

デジタルアーカイブと その社会的活用

第2回

立命館大学 映像学部講義 (2021-04-19)

福田一史

<https://scrapbox.io/fukudakz/21デジタルアーカイブとその社会的活用>



コンテンツ

1. デジタルアーカイブとは
2. デジタルアーカイブをめぐる標準技術

デジタルアーカイブとは

デジタルアーカイブの定義とその意義、および日本の現況

デジタルアーカイブとは

- 「『デジタルアーカイブ』とは、様々な**デジタル情報資源**を**収集・保存・提供する仕組みの総体**」
 - 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.
(2017). 「我が国におけるデジタルアーカイブ推進の方向性」.
http://www.kantei.go.jp/jp/singi/titeki2/digitalarchive_kyougikai/houkokusho.pdf. (accessed 2021-04-05)

デジタルアーカイブの起源

- 1994年
 - 月尾嘉男により「デジタルアーカイブ」という用語が提案される。
 - 月尾嘉男. "デジタル・アーカイブの功罪". <http://www.tsukio.com/denki2.html>
 - “デジタルアーカイブという言葉を生んだ「月尾嘉男」”. https://artscape.jp/artscape/artreport/it/k_0401.html.
- 1996年
 - デジタルアーカイブ推進協議会（JDAA）、設立。
- 2017年
 - デジタルアーカイブ学会、設立。
 - デジタルアーカイブ推進コンソーシアム（DAPCON）、設立。
- より詳細は、以下など参照のこと。
 - デジタルアーカイブ年表 - infocom. <https://service.infocom.co.jp/das/column/column1/chronology.html>.

デジタルアーカイブに関わるコミュニティ

- 情報処理学会・人文科学とコンピュータ研究会
- アート・ドキュメンテーション学会
- 記録管理学会
- 映像情報メディア学会
- 情報メディア学会
- 情報知識学会
- 電子情報通信学会
- 日本情報考古学会
- 画像電子学会
- 日本色彩学会
- デジタルアーカイブ学会
- 文化資源学会
- 人工知能学会
- 日本図書館情報学会
- 日本写真学会
- NPO法人知的資源イニシアティブ
- NPO法人地域資料デジタル化研究会
- NPO法人地域資料情報化コンソーシアム
- NPO法人沖縄デジタルアーカイブ推進協議会
- 全国歴史資料保存利用機関連絡協議会
- 情報保存研究会
- など

なぜデジタルアーカイブが必要か

- **現代社会の基盤は知的文化情報資源のストックであるため**
- それらのアクセシビリティを向上させることで、知的生産性のサイクルは向上するはず。

- 欧米や東アジアなどでは、すでにこの分野で先進的取り組みが数多く展開され、またこのような政策に優先的に資源配分されている。

デジタルアーカイブの活用可能性



図. デジタルアーカイブ社会のイメージ (例)

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「我が国におけるデジタルアーカイブ推進の方向性」(p.3) より

アーカイブ機関のメリット



図. アーカイブ機関のメリット

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「我が国におけるデジタルアーカイブ推進の方向性」(p.4)より

参考書

- 柳与志夫. 2020. デジタルアーカイブの理論と政策：デジタル文化資源の活用に向けて. 勁草書房.

