータの情報
8. デジタルコンテンツへのリンク
9. 権利情報
10. ユニークで固定の識別子

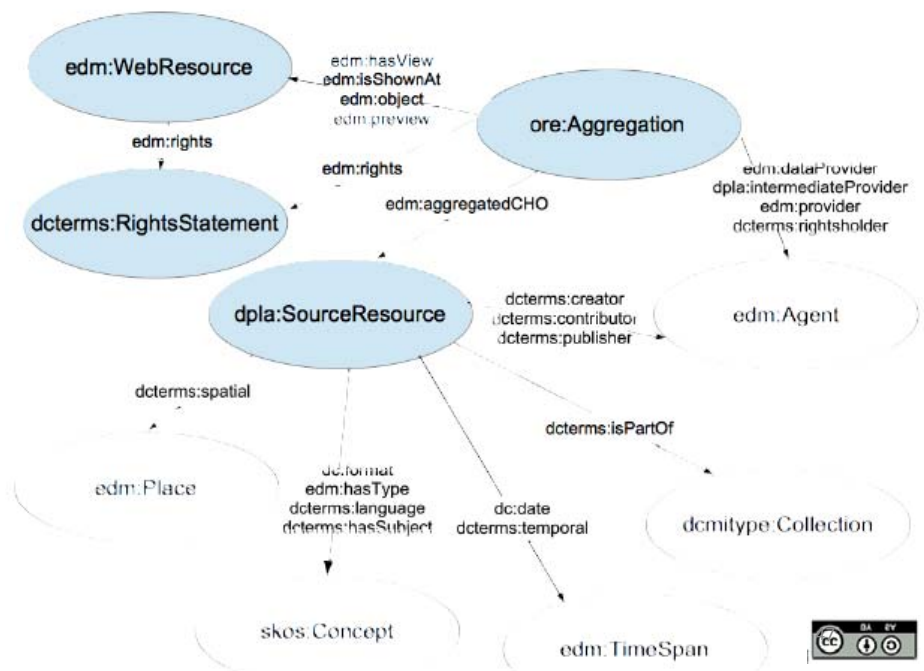
## Digital Public Library of America (DPLA) <https://dp.la/>

- 米国各地の図書館・博物館・文書館等が有するデジタルコンテンツのメタデータを集約し、制約なく市民に提供するプロジェクト
- 米国内40機関、2200万件の画像、テキスト、動画、音声などのメタデータのみ収集（デジタルファイルは収集しない）

## DPLA Metadata Application Profile (MAP)

- EDM をベースに、一部モデルを簡素化
- 2012年にver.1 公開。最新版はver.5 (2017年12月公開)

DPLA内で生成された正規化記述は、`dpla:SourceResource`でまとめられる  
 ソースデータは`originalRecord`として`Aggregation`の下位で持つ



## DPLA MAP の連携

- DPLAにおいて、表記の正規化と場所情報と時間情報の充実が行なわれる
- マッピングやデータ登録ツールの提供

## Schema.org

- 米国のMicrosoft, Google, Yahoo!の検索エンジン大手3社が2011年6月に共同で立ち上げ（のちにロシア最大の検索エンジンYandexも参加）。2012年から2014年に普及。
- ウェブページに含まれた様々な情報を共通の方法でマークアップする方法を定めようとする取り組み
- 2015年からは W3C Schema.org Community Groupで検討されている（最新バージョンは3.4）
- Bibliographic Extention, Architypes Extention

49

## schema.orgによるマークアップ が普及すると？

### 「メタデータによる検索の優れた点」（2010年時点）

谷口祥一. メタデータの「現在」：情報組織化の新たな展開. (勉誠出版, 2010年) pp.10-12. の報告者による言換え

- ① 構造化されているため、精度の高い検索が可能
- ② 値に統制形や記号を用いた網羅的な検索が可能
- ③ 論理的な単位を対象にできる
- ④ 多種多様な情報資源に対して統一した検索が可能
- ⑤ 対象とする情報資源の多様な側面（対象ユーザや利用条件など）を適切に表現可能

検索エンジンによるウェブページの検索も近づいていく

50

Google Library of Congress

約 932,000,000 件 (0.74 秒)

**Library of Congress: Home**  
[www.loc.gov/?](http://www.loc.gov/) このページを訳す  
 The Library of Congress is the nation's oldest federal cultural institution, and it serves as the research arm of Congress. It is also the largest library in the world, with more than 162 million items. The collections include books, sound recordings, ...

loc.gov からの検索結果

<p><b>American Memory</b> American Memory is a gateway to rich primary source materials ...</p> <p><b>Online Catalog</b> The Library of Congress Online Catalog contains approximately ...</p>	<p><b>Exhibitions</b> The Library of Congress offers a wide range on online ...</p> <p><b>Research and Reference</b> Research and Reference Services Library of Congress collections ...</p>
--	--

**アメリカ議会図書館 - Wikipedia**  
<https://ja.wikipedia.org/wiki/アメリカ議会図書館>  
 アメリカ議会図書館（アメリカざかいとしょかん、英語: Library of Congress）は、アメリカ合衆国の国立図書館。蔵書数、予算額、職員数全ての点で世界最大規模の図書館である。略称はLC。所蔵品の点数は数千万冊の書籍や各種資料など一億点を超える。合衆国連邦政府の立法府に属する機関であり、1800年に首都ワシントンD.C.に設立された。ウィキペディア

議会図書館の業務 · 資料の収集 · 保存 · 図書館サービス · 組織

**アメリカ議会図書館**  
 ウェブサイト 経路案内 保存  
 アメリカざかいとしょかん  
 4.7 ★★★★★ Google のクチコミ (766)  
 政府機関

アメリカ議会図書館は、アメリカ合衆国の国立図書館。蔵書数、予算額、職員数全ての点で世界最大規模の図書館である。略称はLC。所蔵品の点数は数千万冊の書籍や各種資料など一億点を超える。合衆国連邦政府の立法府に属する機関であり、1800年に首都ワシントンD.C.に設立された。ウィキペディア

所在地: 101 Independence Ave SE, Washington, DC 20540 アメリカ合衆国

## schema.orgの語彙を利用してLinked Data公開

### World Cat Linked Data

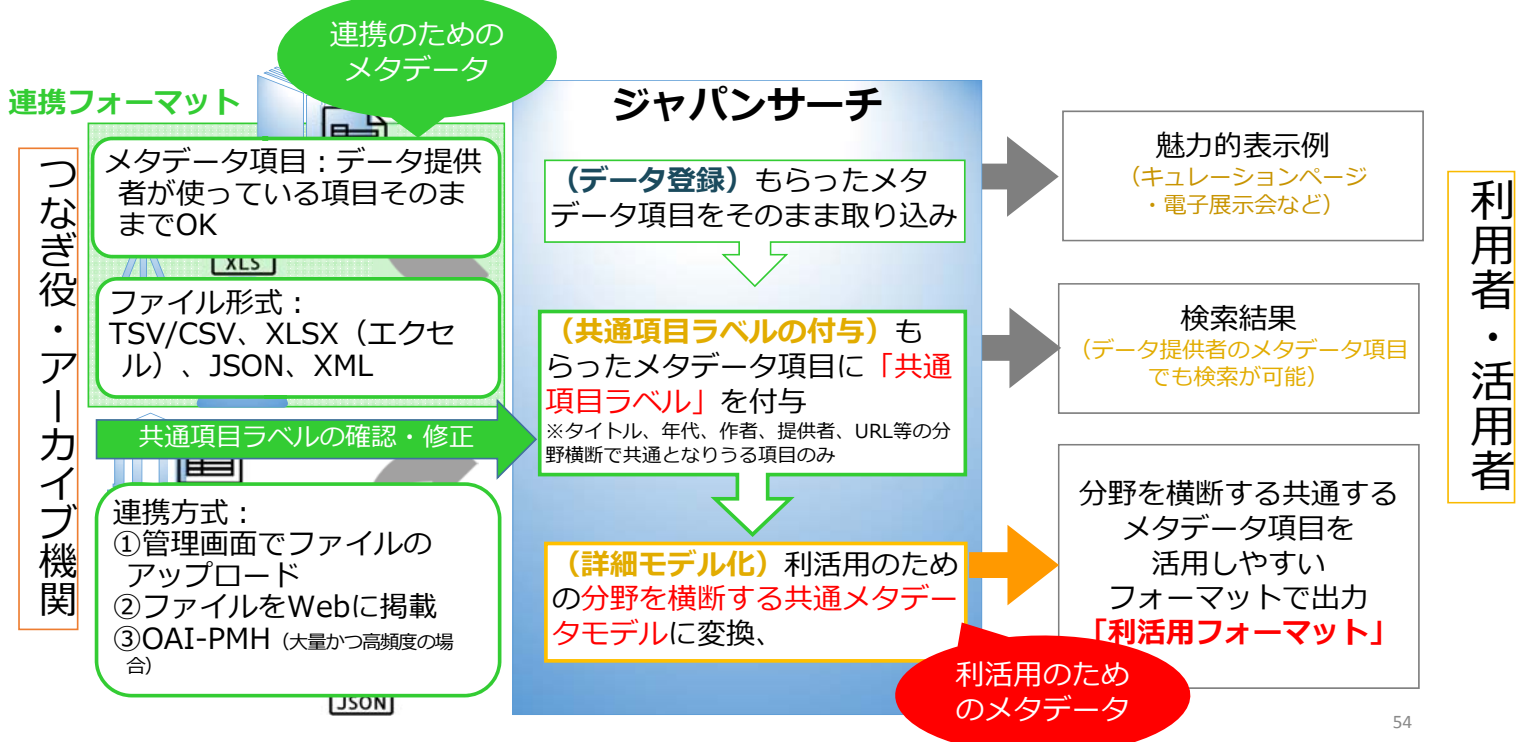
*The Schema.org ontology, extended by other ontologies as needed, is being used as a basis for modeling WorldCat bibliographic data. This provides fullest coverage, improved functionality, and better consumption of this data for our harvest partners and library members looking to do more with structured data exposed within HTML pages (via RDFa 1.1).*

