

# 文化資源特別講義b 文化資源特論III

大阪市立大学・大阪市立大学大学院講義資料

福田一史

fukudakz@gmail.com

<https://scrapbox.io/fukudakz/文化資源学特別講義b・文化資源特論III>



図. 講義ウェブサイトリンク (QRコード)  
※LMSの資料のページにもリンクがあります

# 前回講義のコメントより

- 「文化を計量的に分析することの面白さを感じました」
- 「Voyant toolsは私が使いこなせていないのとまだシステムを確立できていない(?) のとで、「た」とか「して」とかしか出てこなかったのも、あまり意味を感じませんでした。」
- 「紙媒体の文章はデジタル化する必要があるのも、分量が多くなると分析が難しい」
- 「テキストマイニングソフトで谷崎潤一郎の「刺青」を解析してみたのですが、主人公の「清吉」よりも「女」「娘」の方が高い頻度で出てきていたり、「刺青」という語が物語の中盤にはほとんど出てこなかったりと、普段読んでいる時には分からない(気が付かない)ことが可視化されるのが面白かったです。」
- 「助詞を省く機能を付けてくれたらだいぶ助かります。」

# 前回講義のコメントより

- 「TEIについて...とても細かく一つ一つのテキストをタグづけしていると感じたのだが、これは一つ一つ手動で行っているのだろうか。だとしたら、非常に手間のかかる大変なプロジェクトだと感じた。」
- 「Vlog動画の内容分析を行っています。tiktokやYouTubeやBilibiliなど専門的な動画サイト向けのビッグデータソフトウェア(feishu・xinbang)を使って、いいね数、弹幕数、コメント数、または視聴者のメカニズム、弹幕ワールドクラウンなどを統計して分析しています。」
- 「デジタル技術を用いた文化保存、研究は自然災害の多い日本では進展すべき」

# コンテンツ

1. デジタル技術を用いた文化資源の分析（つづき）
2. メディア芸術のアーカイブ構築に関する事例研究：  
メディア芸術データベース

# デジタル技術を用いた 文化資源の分析（つづき）

計量的に文化を記述し、分析する

# テキストマイニング

- テキスト（文章・文字データ・文献など）を対象とするデータマイニング。
- 文章を単語や文節で区切り（形態素解析）、それら単語や文節の出現頻度や共起関係などを計量的に分析する。自然言語処理技術の一種。
- KH Coder
  - テキストマイニングソフトの一つ。  
<http://khcoder.net/>



#	抽出語	品詞/活用	頻度
1	先生	名詞	595
2	K	タグ	411
3	奥さん	名詞	388
4	思う	動詞	296
5	父	名詞C	269
6	自分	名詞	264
7	見る	動詞	225
8	聞く	動詞	219
9	出る	動詞	185
10	人	名詞C	182
11	母	名詞C	170
12	お嬢さん	名詞	168
13	前	副詞可能	163
14	帰る	動詞	155
15	今	副詞可能	139
16	顔	名詞C	135
17	来る	動詞	131
18	考える	動詞	130
19	言葉	名詞	126
20	眼	名詞C	123

図. 夏目漱石「こころ」の抽出語リスト  
<http://khcoder.net/> スクリーンショット集より

# IIIF: International Image Interoperability Framework

- <https://iiif.io/>
- 「**画像へのアクセスを標準化し相互運用性を確保するための国際的なコミュニティ活動**」
  - IIIFを用いた高品質／高精細の画像公開と利用事例 | 人文学オープンデータ共同利用センター. <http://codh.rois.ac.jp/iiif/>
- もしくはそれらにより定義された**API**の仕様
  - [IIIF Image API](#)
  - [IIIF Presentation API](#)
  - など



図. IIIFロゴ



# API

- Application Programming Interface
- プログラムコンポーネント同士を接続するためのインターフェイス、およびその仕様
- バックエンドの一部の機能の公開や共有
- ワールド・ワイド・ウェブを介するAPIは「ウェブAPI」と呼ばれ、HTTP要求とJSONなどによる応答で構成される。

