

第8回東京大学学術資産アーカイブ化推進室主催セミナー「光源氏はデジタルアーカイブの夢を見るか？」

ミニマムなデジタルアーカイブシステム 構築の実践について

2025/01/28(火)

東京大学大学院人文社会系研究科文化資源学研究専攻

阿達 藍留

自己紹介

- 阿達 藍留 (ADACHI Airu)
- 東京大学大学院人文社会系研究科 文化資源学研究専攻 M2
 - 先月、無事に修士論文を提出しました
 - 題目は「持続可能な資料共有を実現するためのデジタルアーカイブ構築手法に関する研究」
- 「大江健三郎文庫」のデジタルアーカイブ構築に参画 (2022.03～)

研究テーマ

- デジタルアーカイブによる、持続可能な資料共有の実現
- デジタルアーカイブを、どうすれば……
 - 「低コスト」に
 - 「だれでも」作れて
 - それでいて「ちゃんと使える」 ものにできるか？

(1) 低コストに

- 「コスト」は、単に予算や金銭の経済的な負担に限られない
 - 情報技術や運用方法を学ぶために割かれる**時間と労力**
 - 外部のエンジニアや業者との、専門用語を含む**やりとりの負担**
 - データの整備や更新、メンテナンスに割かれる**時間**
 - 正常に運用されているか、不具合が生じたらどうするかなど、運用に関わる不確実性やリスクに対する**心理的負荷(ストレス)**

(2) だれでも

- 少人数で資料の整理・保存・公開を行う、**個人や小規模な機関***
- デジタルアーカイブの構築や運用のための専任の人材がおらず、外部委託も難しい場合がある

* 「小規模な機関」:一例として、文部科学省「令和3年度社会教育調査」における「専任学芸員数及び学芸員補数別博物館数」の館あたり人数の集計区分を参考に、資料の整理・保存・公開に携わる専任職員が概ね5名以下の機関を想定

(3) ちゃんと使える

- なぜ資料を公開するのか？→使ってもらいたい(使ってもらう必要がある)から
- 使いやすくないと使ってもらえない→**閲覧や検索の利便性**
- そもそも見つけてもらえなければ使ってもらえない→**外部連携**
 - 元の資料の価値が並外れて広く知られている場合を除けば……

[参考] 福島(2016)の「スリムモデル」

- 慶應義塾大学文学部の福島幸宏先生による、デジタルアーカイブの最小限の要件を定義し、コスト削減を目指したモデル提案
- 「必要最低限」と「理想的な状態」の境界を明示することを重視
- 利用規約の明示・機械可読性の担保・環境に依存しないデータ移行性の担保・アクセシビリティの確保の4要件
- より発展的なデジタルアーカイブの指針として「ノーマルモデル」もある

福島幸宏 (2016) 「ガイドラインに要れるべき要件」

https://www.kantei.go.jp/jp/singi/titeki2/digitalarchive_kyougikai/meta_data/dai2/siryous3_3.pdf

では、具体的に何をすべきか？ の一案 (1/2)

- 静的サイト構成の採用
 - 閲覧者からのリクエストに応じて事前に用意された Web ページ群の中から適切な情報を配信する形態のサイト
 - 主にHTML・CSS・JavaScriptと画像などのリソースで構成される
 - 複雑な処理をするサーバが不要なため、安価かつサイバー攻撃の危険性も小さい
 - 機能が制限されるという欠点も……

では、具体的に何をすべきか？ の一案 (2/2)

- 最新の Web 技術を活用した閲覧性の向上
 - ユーザ側のブラウザで動作するJavaScriptライブラリなどを活用し、静的サイトでも高度な検索、絞り込みや画像閲覧の機能を提供
- 外部連携用メタデータの整備
 - 小規模なデジタルアーカイブでは外部の大規模アーカイブやアグリゲータ(例: ジャパンサーチ)との連携が不可欠。適切な構造化により機械可読性が担保されたメタデータを出力