### Europeana ? (R.Wallis らによる提言論文)

*Europeana Collections is fuelled by metadata describing cultural objects, represented in the Europeana Data Model (EDM). This paper presents the research and consequent recommendations for publishing Europeana metadata using the Schema.org vocabulary and best practices. Schema.org html embedded metadata to be consumed by search engines to power rich services (such as Google Knowledge Graph). Schema.org is an open and widely adopted initiative (used by over 12 million domains) backed by Google, Bing, Yahoo!, and Yandex, for sharing metadata across the web. It underpins the emergence of new web techniques, such as so called Semantic SEO.*  
<http://journal.code4lib.org/articles/12330>

# 4. ジャパンサーチにおけるメタデータの集約と変換

## ジャパンサーチにおけるメタデータの集約と変換



4. メタデータの集約と変換

メタデータ収集、マッピングから検索ができるまで（概念図）

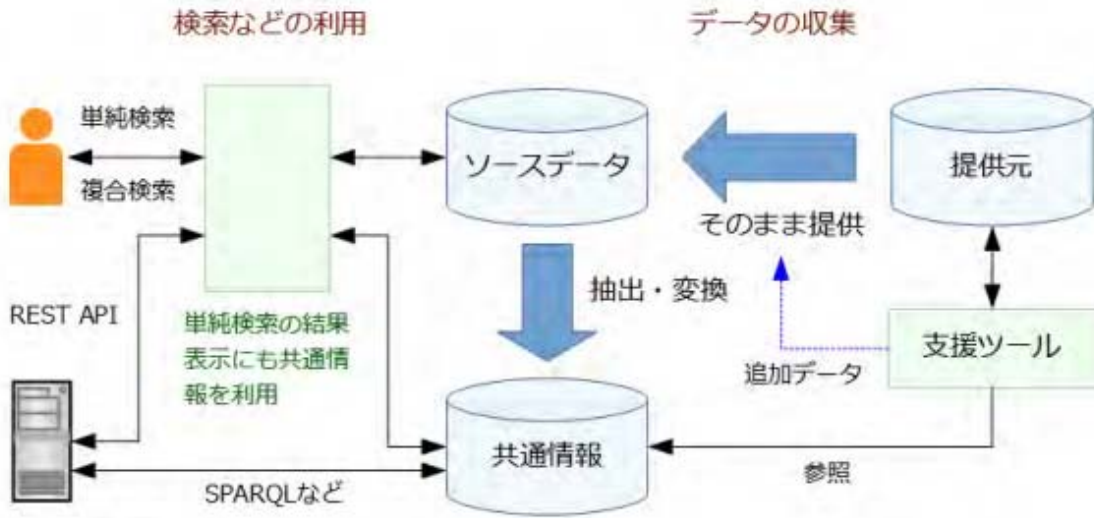


図 3: 提供元からのデータを共通アーカイブ情報に変換して検索などに利用する流れ

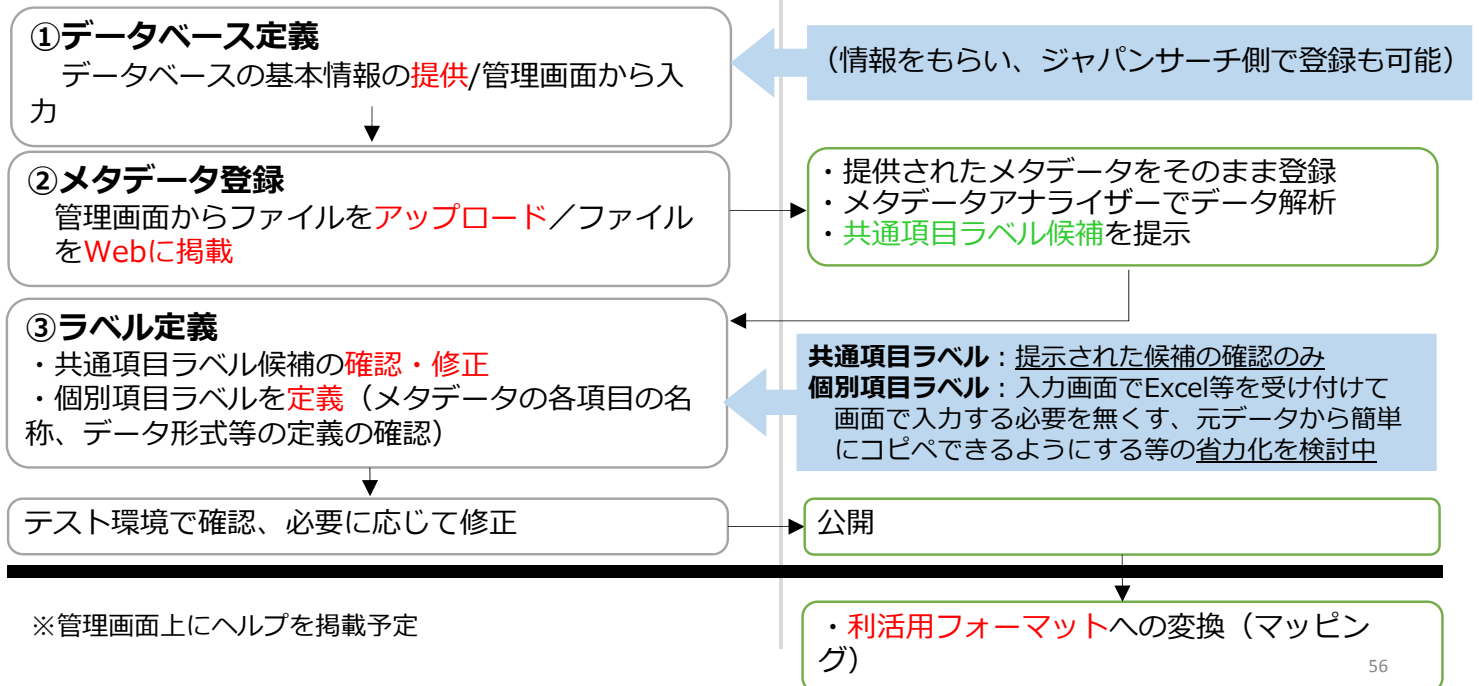
4. メタデータの集約と変換

メタデータ連携の流れ（試験公開版での実装）

データ提供機関側

※赤字がつなぎ役・アーカイブ機関の行う作業

ジャパンサーチ（仮称）



## 4. メタデータの集約と変換

### ① データベース定義

#### ◆ データベース基本情報の登録

The screenshot shows a form for entering database information. Fields include:
 

- データベースのID: test1
- 名称: 国立国会図書館デジタルコレクション
- 名称(ヨミ): こくりつこっかいとしょかんでじたるこれくしょん
- 名称(英語): National Diet Library Digital Collection
- データベースの説明: (empty)
- データベースの説明(英語): (empty)
- カテゴリ: (empty)

管理画面（データベース基本情報）

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（日本語/英語）	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（日本語/英語）	データベースの説明（日本語/英語）	紹介ページ上は100字まで表示（100字以上は折り畳み表示）
必須	データベースのカテゴリ	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
	データベースのサブカテゴリ	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・絞り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	

57

## 4. メタデータの集約と変換

### ② メタデータ登録

#### ◆ 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。

58

## 4. メタデータの集約と変換

### ③ラベル定義

#### ◆ 共通項目ラベルを付与する項目

共通項目ラベル		
ID	必須	国立国会図書館書誌ID
名称/タイトル	必須	タイトル
名称/タイトルヨミ	あれば必須	URL タイトル シリーズ 著者
名称/タイトル英語	あれば必須	出版者 出版年月日等 資料種別
最終更新日	あれば必須	大きさ、容量等 分類 (HDL, HDLC, HDCC等)
URL	あれば必須	請求記号 国立国会図書館書誌ID
サムネイル画像URL	あれば必須	ISBN
提供者	あれば必須	
コンテンツの権利表示	あれば必須	
寄与者	任意	著者 × +

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
任意	提供者	オリジナルのコンテンツの提供者
	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 (複数可)
	時間 (日付) ※	制作年、対象時期など (複数可)
	場所 ※	発行地、制作地など (複数可)