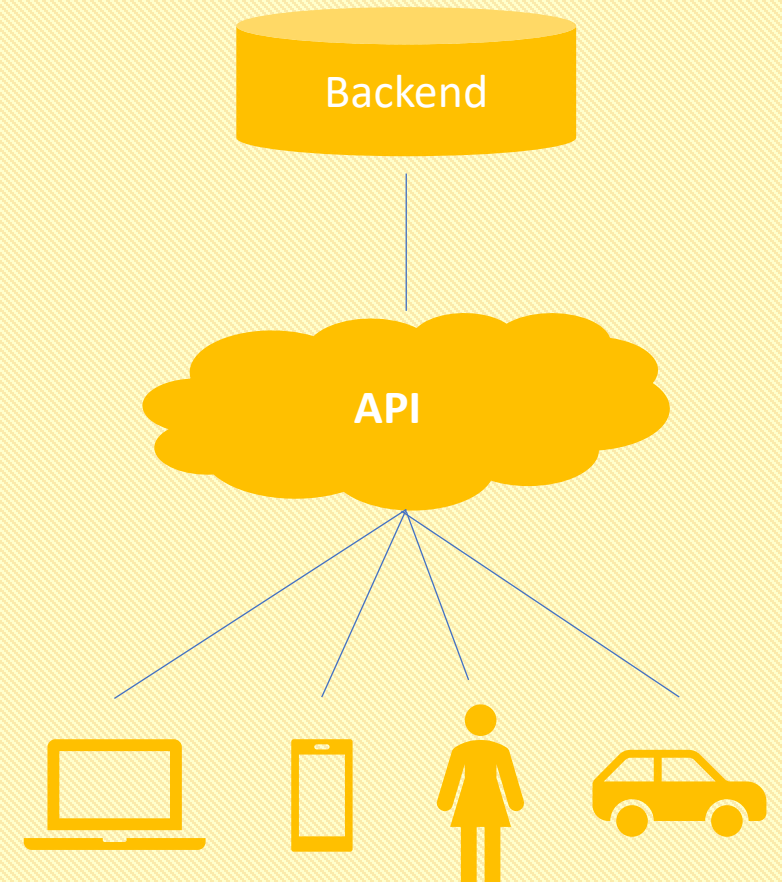


図. APIイメージ

# IIIFの普及と利便性

- IIIFの採用は海外の多くの期間で進んでおり、MLAコミュニティの標準になりつつある
- 標準化が進むことで、機関横断的に情報リソースを利用することができるようになる