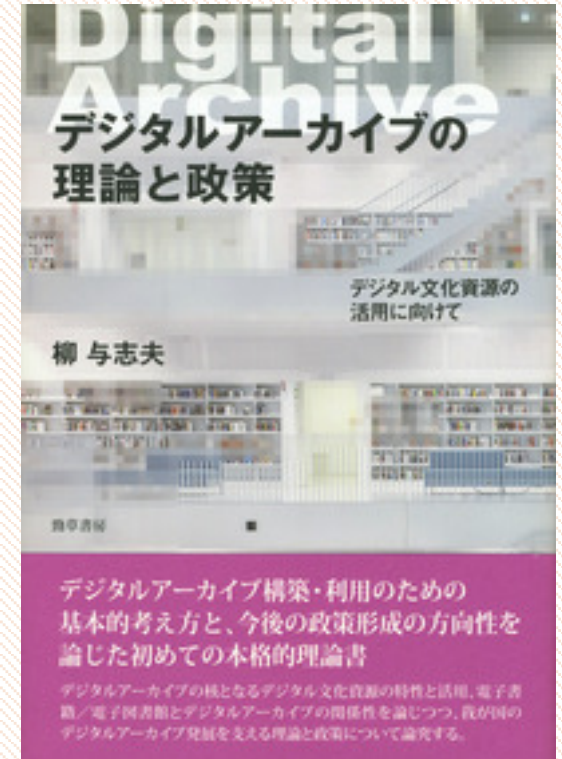


図. 書影

デジタルアーカイブを構成するもの

- 柳は、デジタルアーカイブを構成するものとして以下の7件を提示している（柳 2020, pp. 174-181）。
- 1. デジタルコレクション
- 2. メタデータとデータ構造化
- 3. データの保管・保存
- 4. 検索エンジンとUI
- 5. システム
- 6. 利用
- 7. 運営・マネジメント

デジタルコレクション

- デジタルコンテンツの集合としてのコレクション
- 提供機関の収集・整理の**方針に基づき編成**される
- これまでは有形の資料がデジタル化によりコンテンツとなることが多かった。
 - 原資料とデジタルコンテンツやメタデータをセットで管理する必要性がある。
- 近年のデジタル・ネイティブの資料の増加によりそのあり方が変化しつつある。

メタデータと データ構造化

- メタデータとは「データに関する（構造化された）データ」。
- コンテンツの検索・識別や分析、アクセシビリティの向上に寄与する。
- 図書館情報学における**目録**の考え方を基礎として発展。
- 資料ごとの特徴に基づく要件定義と**オントロジー構築**（データモデリング）が求められる。
- TEIなど、コンテンツのデータ構造化のための標準も存在。

Snorql for Japan Search	
Description of < https://jpsearch.go.jp/data/syozo-100292 >	
rdf:type	type:絵画 (絵画)
rdfs:label	"睡蓮"
schema:name	"Water Lilies"@en "睡蓮"@ja
schema:creator	chname:クロード・モネ (クロード・モネ)
schema:temporal	time:1916 (1916年)
schema:dateCreated	"1916"
schema:description	"form size(cm): 200.5 x 201" "main classification : Paintings" "material(including technique and support): oil on canvas" "mode of acquisition: Matsukata Collection" "name of the museum: The National Museum of Western Art, Tokyo" "position of signature: Signed and dated lower left: Claude Monet 1916" (...show all 15)
schema:image	< https://search.artmuseums.go.jp/jpeg/small/nmwa/0001510006.jpg >
jps:agential	jps:relationType role:制作 (制作) jps:value chname:クロード・モネ (クロード・モネ) schema:description "作家名 (生没年) : モネ、クロード (1840 - 1926)"
jps:temporal	jps:relationType role:制作 (制作) jps:value time:1916 (1916年) schema:description "制作年(和暦) : 1916年" "制作年(検索用) : 1916" "制作年(西暦) : 1916"
jps:accessInfo	< https://jpsearch.go.jp/data/syozo-100292#accessinfo >



図. 睡蓮 - Snorql for Japan Search. <https://jpsearch.go.jp/data/syozo-100292>.

データの保管と保存

- 永続的なアクセス保証のためには、データの保管や保存が前提となる。
- データ形式・ソフトウェア・ハードウェアをセットで考える必要性がある。
- クラウドサーバなどを利用した「3-2-1ルール」の適用など。

検索エンジンとUI

- テキストによるマッチングやインデキシングが基本技術。
 - 様々な検索方法は存在するが、上記が一般的に用いられる。
- ユーザインターフェイス（UI）は多様な形式・実装がありえる。
 - 資料やユーザグループの特徴に基づき、設計・構築する必要性がある。
 - **クラウドソーシング**によるユーザ関与によるコンテンツ強化なども注目されており、その場合、ユーザ管理機能やゲーミフィケーション要素など、特有の機能を実装する必要性が生じる。

システム

- コンテンツの入力・データ処理・データ格納・データベース管理と操作・検索・UI・業務支援などが用いられる。
- これらの基礎機能は特別な要素技術を必要としない場合が多いため、標準的なシステム・技術を用いることが有効。
 - デジタルコレクションのマネジメントシステム：[Omeka](#), [Drupal](#), [AtoM](#), [ADEAC](#) ...
 - 要素技術：TEI（XMLによるテキスト構造化）、IIIF（画像公開API）、Linked Data（統一的なリソース記述法） ...