提案したモデルの実装例: DAKit

- デジタルアーカイブの公開用のWebサイトを作成するツール「DAKit」
 - 阿達藍留, 山田俊幸, 大向一輝. DAKit: 低コストなデータ共有のための 静的デジタルアーカイブジェネレータの提案. 情報知識学会誌. 2022, vol. 32, no. 4, p. 406-409.
<http://doi.org/10.2964/jsik.2022.035>
 - 阿達藍留, 大向一輝. [C13] 静的デジタルアーカイブジェネレータDAKitを用いたメタデータ流通のワークフローに関する検討. デジタルアーカイブ学会誌. 2023, vol. 7, no. s2, p. s79-s82.
<http://doi.org/10.24506/jsda.7.s2.s79>

DAKitの設計方針

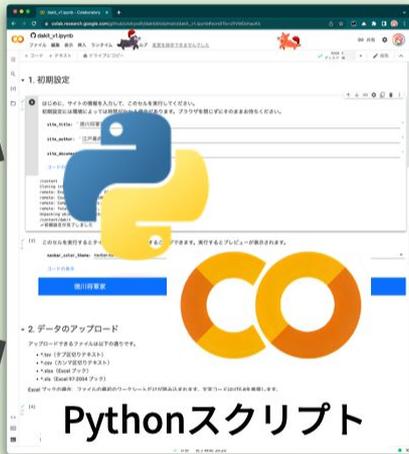
- 資料に関わる当事者自身(資料の専門家であり情報技術の専門家ではない)が利用することを想定
 - 利用者はデータの作成と管理に専念できるようにする
 - 特定の開発環境を自身で用意する必要がないようにする
- 処理内容がブラックボックスとならず、自らの手で改良できることを目指した

	A	B	C	D	E
1	item_id	代	氏名	氏名カナ	生年
2	001	1	徳川 家宣	トクガワ イエヤス	1542
3	002	2	徳川 綱吉	トクガワ ヒデタダ	1579
4	003	3	徳川 家光	トクガワ イエミツ	1604
5	004	4	徳川 家康	トクガワ イエツナ	1641
6	005	5	徳川 綱白	トクガワ ツナヨシ	1646
7	006	6	徳川 家宣	トクガワ イエノブ	1662

資料のデータ

	A	B	C	D	E	F
1	item_id	url				
2	001	https://utdh.mil/dakit/tokugawa_demo/images/001/001_1.jpg				
3	001	https://utdh.mil/dakit/tokugawa_demo/images/001/001_2.jpg				
4	001	https://utdh.mil/dakit/tokugawa_demo/images/001/001_3.jpg				
5	001	https://utdh.mil/dakit/tokugawa_demo/images/001/001_4.jpg				
6	002	https://utdh.mil/dakit/tokugawa_demo/images/002/002_1.jpg				
7	002	https://utdh.mil/dakit/tokugawa_demo/images/002/002_2.jpg				
8	002	https://utdh.mil/dakit/tokugawa_demo/images/002/002_3.jpg				
9	002	https://utdh.mil/dakit/tokugawa_demo/images/002/002_4.jpg				
10	003	https://utdh.mil/dakit/tokugawa_demo/images/003/003_1.jpg				

資料の画像



Google Colaboratoryで動作
手元での環境構築不要



検索・画像表示機能付き
デジタルアーカイブ

- 資料のメタデータ(表形式、Excelブック)
- 資料の画像(JPEG、PNGほか IIIF対応可能)
- データを読み込み、検索ページ・個別ページを自動生成

	B	C	D
①	資料番号	タイトル	別タイトル
	I-01-001	早飛脚廻りにてくわしき所本しらべ大地震	
3	002	I-01-002 江戸大地震	
4	003	I-01-003 関東大地震并處々出火細見	
5	004	I-01-004 聞書諸国大地震并出火	
6	005	I-01-005 江戸大地しんの図	
7	006	I-01-006 東北海岸の海嘯	
8	007	I-01-007 江戸大地震出火場所附	
9	008	I-01-008 諸国大地震之図	
10	009	I-01-009 関東江戸大地震并大火方角場所附	
11	010	I-01-010 安政五年午十一月十五日江戸大火細見図	
12	011	I-01-011 東海道大阪辺大地震津波図	東海道大
13	012	I-01-012 大地震大津波末代噺 二編	
14	013	I-01-013 ゆるがぬ御代要之石寿栄	
15	014	I-01-014 落石山非常明工	