※あれば原則として付与するが、分野の特性に応じて付与しない場合もありうる。

#### ◆ 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。

59

## 4. メタデータの集約と変換

### 「デジタルアーカイブの構築・共有・活用ガイドライン」

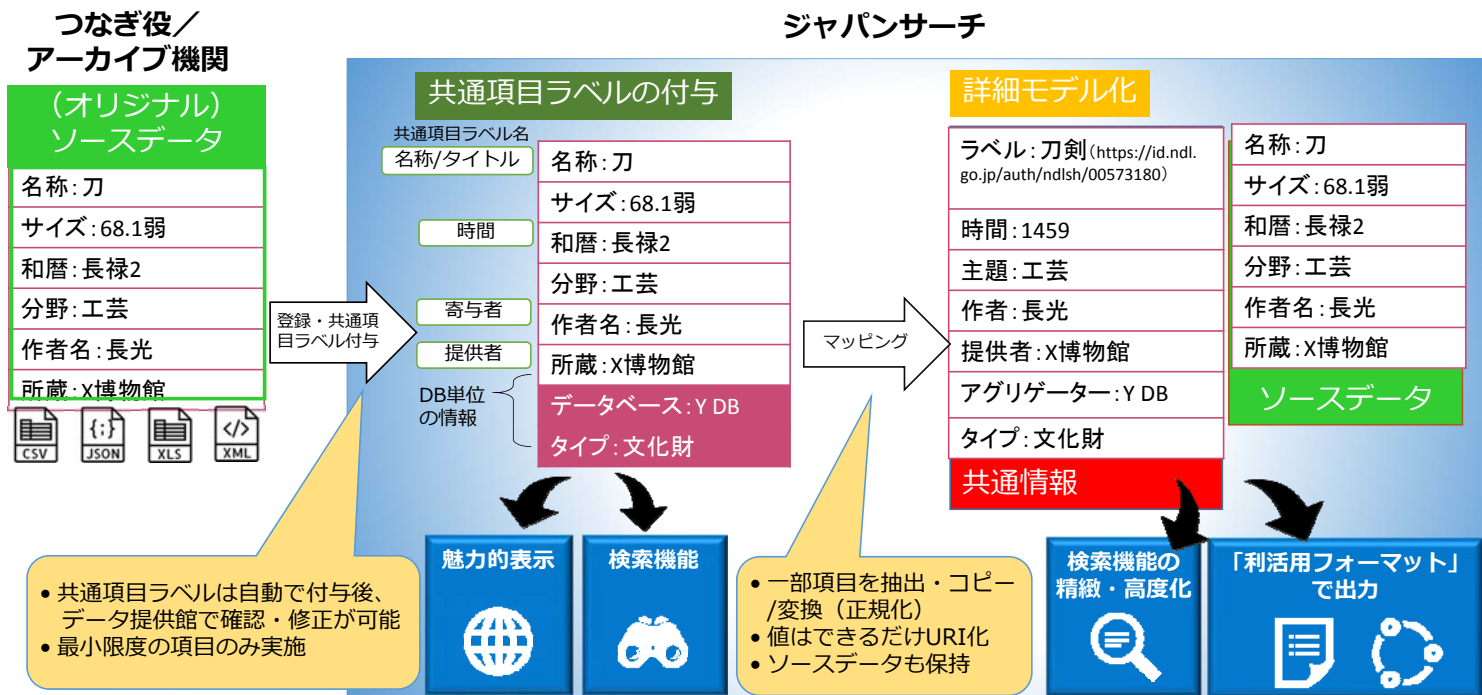
「自らが保有するコンテンツのメタデータを整備して共有する方法には、自らが用意したシステムを用いるほか、分野や地域のコミュニティが提供する既存のプラットフォームを利用する方法がある。また、これらのシステム基盤を利用せずに、参考資料「活用できる表形式のデータとは？」に示す要件を満たす形のCSV等のデータを整備する方法もある。<中略>メタデータの整備においては、次の5つの項目は共有や再利用のために特に重要であり、判明している場合には、必須の情報として記述することが求められる。

- ・タイトル (ラベル)
- ・作者 (人物)
- ・日付 (時代)
- ・場所
- ・管理番号 (表内で重複しない恒久的な識別子) 」

このほか、デジタルアーカイブのデータ共有のためには**コンテンツの権利情報や二次利用条件**といった情報もメタデータの整備において求められる。

#### 4. メタデータの集約と変換

### 共通情報を「利活用フォーマット」として出力



#### 4. メタデータの集約と変換

### 利活用フォーマット概要

#### ◆ 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

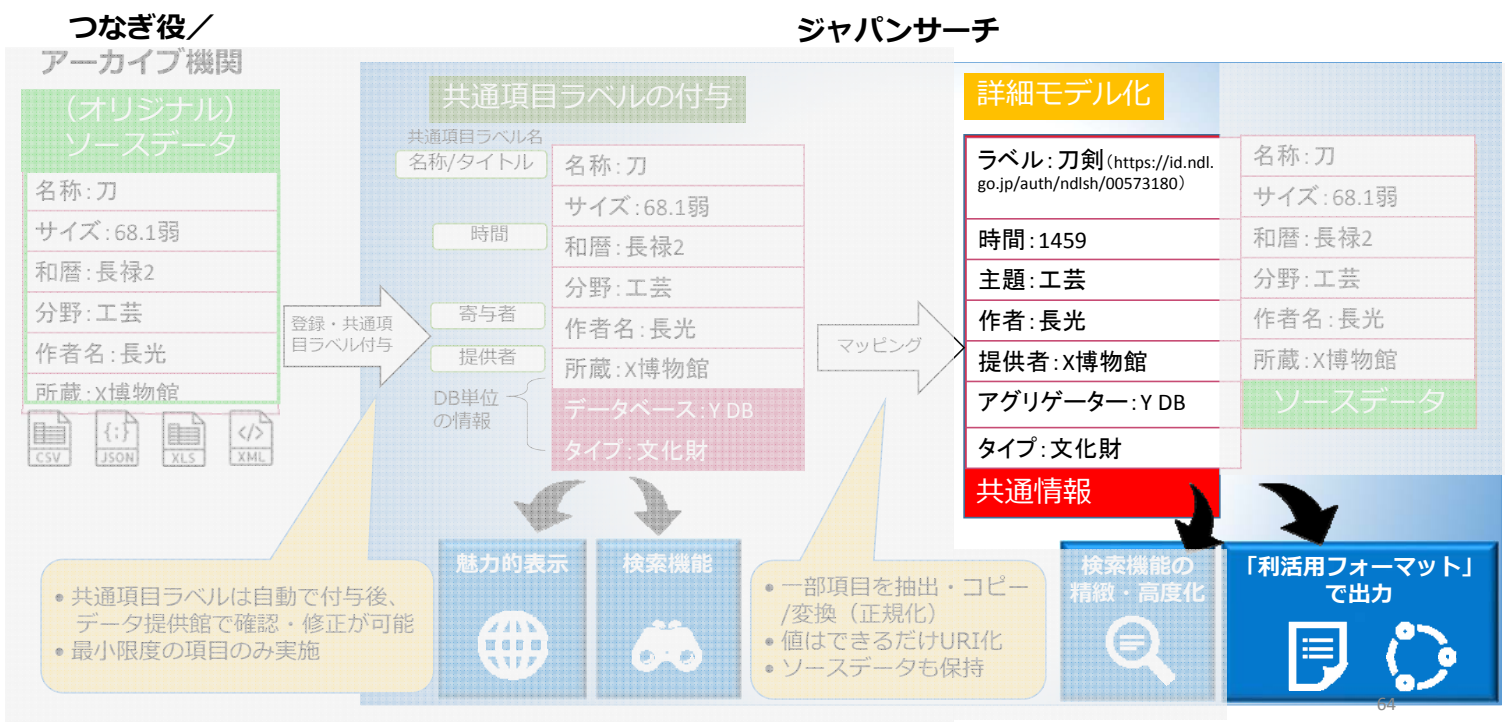
- ✓ 提供元からのソースデータを来歴情報明記の上でそのまま保持
- ✓ ソースデータのうち、利用者の4つのタスク（発見・識別・選択・取得）に特に有益な項目を共通の形式に変換した共通情報を新たに生成し保持

#### ◆ 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

## 5. ジャパンサーチの「利活用フォーマット」

### 共通情報を「利活用フォーマット」として出力



## 平成29年度の実施内容

### 「ジャパンサーチ利活用フォーマット概念モデルと語彙の検討」

平成29年度に国立国会図書館が、ゼノン・リミテッド・パートナーズの支援を受けて検討。

成果物

「概念モデルと語彙検討報告書」

「利活用メタデータフォーマット仕様案」及び仕様案を用いたデータ例

※ <http://www.ndl.go.jp/jp/dlib/standards/jpsformat.html>に掲載

「外部提供API仕様案」

報告書執筆者 神崎 正英（ゼノン・リミテッド・パートナーズ）

当館職員、外部有識者、検討支援受託者から成る検討会を開催（3回）

検討会に参加いただいた外部有識者（五十音順、敬称略）

嘉村 哲郎（東京藝術大学芸術情報センター）

寺澤 正直（内閣府大臣官房公文書管理課）

村田 良二（東京国立博物館学芸企画部）

※ 所属は平成 30 年 3 月現在

### 検討の前提

#### • 「知的財産推進計画2017」に示された統合ポータル像

＜目的＞

分野・地域を超えた知を集約し、学術研究・教育・防災・ビジネスへの利活用に資する海外発信機能を付加・強化することにより、インハウントの促進や海外における日本研究を活性化する