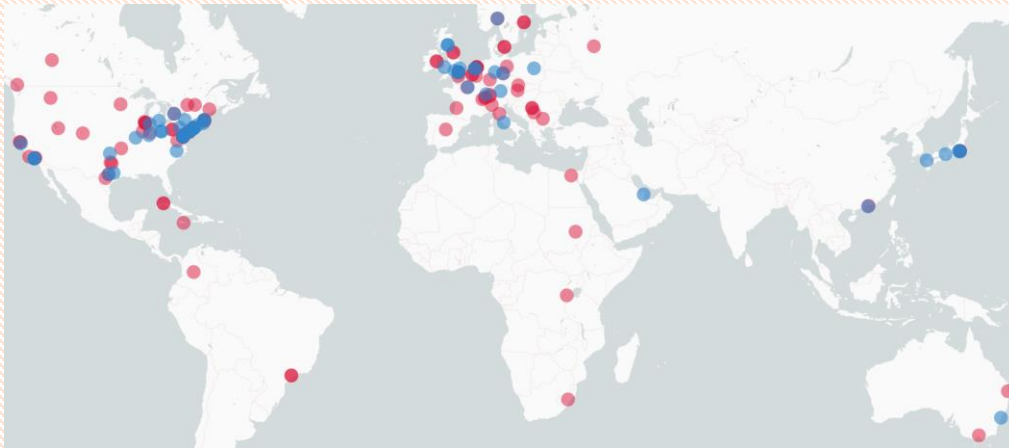


図. IIIFコンソーシアムの広がり  
<https://iiif.io/get-involved/> より

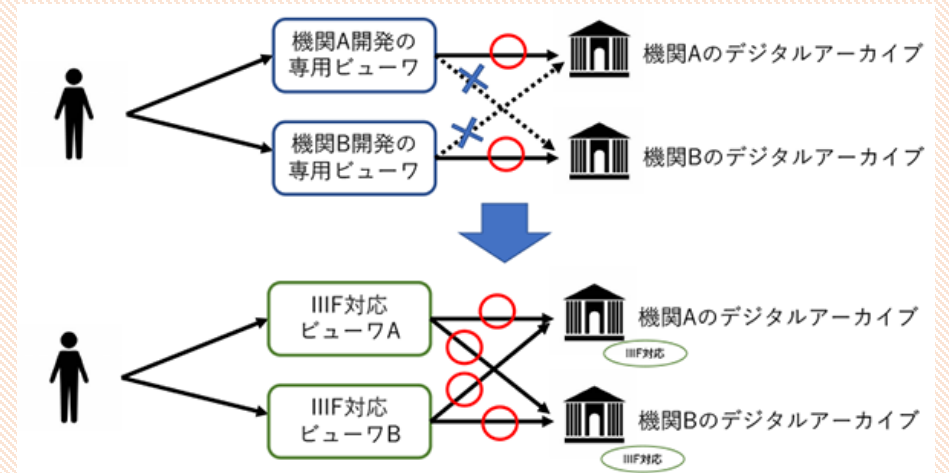
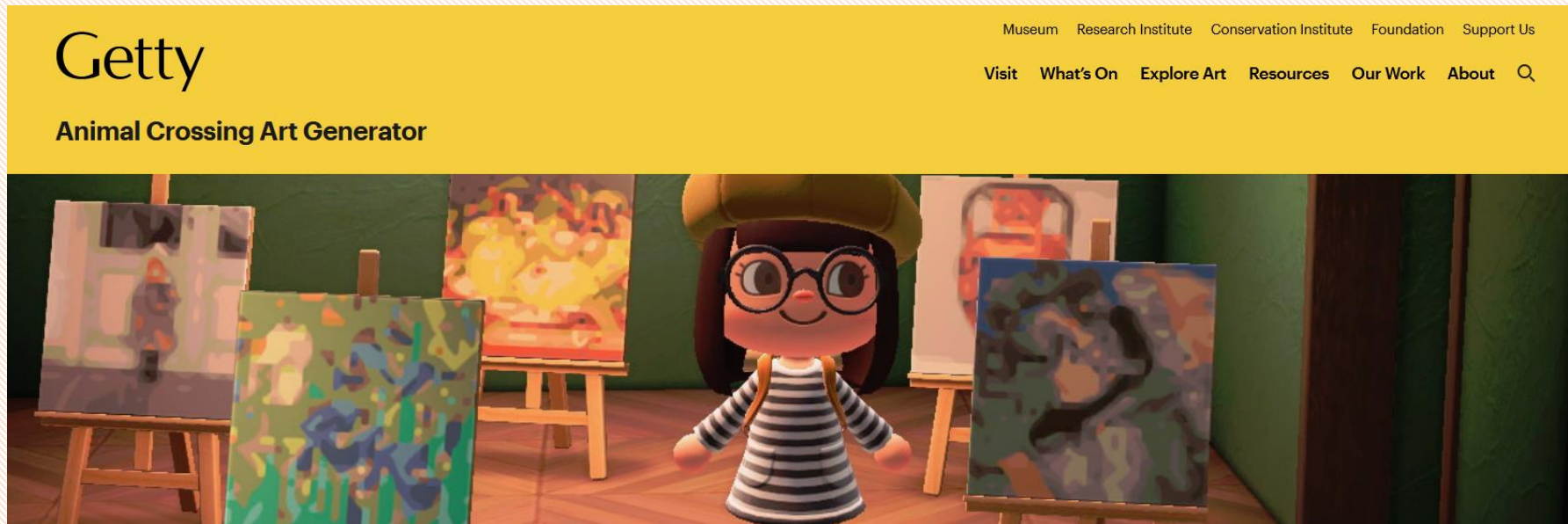