	B	C	D	E
②	path			
	001	https://adachi-a.github.io/dakit2/items/images/001.jpg		
3	I-01-001	https://adachi-a.github.io/dakit2/items/images/002.jpg		
4	I-01-002	https://adachi-a.github.io/dakit2/items/images/003.jpg		
5	I-01-002	https://adachi-a.github.io/dakit2/items/images/004.jpg		
6	I-01-002	https://adachi-a.github.io/dakit2/items/images/005.jpg		
7	I-01-003	https://adachi-a.github.io/dakit2/items/images/006.jpg		
8	I-01-004	https://adachi-a.github.io/dakit2/items/images/005.jpg		
9	I-01-004	https://adachi-a.github.io/dakit2/items/images/004.jpg		

①資料のメタデータ

②資料と画像ファイルの対応表

③資料画像



dakit_v3.ipynb - Colab

colab.research.google.com/drive/1vy4-oYme7Ggtqx1W5UICKZrYHEC-p_m#scrollTo=bDjm-ieYU10U

dakit_v3.ipynb ☆

ファイル 編集 表示 挿入 ランタイム ツール ヘルプ 最終編集: 1月15日

+ コード + テキスト

接続 Gemini

下のセルを実行すると、ページが生成されます。

```
display_rows = setting_df[setting_df[2] == "1"][0].values
filter_rows = setting_df[setting_df[3] == "1"][0].values
sort_rows = setting_df[setting_df[4] == "1"][0].values
search_target_rows = setting_df[setting_df[5] == "1"][0].values

dakit_item_id_column_name = id_row
dakit_item_title_column_name = title_row

fuse_keys = list(search_target_rows)

# ソートの設定
itemsjs_sortings = {}
for row in sort_rows:
    itemsjs_sortings[f"{row}(昇順)"] = {"field": row, "order": "asc"}
    itemsjs_sortings[f"{row}(降順)"] = {"field": row, "order": "desc"}

# フィルタの設定
itemsjs_aggregations = {}
for row in filter_rows:
    itemsjs_aggregations[row] = {"title": row, "conjunction": False}

# 検索ページの表示方法の設定
dakit_display_keys = list(display_rows)
dakit_display_keys_string = ",".join(dakit_display_keys)
```

https://adachi-a.github.io/dakit2/

DAKit Test Site

Search...

並び替え: 資料番号 asc

1ページあたりの表示数: 10

内容分類

災害

37

その他

2

名所案内

1

資料分類

かわら版

35

その他

3

引札

1

番付

1

印刷分類

木版(墨一色)

25

木版(多色)

14

40件の検索結果

[早飛脚廻りにてくわしき所](#)

災害 かわら版 木版(多色) 嘉永7年6月

[江戸大地震](#)

災害 かわら版 木版(墨一色) 安政2年10月

[関東大地震并處々出火細見](#)

災害 かわら版 木版(墨一色) 安政2年10月

[聞書諸国大地震并出火](#)

災害 かわら版 木版(墨一色) 嘉永7年6月

[江戸大地しんの図](#)

災害 かわら版 木版(多色) 安政2年10月

[東北海岸の海嘯](#)

災害 かわら版 木版(多色) 明治29年6月

[江戸大地震出火場所附](#)

災害 かわら版 木版(墨一色) 安政2年10月

[諸国大地震之図](#)

災害 かわら版 木版(墨一色) 嘉永7年11月

[関東江戸大地震并出火](#)

DAKit Test Site

早飛脚廻りにてくわしき所本しらべ大地震



id	1
資料番号	I-01-001
タイトル	早飛脚廻りにてくわしき所本しらべ大地震
別タイトル	
内容分類	災害
資料分類	かわら版
印刷分類	木版(多色)
年月日	嘉永7年6月4日
サイズ	35x49
備考	

採用した技術

- アイテム別ページ: Python (テンプレート言語: Jinja2)
- 検索ページ (<https://github.com/adachi-a/itemsjs-svelte-demo>)



Svelte: 高速で軽量なJavaScriptフレームワーク(正確にはコンパイラ)