＜対象資料＞

知的・文化的資源（デジタル、非デジタルを問わず）

＜連携の体制＞

「各アーカイブ機関」「つなぎ役」「国」がそれぞれの役割を果たす

### ジャパンサーチにおけるメタデータの役割と基本要件

各機関が保有するCHOを、検索・表示・提供するための情報。

- 参加機関の名称等を検索・表示・提供時に出力できるようにする（参加機関のインセンティブ向上のため）。
- デジタルコンテンツの有無、ライセンス情報の区分等で検索や絞り込みができるようにする。
- 従来の図書館資料で扱ってこなかった各種文化財についても、適切な形で検索・表示・提供できるように、必要な情報を記録できることが望ましい。例えば、材質・構造・技法等のモノに関する情報、その由来情報、地理的な情報、いわゆる著者以外の関係者に係る情報、などが考えられる。

### ジャパンサーチ利活用フォーマットのデータモデル策定における考え方

提供機関から収集したメタデータはそれぞれ異なる目的のために設計されたもので、その意図を十分生かしながら共通のメタデータにマッピングすることは難しい。

- (1)提供元からのソースデータは来歴情報明記の上そのまま保持
- (2)利用者の4つのタスク（発見、識別、選択、入手）の観点で共通情報を生成

5. 利活用フォーマット

共通情報

書誌レコードの機能要件（FRBR）の利用者のタスクの観点で、共通情報に求められるメタデータ項目（プロパティ）を検討した。

付加価値を持つ再利用のため、Linked Dataで提供する。

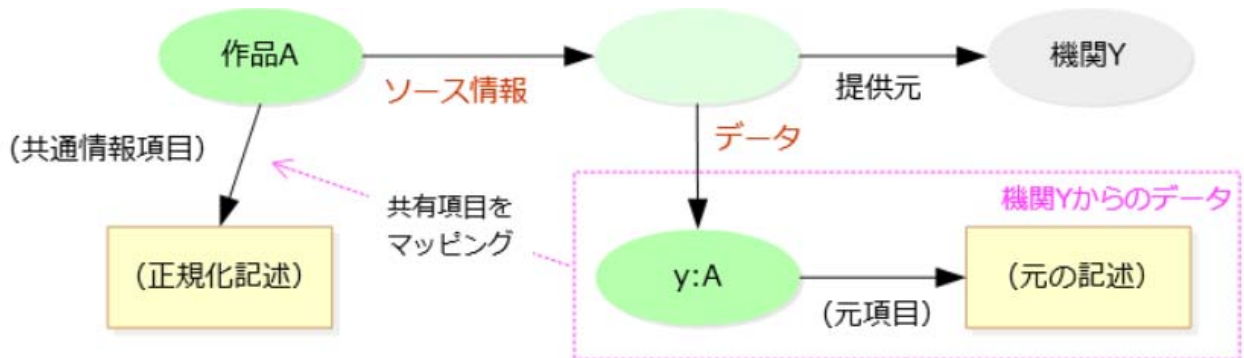
表：利用者タスクに応じたメタデータ項目の検討過程

タスク（機能）	必要な情報	設計方針
発見(検索)	キーワードなどで検索するための情報（領域の知識を前提とせずに）	「いつ」「どこで」「誰が」「何を」で検索可能にする
識別(表示)	示されている対象か何なのか、求めている（既知の）資料かどうか判断できる情報	時間、場所、人に関する情報に単純プロパティと構造化プロパティを併用する
選択(表示)	示されている対象が求めている（未知の）資料かどうかを判断でき、比較検討かを可能にするための情報	
入手(提供)	個別リソースあるいは詳細メタデータにアクセスするための情報	アクセス情報の充実

5. 利活用フォーマット

共通情報とソースデータ、「ソース情報分離モデル」

- ソースデータに含まれる情報の一部を、共通情報にマッピング（正規化記述）
- 共通情報の項目の1つとして「ソース情報」を持ち、ソースデータの内容をそのまま保持
- ソースデータの形式はJSONに変換し、単純検索や全文検索などで利用



## Europeana Data Model (EDM)

	要件	実現した方法
1	提供された文化遺産機関の資料（以下「オブジェクト」）とそのデジタル表現の分離	クラス [edm: ProvidedCHO]と[edm: WebResource]
2	オブジェクトとそのレコードの分離	クラス [edm: ProvidedCHO]と 記述的プロパティ群
3	同一のオブジェクトに対する複数のレコードの許容	クラス [ore: Aggregation] と[ore: Proxy]
4	他のオブジェクトから成るオブジェクトのサポート	階層関係や順序を表現するプロパティ群
5	抽象レベルの異なる記述との互換性と維持	オブジェクトとイベントに焦点を当てたモデル
6	詳細化も可能な標準的メタデータ形式	既存の語彙の利用と既存の標準とのマッピング
7	統制語彙由来の概念など文脈的リソースのサポート	セマンティックウェブ対応とツール

参考：

Europeana Data Model Premier.

[https://pro.europeana.eu/files/Europeana\\_Professional/Share\\_your\\_data/Technical\\_requirements/EDM\\_Documentation/EDM\\_Primer\\_130714.pdf](https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf)

福山樹里. Europeanaのメタデータ：デジタルアーカイブの連携の基盤. 情報の科学と技術. Vol.7. 2017. no.2

71

## Europeana Data Model (EDM)の鍵となる概念

**アグリゲーション (ore: Aggregation)**

- 単一のリソースとして扱えるように1つに集められた、関連リソースの集合体
- EDMにおいては、一つのデータプロバイダから提供された、一つのオブジェクトに関連するリソースを集合として扱うための実体。

**プロキシ (ore: Proxy)**

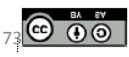
- ある特定の集合体 (Aggregation) の中にある特定のリソースを表す。
- EDMにおいては、オブジェクトの代理となつて、特定のデータプロバイダから見た特定のオブジェクトについての記述の対象になる。
- 同一のオブジェクトに対して複数のプロバイダのメタデータレコードを許容するため、また元データを維持しつつ、提供されたメタデータに文脈を付加したEuropeana独自のデータを追加的に持つために必要な実体。

3. 海外の統合ポータル

Europeana内で変換され、文脈情報を付加されたメタデータ



Fig. 16- Europeana aggregation—simple case with only one provider for the object



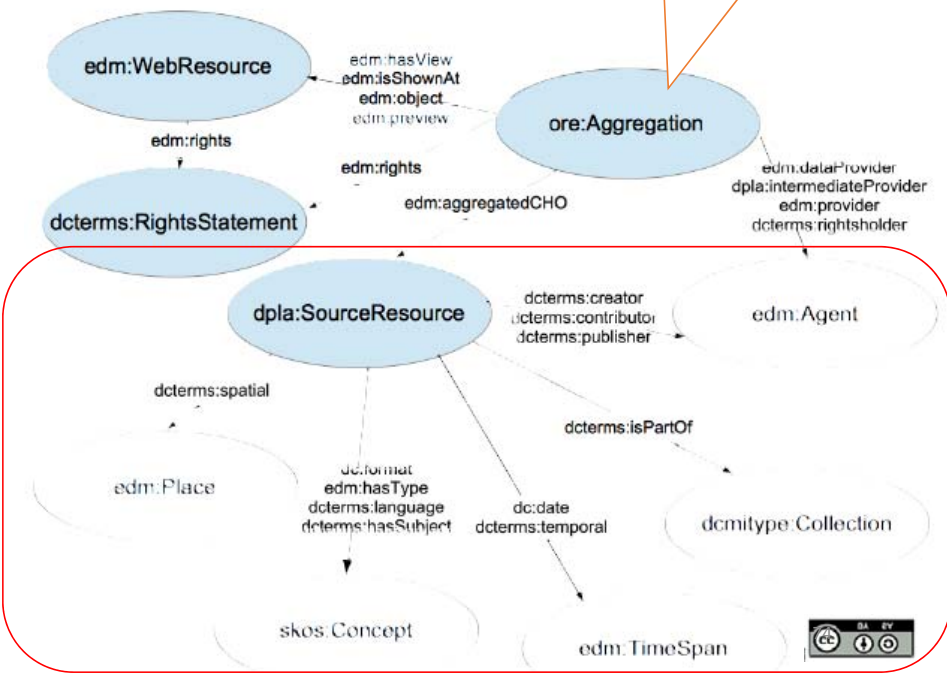
3. 海外の統合ポータル

DPLA Metadata Application Profile (MAP)

- EDMをベースに、モデルを簡素化
- 2012年にver.1 公開。最新版はver.5 (2017年12月公開)

dpla:originalRecord  
に提供されたままのレコードをそのまま格納

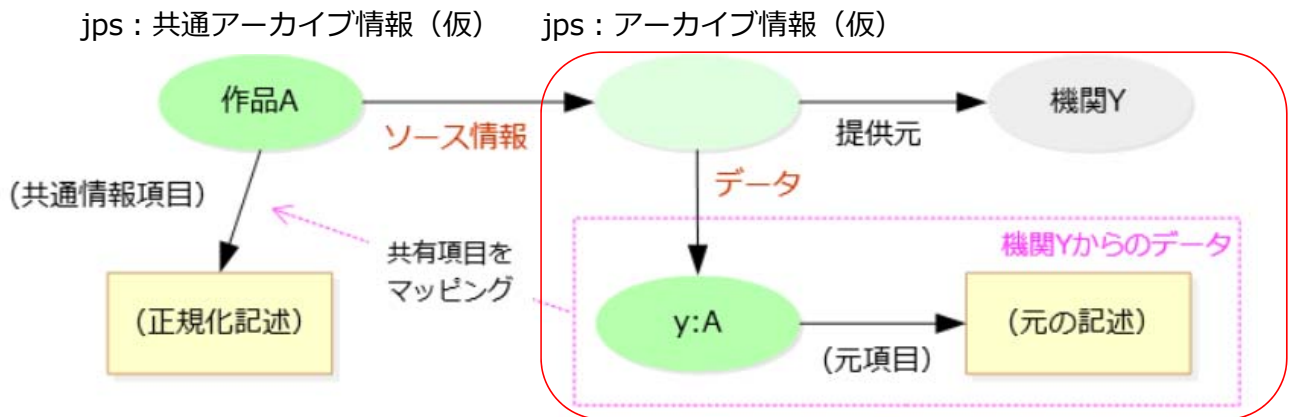
DPLA内で生成された正規化記述は、  
dpla:SourceResourceでまとめられる