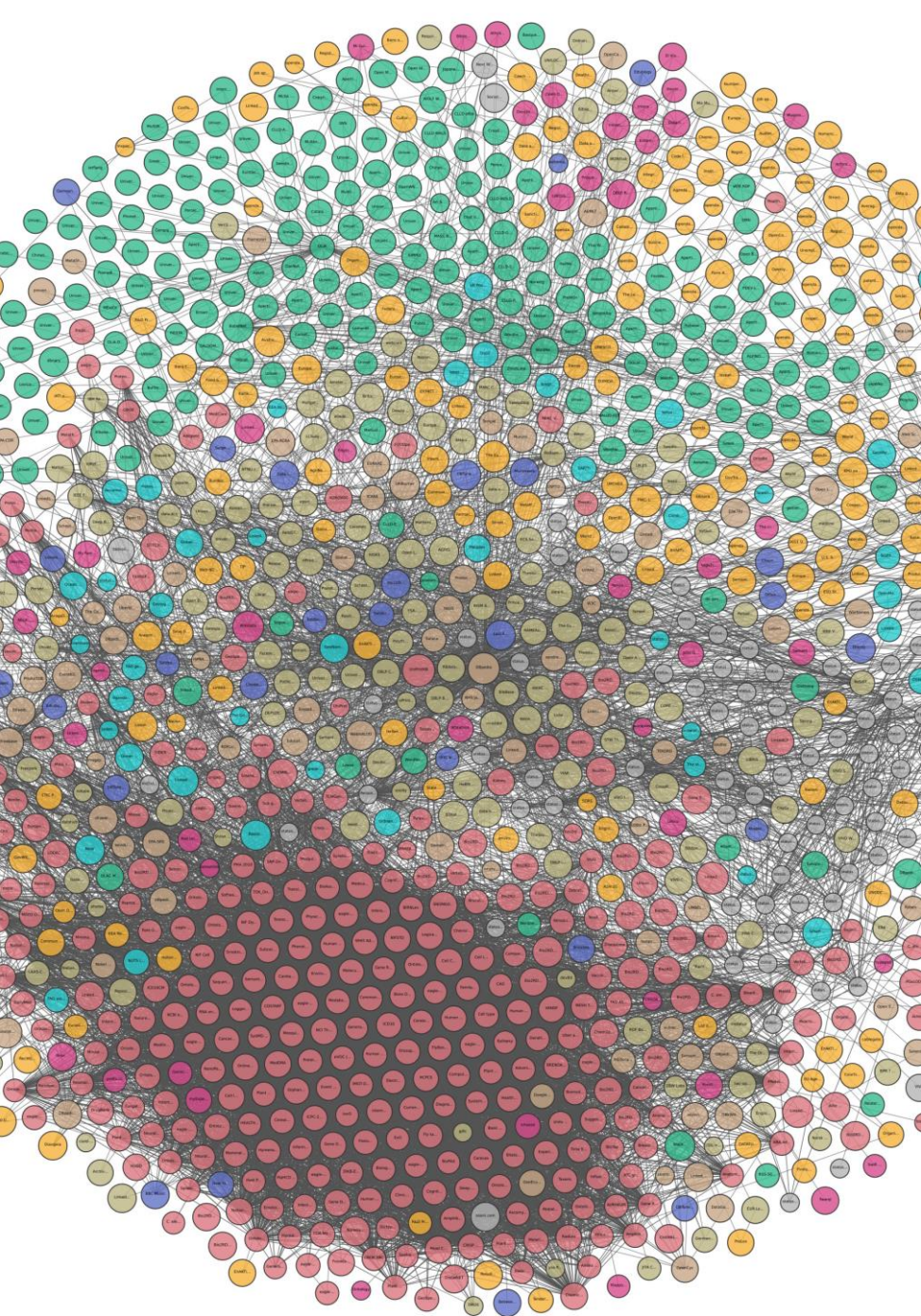
図. 従来型のビューワとIIIFビューワの違い  
<https://ura.sec.tsukuba.ac.jp/archives/24153> より

# 事例：どうぶつの森

- [Getty Animal Crossing Art Generator](#)
- 文化財活用センター『[「あつ森」で考える、ゲームの学術的な可能性](#)』







# Linked Data

- リンクト・データ
- ワールド・ワイド・ウェブを通じて機械可読性の高いデータを公開するための技術。
- 特にオープンデータとして公開するものを**Linked Open Data (LOD)**と呼ぶ。
- セマンティックウェブを構成するための主要技術の一つ。

# Linked Data

- 文書のウェブからデータのウェブへの転換
- Tim Berners-Leeの提案するLinked Dataの原則
  1. Use **URIs** as names for things
  2. Use **HTTP URIs** so that people can look up those names.
  3. When someone looks up a URI, **provide useful information**, using the **standards** (RDF\*, SPARQL)
  4. **Include links to other URIs**. so that they can discover more things.
    - Linked Data – Design Issues. <https://www.w3.org/DesignIssues/LinkedData.html>.
- より詳しくは以下論文など参照のこと
  - 大向一輝. 2013. “オープンデータ活用：1. オープンデータとLinked Open Data.” 情報処理 54 (12): 1204–10. <http://id.nii.ac.jp/1001/00095922/>.