利用

- サービス構築にあたっては、その設計のためユーザ像を指定する必要がある。ただし、想定しないユーザ群や使われ方をすることも多い。そのような場合、より発展性があるともいえる？
- 継続的なシステムの改修やコンテンツ追加・更新を行い、ユーザコミュニティの形成・発展により、利用は促進される。
 - データソン・アイデアソンなどによるコミュニティ形成支援が有効。
- Linked Dataや可視化ツールなどによる、分析環境の提供も有効。

運営・マネジメント

- 単にデジタルアーカイブを公開しただけでは、うまくいかない場合が多い。しかも、デジタルアーカイブはプロジェクト型、つまり有期的な運営である場合が多い。
- 継続的なアクセスの保証や利活用の支援のため、運営や体制づくりが求められる。
- 運営のための資金の獲得のためにはいくつかの方法論がある。
 - e.g. 研究費や補助金の獲得、寄付金、サービス利用料、広告収入、クラウドファンディング、非営利法人化による会費収入、関連グッズ開発

日本におけるデジタルアーカイブ概況

- 国内でも文化・知的資源の提供は様々行われている
- ただし、欧米や中国・韓国などと比べても「**量的に十分な状況とは言えない**（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., p. 10）」
 - 結果として、海外の日本研究も減少傾向に
 - 海外の動向については次回以降で紹介

日本における課題

- **オープンな標準（Open Standards）**に関するノウハウの構築・共有
 - APIの提供が限定的
 - 専用アプリケーションが必要な場合がおおく、汎用性が低い
- 再利用性・アクセシビリティの高い**コンテンツデータ**の提供
 - 画像のみでテキスト化などが不十分、そのためアクセス性が低い
- **二次利用**規定の整備
 - 提供するデータの再利用性が低く、結果としてサービスの価値が向上しない

政策形成における必要性

1. 欧米さらには東アジアなどでは、文化資源政策は優先的に取り組まれている。文化力を背景とする、国の政治的・経済的戦略の構築が改めて重要視されている。
2. 出版などプラットフォームの寡占化が進み、各国の産業は強く影響を受けている。
3. デジタル化による文化資源管理の制度・技術の変容が進んでおり、今後の産業構造に影響する。
4. ハコモノ行政からの脱却の代表的アプローチとして、知的・文化資源政策に注目が集まっている。
 - 『デジタルアーカイブの理論と政策』（pp. 7-8）より

政策的アプローチの問題点

- 旧来の文化資源に関する政策は、各省庁さらには自治体など、**個別のアプリケーション**による、独自路線で進められてきた。そのため**小規模で再利用性の低い文化資源アーカイブが乱立**。
 - 縦割り行政的弊害
- 結果として**電子図書館構築で失敗**。
 - 中井万知子. 2019. “「電子図書館」再考(前篇).” 立正大学図書館司書課程年報 = *Annual Report of the Librarian Course, Rissho University*, no. 5: 2–23. <https://doi.org/info:doi/10.34386/00007710>.
 - 中井万知子. 2020. “「電子図書館」再考(後篇).” 立正大学図書館司書課程年報 = *Annual Report of the Librarian Course, Rissho University*, no. 6: 17–44. <https://doi.org/info:doi/10.34386/00007671>.