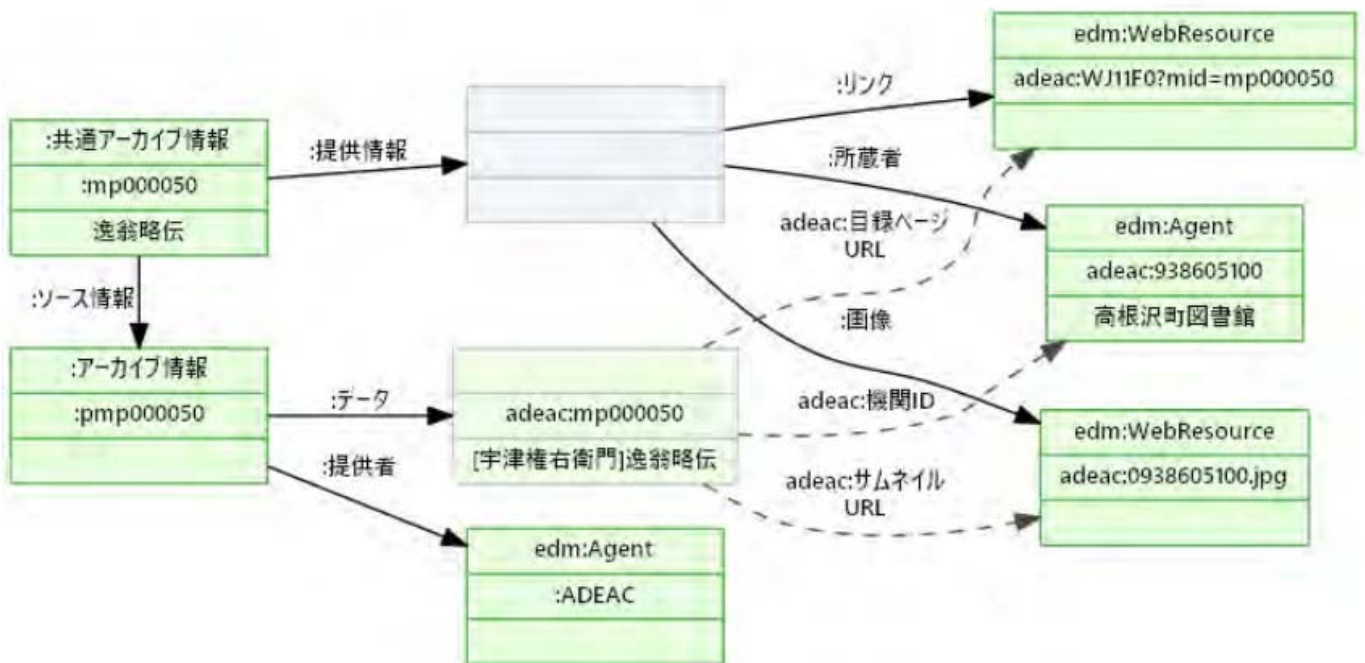
5. 利活用フォーマット

ジャパンサーチの「ソース情報分離モデル」

ore: Aggregationというクラスは設定せず、EDMでいうore: Proxy、DPLA MAPでいうdpla:SourceResourceに対応するクラスとして「共通アーカイブ情報（仮）」を設ける。  
 ソースデータやソースデータのプロバイダに関する情報は「アーカイブ情報（仮）」としてまとめる。



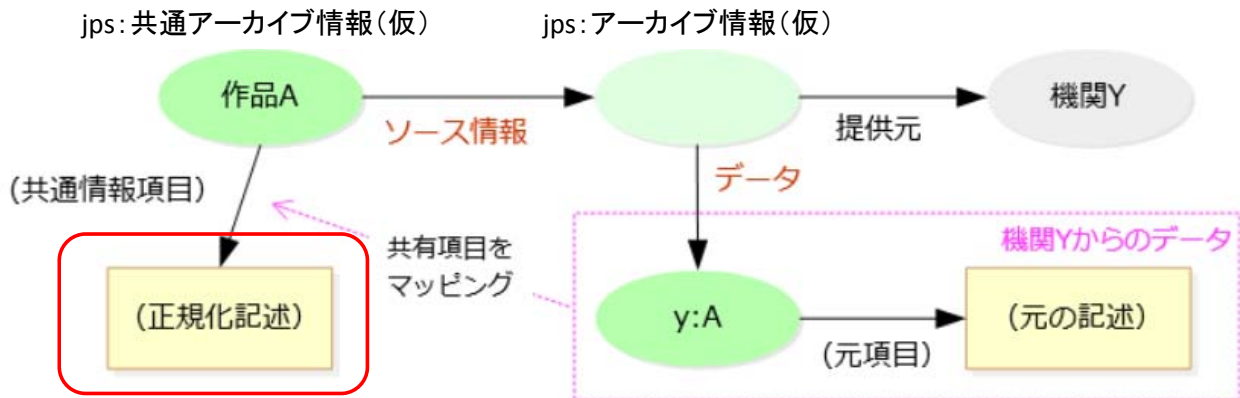
5. 利活用フォーマット



5. 利活用フォーマット

ジャパンサーチの「ソース情報分離モデル」 (再掲)

ore: Aggregationというクラスは設定せず、EDMでいうore: Proxy、DPLA MAPでいうdpla:SourceResourceに対応するクラスとして「共通アーカイブ情報(仮)」を設ける。  
 ソースデータやソースデータのプロバイダに関する情報は「アーカイブ情報(仮)」としてまとめる。



平成29年度の「利活用フォーマットの語彙」の検討範囲

5. 利活用フォーマット

共通情報の項目 (タスクに応じ)

語彙名称	語彙	内容	発見	識別	選択	入手
発見タスクに重要な基本項目-「何」「誰」「いつ」「どこ」についてschema.orgの語彙で発見可能にした						
タイプ	rdf:type	コンテンツの基本区分(書籍、文化財など)と共通アーカイブ情報であることを表す型	✓		✓	
ラベル	rdfs:label	一覧などに表示しコンテンツを識別するための名前。提供されない場合はIDなどから生成する。		✓✓	✓	
名称	schema:name	タイトル、別名、読みなど検索対象とする名前、読み、英語名称は言語タグで区別する。	✓✓	✓	✓	
識別子	schema:identifier	コンテンツを特定するための識別子 (ISBNなど、記録された媒体種別レベルのID、個別アイテムのIDではない)。	✓	✓		✓
寄与者	schema:contributor	コンテンツに寄与した人や組織。出演者等を含む。寄与の意味の違いはココでは区別せず、構造化プロパティで判断する。schema:creator, schema:publisherとして記述した値は、ここには重複して記述しない。	✓	✓	✓	
時間	schema:temporal	コンテンツの制作、出版、主題などに関する時間。時間の意味の違いはここでは区別せず、構造化プロパティで判断する。	✓	✓	✓	
場所	schema:spatial	コンテンツの制作、出版、主題などに関する場所。場所の意味の違いはここでは区別せず、構造化プロパティで判断する。	✓	✓	✓	

## 5. 利活用フォーマット

語彙名称	語彙	内容	発見	識別	選択	入手
発見タスクに重要な基本項目-「何」「誰」「いつ」「どこ」についてschema.orgの語彙で発見可能にした(続き)						
作者	schema:creator	作品制作の中心となった人や組織。寄与者[関係-制作]のショートカットとして位置付ける。	✓✓	✓✓	✓	
発行者	schema:publisher	作品の頒布の中心となった人や組織。出版者、制作会社、配給など。寄与者[関係-出版]相当のショートカットとして位置付ける。	✓	✓	✓	
作成日	schema:dateCreated	制作時の時間。時間[関係-制作]のリテラル値。	✓	✓	✓	
出版日	schema:datePublished	出版・発行時の時間。時間[関係-出版]相当のリテラル値。	✓	✓	✓	
識別、選択タスクに重要な基本項目-「何」に関連したより詳しい情報を構造化せずに表示する。schema.orgの語彙を使用。						
主題・区分	schema:about	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)。検索結果の絞り込みに利用できる用語。	✓	✓	✓	
言語	schema:inLanguage	コンテンツの記述言語(id.loc.govによるISO639-2のURIを用いる。)	✓	✓✓	✓✓	
画像	schema:image	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	schema:description	コンテンツの物理的特徴・素材等の記述や説明文など個別項目に収録できない情報	✓	✓	✓✓	
上位資料	schema:isPartOf	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)	✓	✓	✓	