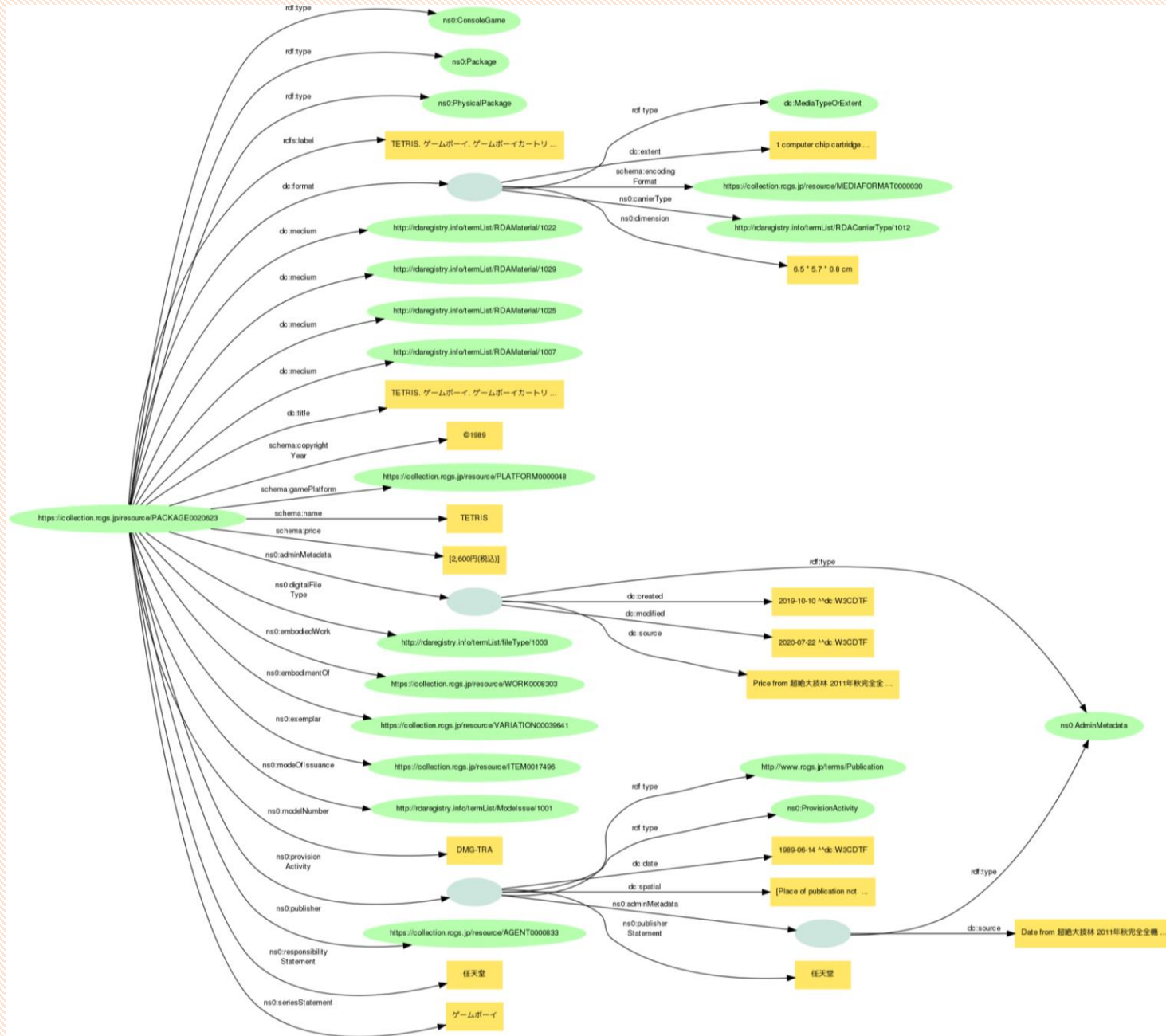


図. ゲームボーイ版テトリスのRDFによる書誌データのグラフ  
The Web of Kanzaki – RDFグラフの視覚化により生成

# 機械学習

- **KuroNetくずし字認識ビューア**

- 人文学オープンデータ共同利用センターが提供する、ディープラーニングを用いたくずし字のOCRサービス。
- <http://codh.rois.ac.jp/kuronet/>