図書館情報学アプローチの有効性

- 柳（2020, pp. 12-14）は以下の論点から、**図書館情報学アプローチ**による文化資源政策に対する**有効性**を指摘した。
 1. 学問体系としての歴史
 2. アクセス向上という目的に対する普遍性・体系性と真正性
 3. 「場」に関する関心の高さ
 4. 利用者サービスを一体的に論じる
 5. 多様性の保証

今後の方向性

- コンテンツ作成・データ提供の方法論の**共通化**（標準化）
- 機関・個人の**連携強化**
 - 学術・技術をテーマとする会議や研究会の開催
- **専門人材の育成**
- 国際会議の参加、留学生の行き来、国際プロジェクトの推進など海外との交流を通じた最新動向の共有
- 安定的な予算の確保
 - プロジェクト型活動の有期性についての課題

デジタルアーカイブをめぐる 標準技術

代表的な標準化技術の紹介

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。

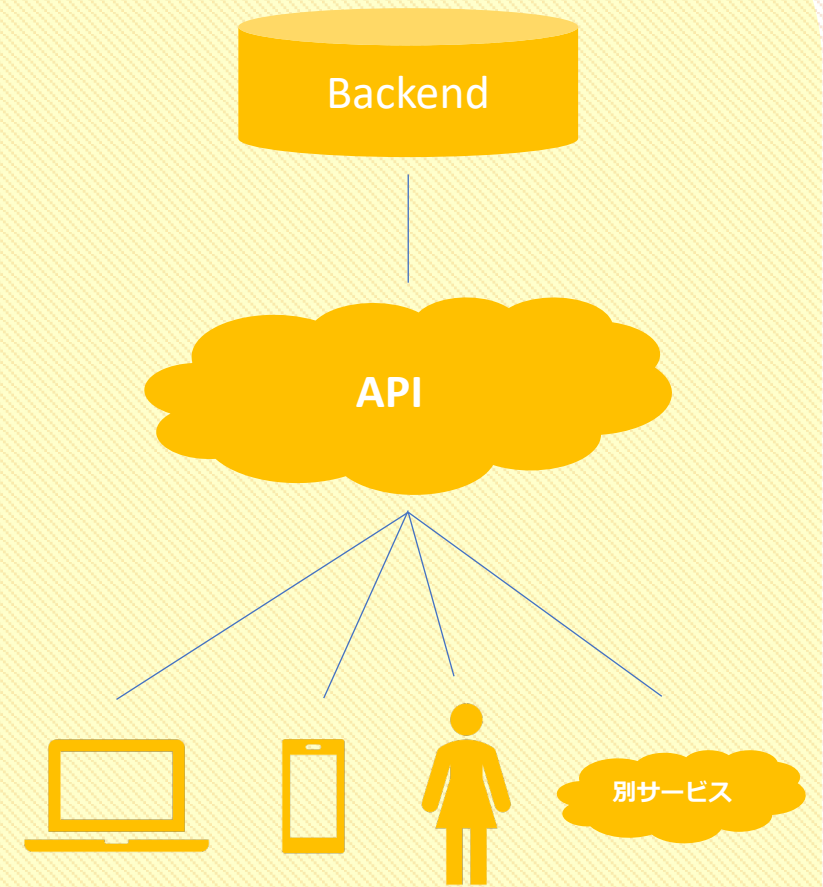


図. APIイメージ

API

- 「Web 2.0」 (2005～) の基礎要素としても注目される
- **APIファースト**
 - ウェブサービス開発において、まずAPIの設計や開発を進めてから、ビューや他のシステムの接続などの開発を進めるという開発法
- APIを提供することで、それを軸とするエコシステム (≒APIエコノミー) が形成される。それにより、API提供サービスの価値向上につながる。
 - **ネットワーク外部性**が強く影響する