## 5. 利活用フォーマット

語彙名称	語彙	内容	発見	識別	選択	入手
元データの情報も生かした識別や選択をするための構造化プロパティ-「誰」「いつ」「どこ」について独自語彙で充実させる。						
寄与関係	v:contribution	制作に関与した人や組織の関係。構造化値。	✓	✓	✓	
寄与タイプ	v:relationType	寄与のタイプ。領域を超えて一般化できる制作、編集、翻訳など	-	-	-	-
寄与者	schema:agent	寄与した人物・団体	-	-	-	-
役割	schema:description	領域固有の役割名、キャストの配役など補足記述	-	-	-	-
リンク	schema:url	関連リンク	-	-	-	-
時間関係	v:temporal	作成、公開、発見、主題など資料に関する時間	✓	✓✓	✓✓	
時間タイプ	v:relationType	時間関係のタイプ。「制作」、「主題」など。	-	-	-	-
時間	v:temporalValue	時間関係における時間(範囲)を示す(西暦年単位で正規化したURI)	-	-	-	-
時代	v:era	時間関係における時代を示す(時間オントロジーで定義するURI)	-	-	-	-
注記	schema:description	月日ほか補足情報	-	-	-	-

## 5. 利活用フォーマット

語彙名称	語彙	内容	発見	識別	選択	入手
元データの情報も生かした識別や選択をするための構造化プロパティ-「誰」「いつ」「どこ」について独自語彙で充実させる。						
場所関係		作成、公開、主題など資料に関係する場所・地域		✓	✓✓	
場所タイプ	v:relationType	場所関係のタイプ。制作、主題など (URI)	-	-	-	-
場所	v:spatialValue	場所関係における場所を示す (URI)	-	-	-	-
注記	schema:description	より詳細な地名・住所など補足記述 (文字列)	-	-	-	-
緯度経度	schema:geo	緯度経度などの位置情報 (schema:GeoCoordinates)	-	-	-	-
上位資料関係		上位コンテンツとの関係 (掲載誌のページ) などを記述する		✓	✓	
階層タイプ	v:relationType	上位コンテンツとの関係のタイプ。掲載など (URI)	-	-	-	-
上位資料	v:source	上位コンテンツ関係における上位コンテンツ本体を示す (URI)	-	-	-	-
部分	v:selector	上位コンテンツの特定部分を示す (文字列)	-	-	-	-
注記	schema:description	補足記述 (文字列)	-	-	-	-

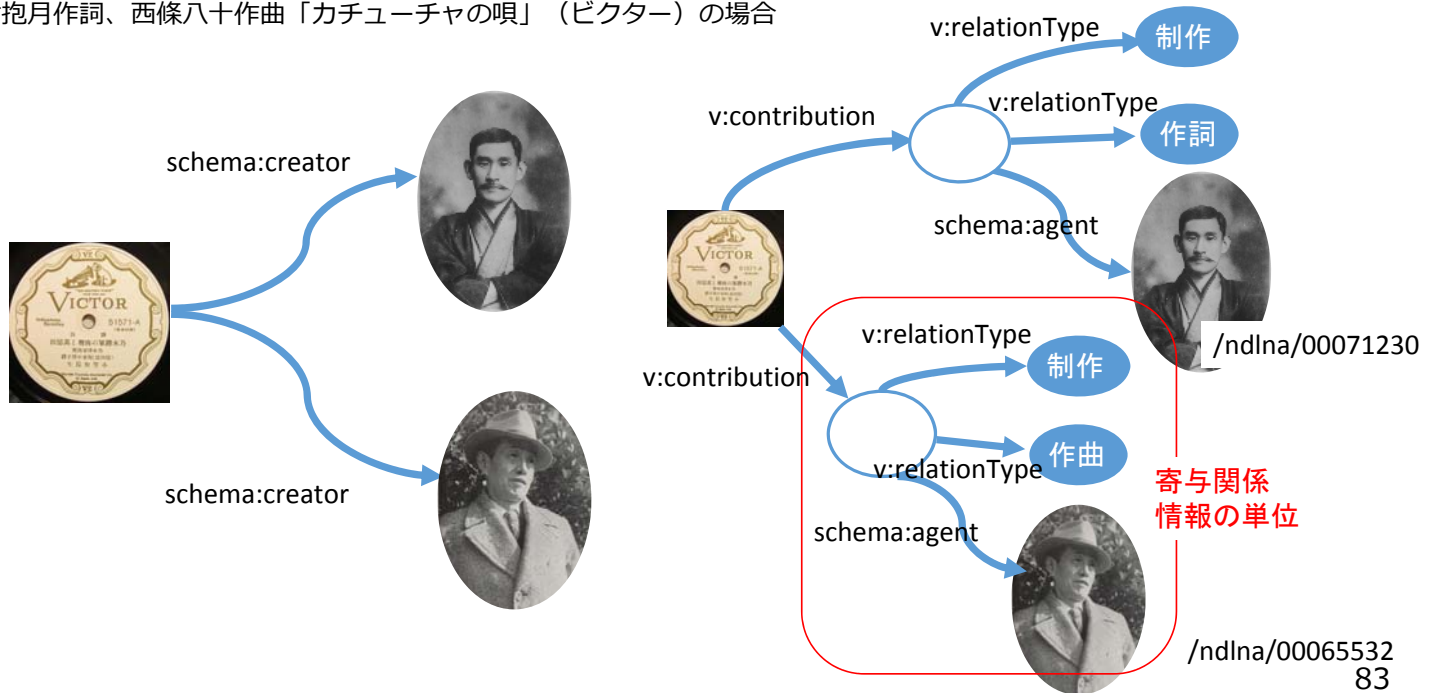
81

語彙名称	語彙	内容	発見	識別	選択	入手
コンテンツにアクセスするための情報を構造化プロパティ-「提供情報」と「ソース情報」に分けて提供する						
提供情報	v:accessInfo	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	schema:provider	コンテンツ (に関する情報) の提供者 (URI)		✓	✓	✓✓
保管者	v:contentHolder	コンテンツ (に関する情報) の保管者 (提供者と別の場合) (URI)		✓	✓	✓✓
リンク	schema:url	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 (URL)				✓✓
オブジェクト	v:digitalObject	コンテンツのデジタル画像や音声・動画。ライセンス情報を伴う (URI)			✓	✓✓
権利情報	v:contentRights	コンテンツ自身の利用に関するライセンス及び権利に関する情報			✓	✓
個別識別子	v:contentId	提供者・所有者が管理するアイテムとしての識別子 (請求記号等)	✓	✓		✓✓
注記	schema:description	補足記述			✓	✓
ソース情報	v:sourceInfo	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	schema:provider	ソースデータの提供者 (つなぎ役) (URI)				✓
データ	v:sourceData	ジャパンサーチが保持・提供するソースデータ (URI)				✓
リンク	schema:url	つなぎ役におけるソースデータの掲載ページ (URL)				✓
参照	rdfs:seeAlso	ソースデータがRDFである場合 (URI)				✓
更新日	schema:dateModified	収集元データの更新日又はつなぎ役による収集日 (日付型)	✓✓	✓✓	✓	✓
注記	Schema:description	補足記述			82 ✓	✓