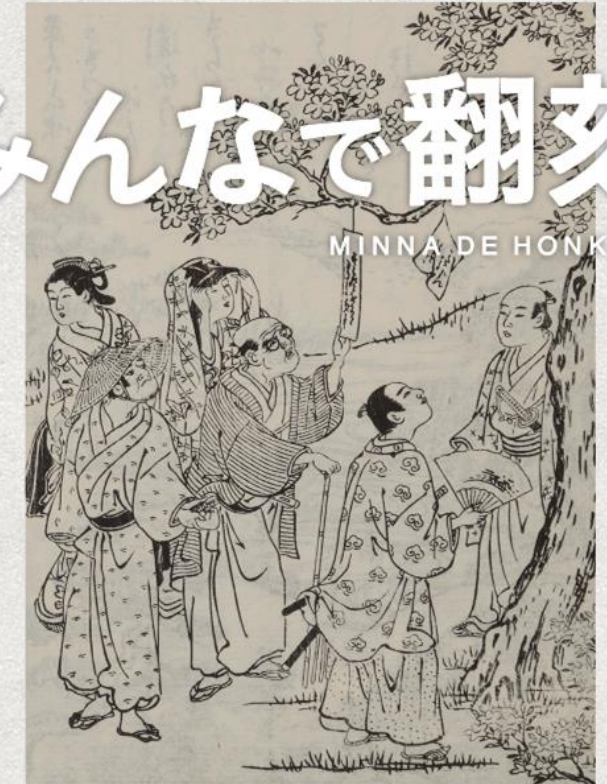
# 集合知・クラウドソーシング

- みんなで翻刻
- <https://honkoku.org/>
- 「すでに『みんなで翻刻』では、5000人の人々により600万文字もの史料が翻刻されています。」
  - <https://honkoku.org/>
- コミュニティ維持のために**ゲーミフィケーション**の方法論を採用。



## みんなで翻刻

MINNA DE HONKOKU



と、AIと協力して、  
書解読に参加しよう！

JOIN  
参加する

※現在新バージョンの試験運用中です。  
安定稼働するまで旧バージョンも併せてご利用ください。

いいね! ツイート

B!ブックマーク LINEで送る



# DHの方向性

- 既存の文化研究の**ブラックボックス化を危惧**し、デジタル技術を用いて**横串を指す** / **透明化**をすすめる。
- **オープン化**や**標準化**により、文化資源へのアクセシビリティを向上させ、持続的に提供可能とし、**知識基盤**を確立する。
  - 特にオープンな標準や、コミュニティで維持するオープンソースを指向する。**方法論の共有地**（Methodological Commons）。
- アクセシビリティが向上することにより、これまで想定していなかった**新しい方法論**や**知見**が登場。
  - 「デジタル・ヒューマニティーズ入門」
  - デジタル・ヒューマニティーズー人文学と情報学の接点が導く新たな知識の世界。  
<https://www.youtube.com/watch?v=aemYqznZM-8> (accessed 2021-04-15)

# メディア芸術のアーカイブ構築に 関する事例研究： メディア芸術データベース

マンガ・アニメ・ゲーム・メディアアートの総合目録

# そもそも、メディア芸術の資源化って必要？

- メディア芸術 = マンガ・アニメ・ゲーム・メディアアート
  - 現代的な視覚文化・視覚メディアと言い換えることもできる
  - 特にマンガ・アニメ・ゲームが特定の文化的まとまりとして認識されているといのは確か（メディアアートが微妙に仲間外れ？）
- これらは...
  - 教育の対象となり得るのか？
  - 調査や研究の対象となり得るのか？
  - 社会そのものや人々の考え方・認識に強い影響があるのか？
  - 誇らしいものなのか？ 恥ずかしいものなのか？
  - 資源化されてきていないからこそ、文化として認められていない？

# アーカイブの対象としてのメディア芸術の特徴

- 著作権がある (In Copyright) 作品が多い
  - コンテンツを複製して公開できない > 収集・アクセスの提供に焦点化
  - 継続的にビジネス的価値が見いだされており、企業によるアーカイブも一部において活発