APIの例

- 小売 : Amazon MWS API
- 金融 : Moneyforward, freee, 各銀行API
- 気象 : Open Weather Map API, Dark Sky API
- 地理 : Google Maps API, 電子国土web
- SNS : Twitter API, Facebook API
- 図書・論文 : NDL (国立国会図書館), OpenBD, CiNii



```
JSON 生データ ハッター
保存 コピー すべて折りたたむ すべて展開 JSON を検索
▼ @context:
  rdfs: "http://www.w3.org/2000/01/rdf-schema#"
  dc: "http://purl.org/dc/elements/1.1/"
  dcterms: "http://purl.org/dc/terms/"
  foaf: "http://xmlns.com/foaf/0.1/"
  prism: "http://prismstandard.org/namespaces/basic/2.0/"
  con: "http://www.w3.org/2000/10/swap/pim/contact#"
  cinii: "http://ci.nii.ac.jp/ns/1.0/"
  bibo: "http://purl.org/ontology/bibo/"
  @id: "https://ci.nii.ac.jp/naid/40022512356.json"
▼ @graph:
  ▼ @:
    @id: "https://ci.nii.ac.jp/naid/40022512356#article"
    @type: "bibo:Article"
    foaf:isPrimaryTopicOf:
      @id: "https://ci.nii.ac.jp/naid/40022512356.json"
    dc:title:
      ▼ @:
        @value: "MAGIA sessions : マンガ・アニメーション・ゲーム・メディアアートのアーカイブの現在"
    dc:publisher:
      ▼ @:
        @value: "美術出版社"
    prism:publicationName:
      ▼ @:
        @value: "美術手帖"
    prism:issn: "0287-2218"
    prism:volume: "73"
    prism:number: "1087"
    prism:startingPage: "138"
    prism:endingPage: "145"
    prism:publicationDate: "2021-04"
    dc:date: "2021-04"
    cinii:naid: "40022512356"
    cinii:ncid: "AN10310657"
    cinii:ndljpi: "031336645"
    dc:language: "JPN"
    dc:source:
      ▼ @:
        @value: "NDL"
    dcterms:isPartOf:
      @id: "https://ci.nii.ac.jp/ncid/AN10310657#entity"
      dc:title: "美術手帖"
    rdfs:seeAlso:
      ▼ @:
```

図. CiNiiのJSONレスポンス

```

version="1.0" encoding="UTF-8"?>
model href="tei_all_ja_20191012_4.rnc" type="application/relax-ng-compact-syntax"
xmlns="http://www.tei-c.org/ns/1.0" xmlns:eaj="http://www.example.org/ns/ejaTEI"
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>走れメロス</title>
      <author>太宰治</author>
    </titleStmt>
    <publicationStmt>
      <distributor>青空文庫</distributor>
      <authority>金川一之</authority>
      <authority>高橋美奈子</authority>
      <date when="2011-01-17"> 2011年1月17日</date>
    </publicationStmt>
    <sourceDesc>
      <bibl>
        <author>太宰治</author>
        <title>走れメロス</title>
        <publisher>筑摩書房</publisher> 「太宰治全集3」 ちくま文庫、 <date when="
          >1988 (昭和63) 年10月25日</date>初版発行 <date when="
          >1998 (平成10) 年6月15日</date>第2刷 </bibl>
      </sourceDesc>
    </fileDesc>
    <revisionDesc>
      <list>
        <item>
          <date when="2011-01-17">2011年1月17日</date>修正 </item>
        <item> 入力<persName>金川一之</persName> 校正<persName>高橋美奈子</persName>
          <date when="2000-12-04"> 2000年12月4日</date>作成 </item>
      </list>
    </revisionDesc>
  </teiHeader>
  <text>
    <front>
      <p>
        <listPerson>
          <person xml:id="メロス">
            <persName>メロス</persName>
            <occupation>牧人</occupation>
          </person>
        </listPerson>
      </p>
    </front>
  </text>

```

TEI

- Text Encoding Initiatives
- 人文学資料の適切なマークアップ (XML) のための仕様・標準、それを構築するためのプロジェクト。
- The TEI Guidelines
 - <https://www.tei-c.org/release/doc/tei-p5-doc/ja/html/index.html>
- 日本語向けTEIガイドライン (作成中・試行版)
 - https://github.com/TEI-EAJ/jp_guidelines/wiki