## 5. 利活用フォーマット

### 「単純プロパティ」と「構造化プロパティ」

(例) 島村抱月作詞、西條八十作曲「カチューチャの唄」(ビクター)の場合



## 5. 利活用フォーマット

### 時間関係情報

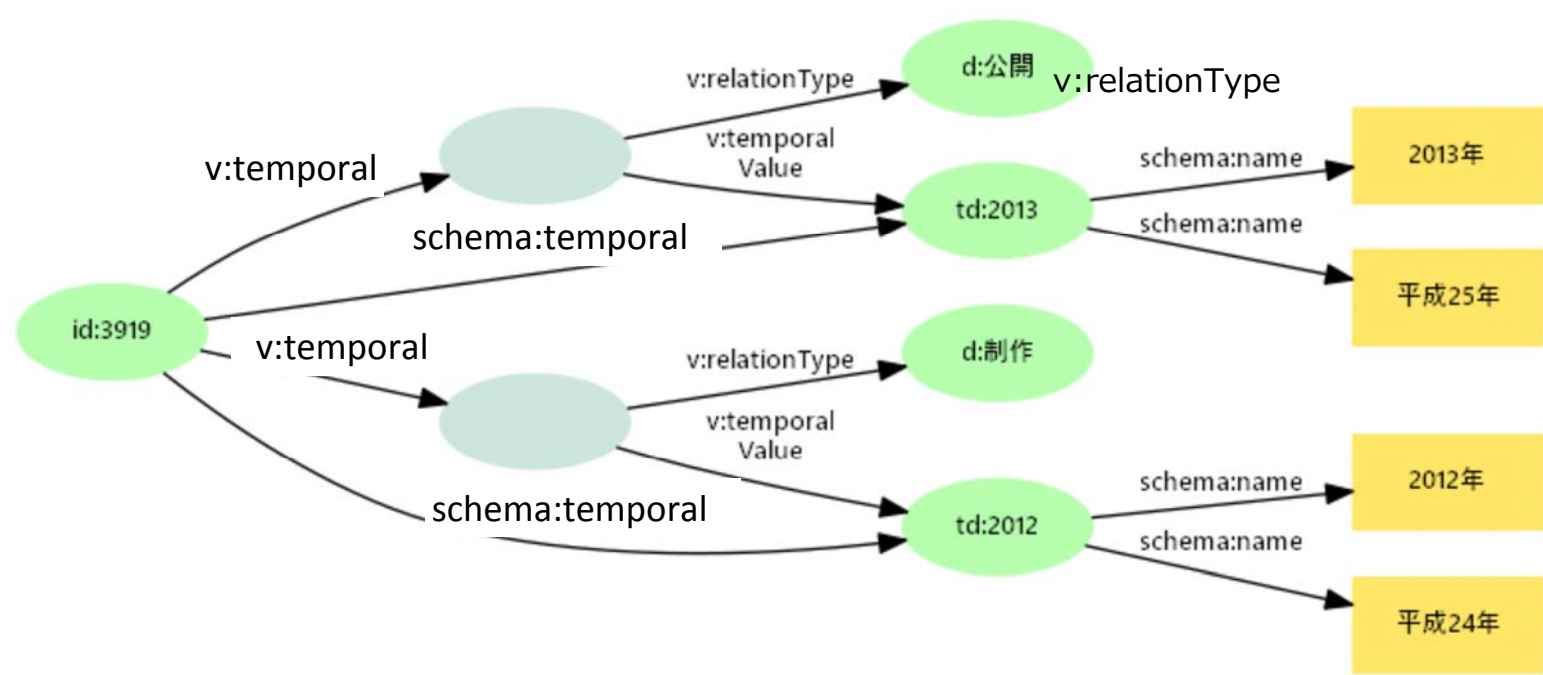
#### 「いつ」に関する情報を集約する

- 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) - `schema:temporal`
- 構造化プロパティで関係の違いを示す(識別、選択) - `v:temporal`で構造化

#### 正規化について

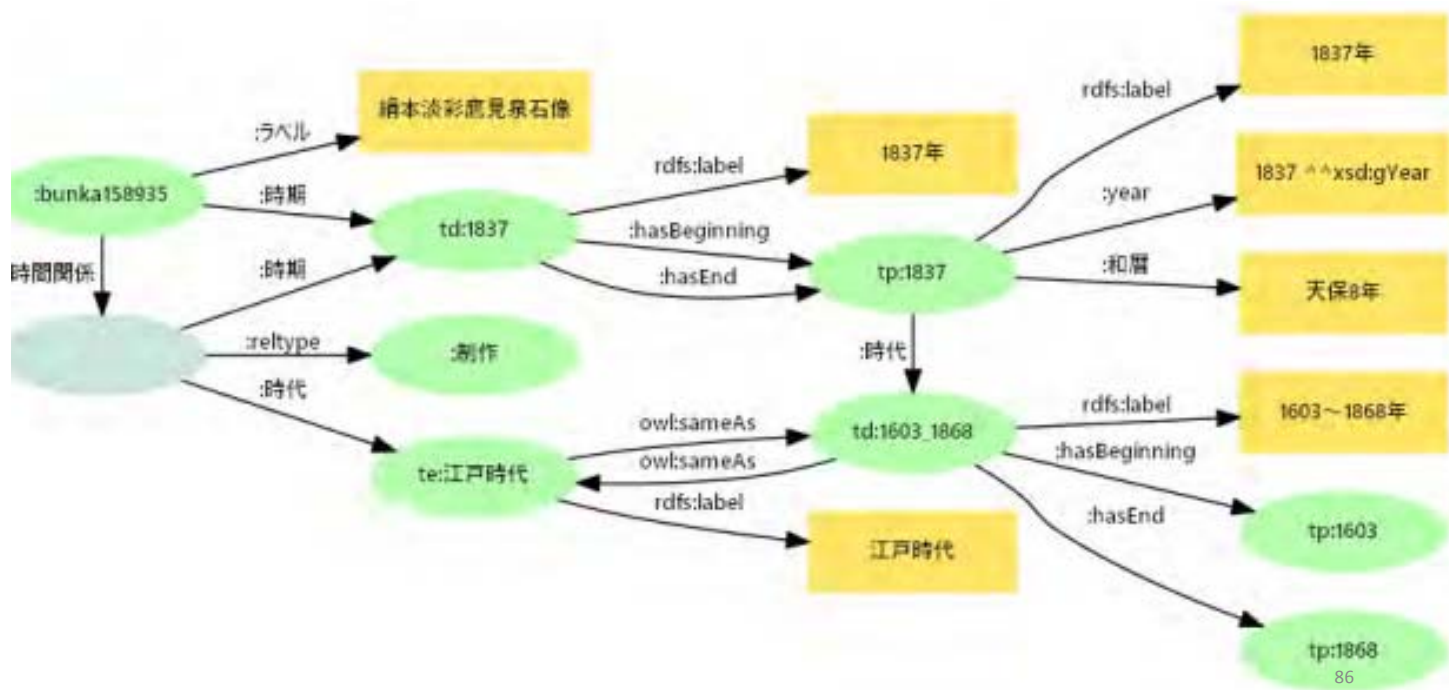
- 値は年を単位に正規化し、元記述が月日まで含むなどの場合は構造化情報に保持する
- 時間を文字列ではなくURIで表し、西暦と和暦、時代区分の対応などを「時間オントロジー」としてまとめる
- 時間範囲も`td:1868_1913`のようなURIを介して開始年、終了年を構造化することで、単一年の値と同様に処理できる(DPLAでも時間範囲は構造化)

5. 利活用フォーマット



5. 利活用フォーマット

時間オントロジーの考え方

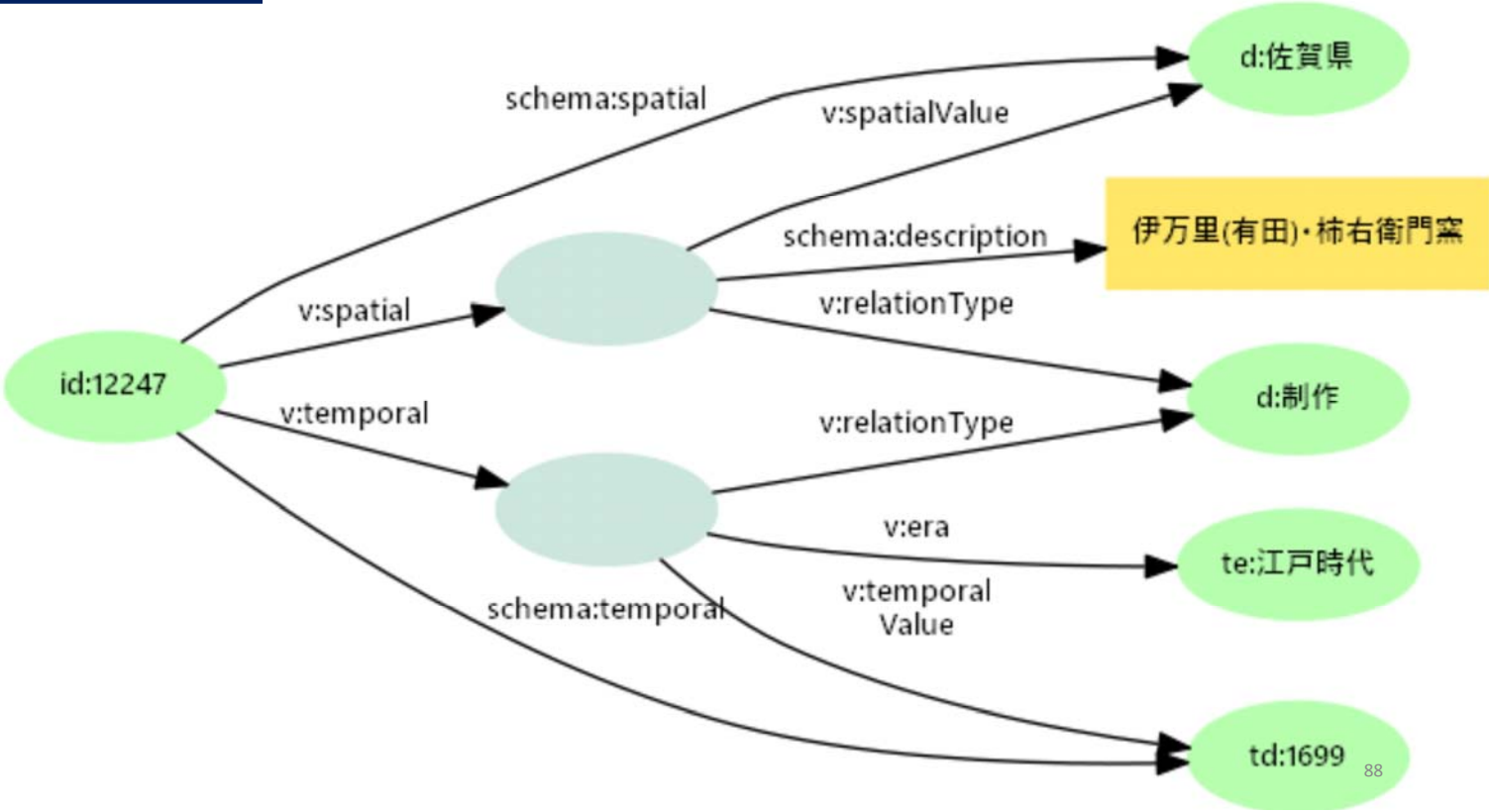


場所関係情報

「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） - schema:spatial
- 関係の違いは構造化プロパティで示す（識別） -v:spatialで構造化
- 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