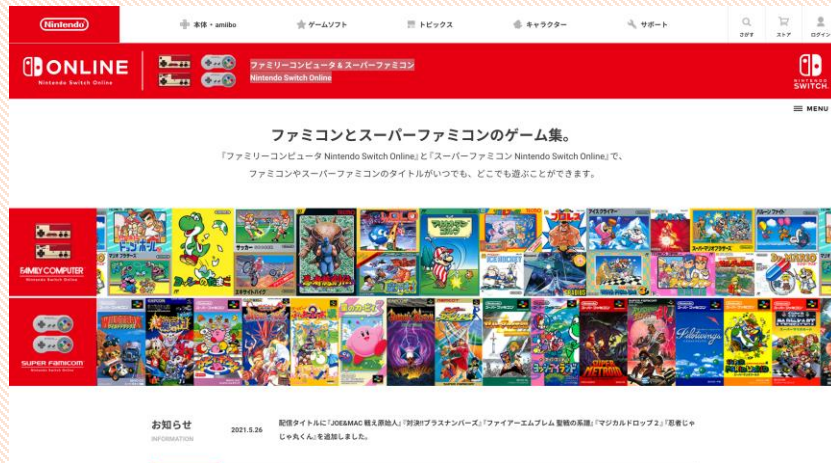


図. [ファミリーコンピュータ&スーパーファミコン Nintendo Switch Online](https://www.nintendo.com/online)



図. [BOOK☆WALKER](https://www.bookwalker.jp/)



図. [ピッコマ](https://www.kakao.co.jp/service/)

# アーカイブの対象としてのメディア芸術の特徴

- 複製され頒布される
  - 図書資料に近い性質
  - ただし専用の電子装置が必要とされる場合も多く、メディアの形式が多様
    - 保存や利用が課題に
- 放送やオンラインなど形のないものも近年増加傾向にある



図. さまざまなメディア芸術のメディア

	モンスターハンターストーリーズ2 ~破滅の翼~	2021年7月9日		¥ 6,990
	Sekiro™: Shadows Die Twice - GOTY Edition	2019年3月22日	-50%	¥ 8,360 ¥ 4,180
	天穂のサクナヒメ	2020年11月11日	-30%	¥ 5,478 ¥ 3,834
	NieR Replicant™ ver.1.22474487139...	2021年4月24日	-25%	¥ 8,580 ¥ 6,435
	ペルソナ4 ザ・ゴールデン	2020年6月14日	-35%	¥ 1,980 ¥ 1,287
	Hades	2020年9月17日	-30%	¥ 2,570 ¥ 1,799
	オクトパストラベラー	2019年6月8日	-50%	¥ 7,480 ¥ 3,740
	モンスターハンターストーリーズ2 ~破滅の翼...			¥ 8,065

図. オンライン配信プラットフォーム「[Steam](https://store.steampowered.com/)」

# アーカイブの対象としてのメディア芸術の特徴

- メディアミックスによりメディア横断的に展開する
  - マンガ原作のアニメ、ゲーム原作のアニメなど
- 日本発で、全世界的な影響力



図. メディアミックスに関する研究書



# メディア芸術データベース

- <https://mediaarts-db.bunka.go.jp/>
- マンガ・アニメ・ゲーム・メディアアートの作品・所蔵情報のデータベース（DB）。
- 2015年より試験版、2019年よりベータ版が公開される。
- マンガ所蔵館やゲーム所蔵館などの協力機関の作成する書誌データを収集する他、独自に作品情報などを作成・整理し公開。