図. マークアップサンプルファイル「走れメロス」.
https://www.dhii.jp/nagasaki/dazai_all_20191012.xml

Gettyと「どうぶつの森」

Getty

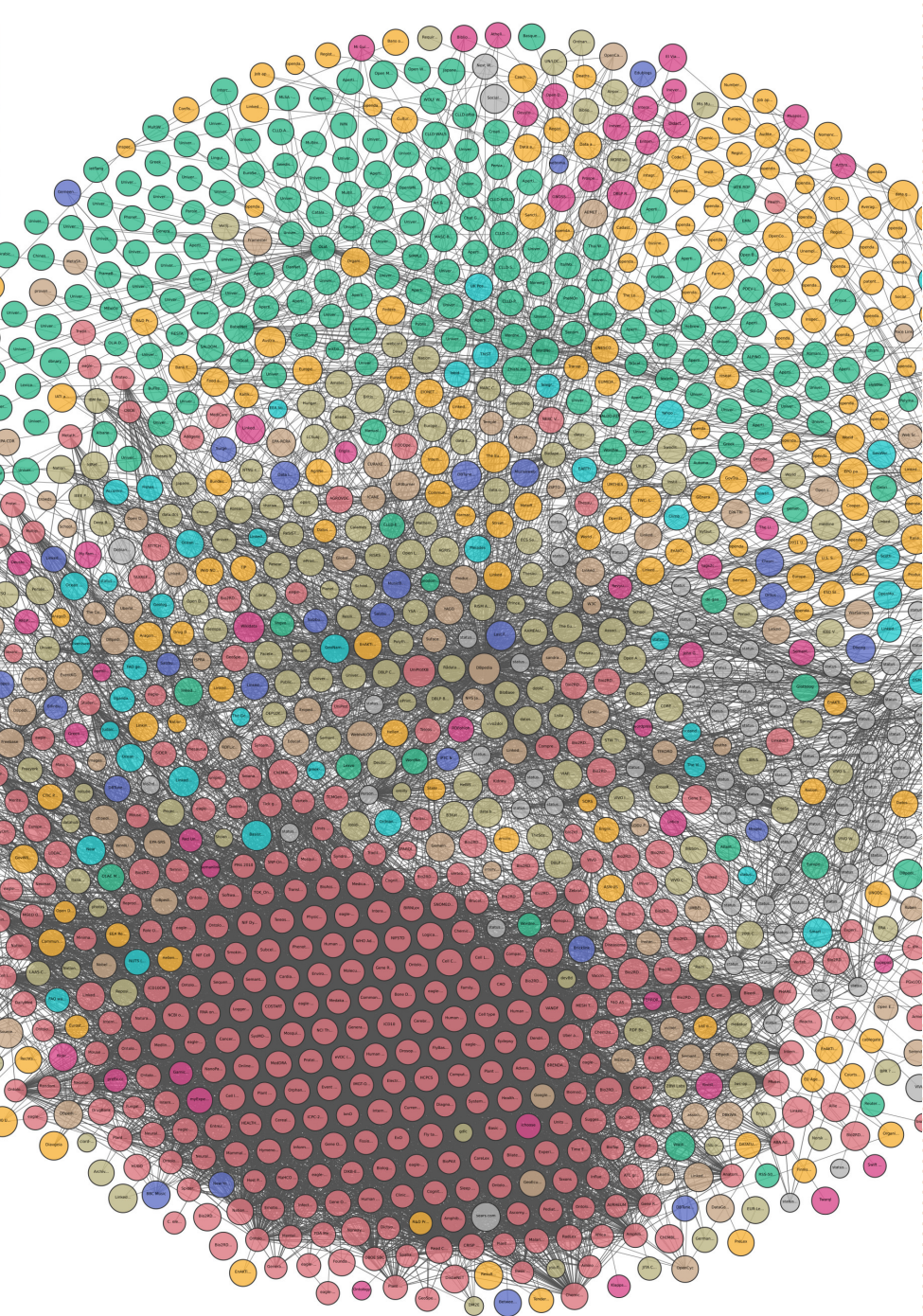
Animal Crossing Art Generator

Museum Research Institute Conservation Institute Foundation Support Us

Visit What's On Explore Art Resources Our Work About 



 Animal Crossing Art Generator <https://experiments.getty.edu/ac-art-generator#iifloader> (accessed 2020-06-29)



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/>.