※ただし、「場所」の正規化は時間以上に課題が多い



## ライセンス情報の記述（検討中）

**コンテンツ自身のライセンス -v:contentRights**

- 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） -schema:licence**

- 提供元毎などに標準ライセンスを決めて一括記述
- 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表示する）

**メタデータに関するライセンス（案） -原則CC0のため項目なし？**

- 原則として共通情報はすべてCC0
- ソースデータもジャパンサーチで提供できるものはCC0？

89

## 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

90

共通情報整備の課題

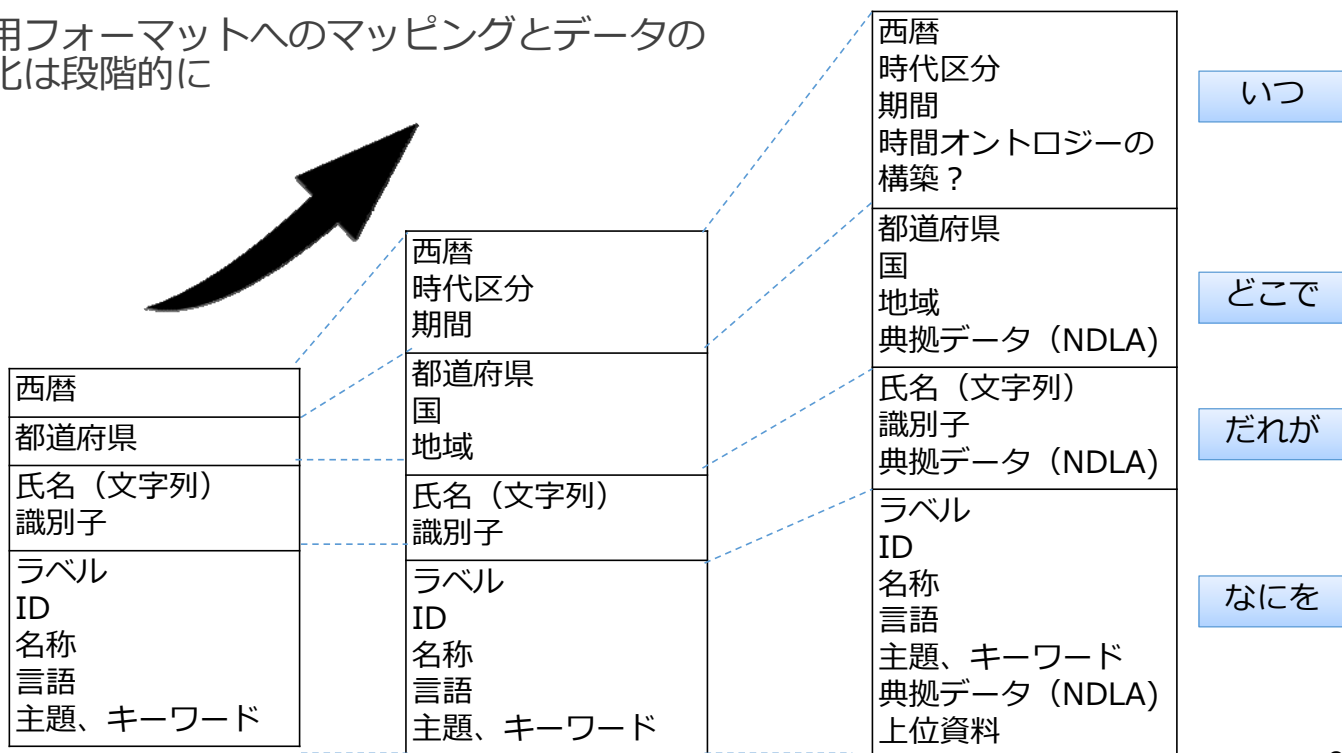
利活用のための提供データの調整と補完

- 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

運用のためのドキュメント整備

- 既存語彙との関係整理、利活用者に向けた説明資料などを整備

利活用フォーマットへのマッピングとデータの正規化は段階的に



## 6. まとめ

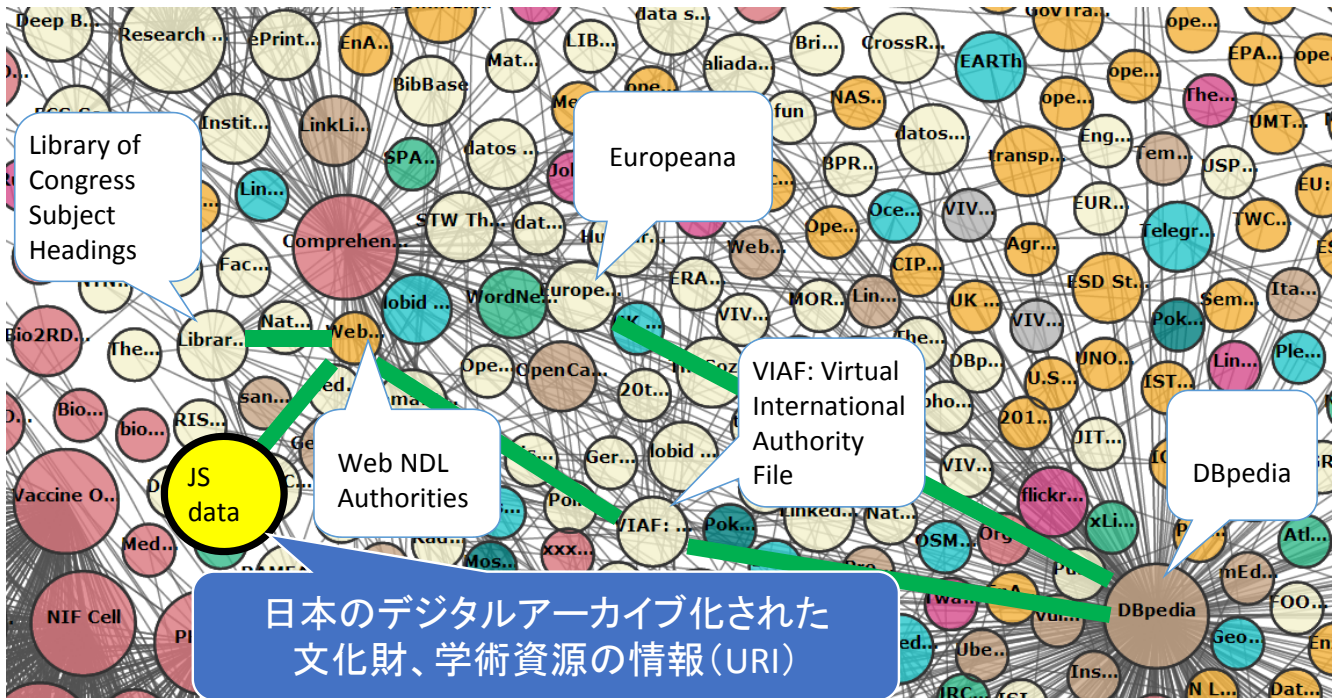
## どこから来て、どこに行くのか（私見）

- ◆「電子図書館」構想／計画における統合検索サービス  
ウェブ上の情報資源へのナビゲーション  
デジタル一次情報資源へのポータルから書誌情報基盤へ
- ◆ポスト電子図書館「知識インフラ」「利活用基盤」としてのジャパンサーチ
- ◆図書館的要素のリ・デザイン  
典拠コントロール、アクセスポイントから文脈の付与へ  
情報探索支援における透明性と信頼性の保証とリテラシーへのまなざし



「図書館」のリ・デザイン？（図書館は「知識インフラ」になれるか／を作れるか）

## 使えるデータを作る、日本情報のハブへ



## 参考URI (全体を通して)

- [https://www.kantei.go.jp/jp/singi/titeki2/digitalarchive\\_suisiniinkai/index.html](https://www.kantei.go.jp/jp/singi/titeki2/digitalarchive_suisiniinkai/index.html)
- [https://www.kantei.go.jp/jp/singi/titeki2/digitalarchive\\_kyougikai/houkokusho.pdf](https://www.kantei.go.jp/jp/singi/titeki2/digitalarchive_kyougikai/houkokusho.pdf)
- [http://www.kantei.go.jp/jp/singi/titeki2/digitalarchive\\_kyougikai/guideline\\_gaiyou.pdf](http://www.kantei.go.jp/jp/singi/titeki2/digitalarchive_kyougikai/guideline_gaiyou.pdf)
- [https://www.kantei.go.jp/jp/singi/titeki2/digitalarchive\\_suisiniinkai/jitumusya/2017/torimatome.pdf](https://www.kantei.go.jp/jp/singi/titeki2/digitalarchive_suisiniinkai/jitumusya/2017/torimatome.pdf)
- <http://www.ndl.go.jp/jp/dlib/project/history.html>
- <https://pro.europeana.eu/page/issue-3-open-data-and-data-re-use>
- <https://pro.dp.la/hubs/metadata-application-profile>
- <https://pro.europeana.eu/page/edm-documentation#model-definition>
- <https://www.oclc.org/developer/develop/linked-data/worldcat-vocabulary.en.html>
- <http://www.ndl.go.jp/jp/dlib/standards/jpsformat.html>