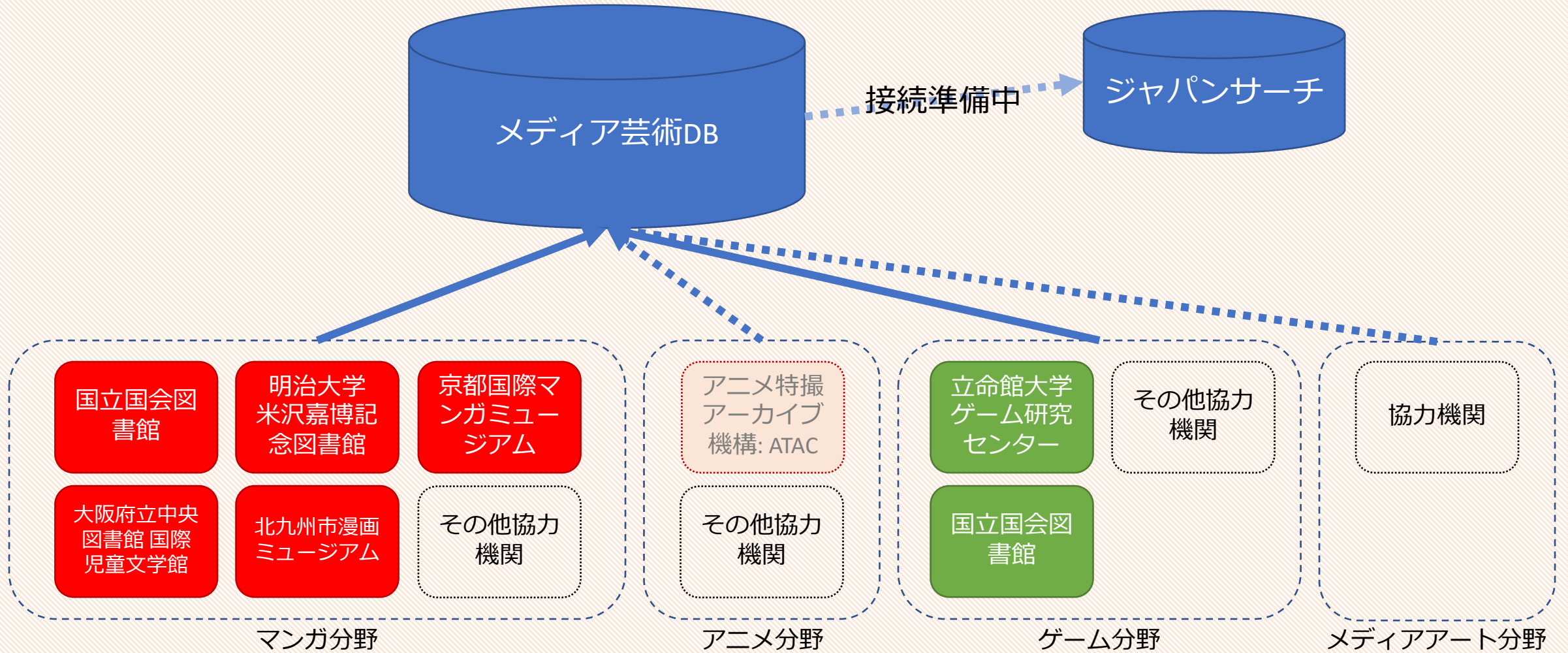
## メディア芸術データベース ベータ版

# 位置づけ

- 文化芸術振興基本法の第九条「メディア芸術の振興」に基づき展開される施策
  - e-Gov <https://elaws.e-gov.go.jp/document?lawid=413AC1000000148>
- メディア芸術の振興の施策（[文化庁ウェブサイト](#)より）
  1. 文化庁メディア芸術グローバル展開事業
  2. メディア芸術の人材育成
  3. メディア芸術連携基盤等整備推進事業



# メディア芸術のアーカイブ集約



# メディア芸術DB： 参考資料・データ

- パンフレット「よくわかる メディア芸術データベース ～つながるデータで構築する未来へのアーカイブ～（右図）」（[PDF](#)）
- LODデータセット（[Github](#)）
- メタデータスキーマ仕様書（[PDF](#)）
- メディア芸術データベースの説明動画
  - 「メディア芸術データベース活用コンテスト2021説明会」（[Youtube](#); 15:50ごろから）



図.パンフレット「よくわかるメディア芸術データベース」

メディア芸術データベース データベース	
メディア芸術データベース メタデータスキーマ Media Arts Database Metadata Schema	
Ver. 1.0 2021-03-22 	
<b>公開年月日</b>	公開年月日 ラベル 出版・修正した日付 定義 当該リソースが初めて公開された日付が記録される。 コメント 領域 日付 タイプ プロパティ URI <a href="http://schema.org/datePublished">http://schema.org/datePublished</a>
<b>作者</b>	作者 ラベル 作成した責任主体 定義 [責任主体] 領域 タイプ プロパティ URI <a href="http://purl.org/dc/terms/creator">http://purl.org/dc/terms/creator</a>
<b>作成名</b>	作者名 ラベル 作成した責任主体の名称 定義 役割付きで記録することが推奨される。 コメント 領域 リテラル タイプ プロパティ URI <a href="http://schema.org/creator">http://schema.org/creator</a>

図.メタデータスキーマ仕様書

# メディア芸術DBの現況（とりわけ課題面）

- メディア芸術分野の目録やアーカイブの専門家は少ない
  - 研究者もメディア研究の専門家で大部分が構成される
- マンガ分野以外のデータ登録機関が少ない
  - マンガは「図書」であるため既存の図書館所蔵管理（データ作成）システムの多くを援用可能
  - ゲーム分野の立命館大学ゲーム研究センターとアニメ分野のATAC（予定）以外はまだ明確なデータ登録機関がない
  - 1) データの入力系の整備（**データ登録簡易化**）と、2) 関連機関の**専門性の高度化**、の両方を充実させていく必要性がある