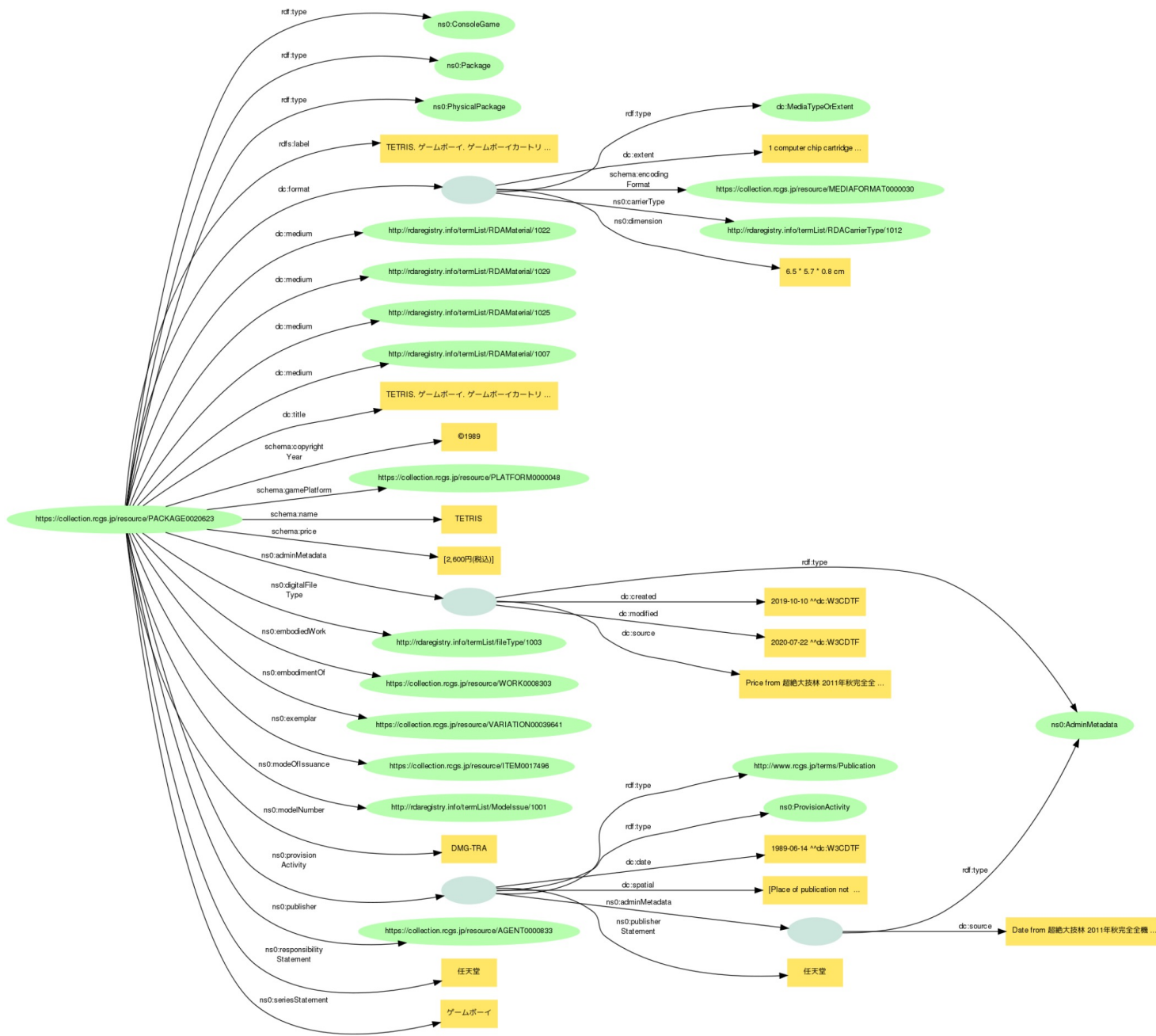


図. ゲームボーイ版テトリスのRDFによる書誌データのグラフ
The Web of Kanzaki – RDFグラフの視覚化により生成