- ItemsJS: クライアントサイドで絞り込み検索を実現するライブラリ



Vite: 高速なビルドツール

- 実行環境: Google Colaboratory を想定

活用事例

- 東京学芸大学附属竹早小学校「竹早アーカイブ」
<https://www.slideshare.net/slideshow/10-steps-for-developing-school-digital-archives/271638862>
 - 児童が授業中に制作した学習成果物をデジタルアーカイブ化し、記念や将来の指導資料として活用する
 - **公開範囲のコントロールが必須、低コストでの構築・保守が可能、専門知識がなくても維持できることから採用**
- 大江健三郎文庫自筆原稿デジタルアーカイブ（DAKitをベースにシステムを開発）

竹早アーカイブ

Search...

並び替え: **作品番号(昇順)** ▼
1ページあたりの表示数: **10** ▼

学年

- 6年 **18**
- 5年 **16**

教科

- たけのこタイム **34**

カテゴリ

- 園工・美術 **21**
- 家庭科 **6**
- 音楽 **3**
- その他 **2**
- 家庭科・手芸 **2**

34件の検索結果

ロボットAタイプ
作品番号: 2023-0001 / 作品名: ロボットAタイプ / 作成者: [不明] / 学年: 6年 / 作成年: 2023 / 教科: たけのこタイム / カテゴリ: 園工・美術 / カテゴリ-2: 折り紙 / グループ名: シン折り紙クラブ / 作品形式: 立体製作物 / キーワード1: ロボット / キーワード2: 創作 / キーワード3: 折り紙 / キーワード4: / キーワード5:

ロボットBタイプ
作品番号: 2023-0002 / 作品名: ロボットBタイプ / 作成者: [不明] / 学年: 6年 / 作成年: 2023 / 教科: たけのこタイム / カテゴリ: 園工・美術 / カテゴリ-2: 折り紙 / グループ名: シン折り紙クラブ / 作品形式: 立体製作物 / キーワード1: ロボット / キーワード2: 創作 / キーワード3: 折り紙 / キーワード4: / キーワード5:

カムシ
作品番号: 2023-0003 / 作品名: カムシ / 作成者: [不明] / 学年: 6年 / 作成年: 2023 / 教科: たけのこタイム / カテゴリ: 園工・美術 / カテゴリ-2: 折り紙 / グループ名: シン折り紙クラブ / 作品形式: 立体製作物 / キーワード1: カムシ / キーワード2: 昆虫 / キーワード3: 折り紙 / キーワード4: / キーワード5:

フクロウ
作品番号: 2023-0004 / 作品名: フクロウ / 作成者: [不明] / 学年: 6年 / 作成年: 2023 / 教科: たけのこタイム / カテゴリ: 園工・美術 / カテゴリ-2: 折り紙 / グループ名: シン折り紙クラブ / 作品形式: 立体製作物 / キーワード1: フクロウ / キーワード2: 鳥 / キーワード3: 折り紙 / キーワード4: / キーワード5:

死神
作品番号: 2023-0005 / 作品名: 死神 / 作成者: [不明] / 学年: 6年 / 作成年: 2023 / 教科: たけのこタイム / カテゴリ: 園工・美術 / カテゴリ-2: 折り紙 / グループ名: シン折り紙クラブ / 作品形式: 立体製作物 / キーワード1: 死神 / キーワード2: / キーワード3: 折り紙 / キーワード4: / キーワード5:

シロサイ
作品番号: 2023-0006 / 作品名: シロサイ / 作成者: [不明] / 学年: 6年 / 作成年: 2023 / 教科: たけのこタイム / カテゴリ: 園工・美術 / カテゴリ-2: 折り紙 / グループ名: シン折り紙クラブ / 作品形式: 立体製作物 / キーワード1: シロサイ / キーワード2: 動物 / キーワード3: 折り紙 / キーワード4: / キーワード5:

巾着袋
作品番号: 2023-0007 / 作品名: 巾着袋 / 作成者: [不明] / 学年: 6年 / 作成年: 2023 / 教科: たけのこタイム / カテゴリ: 家庭科 / カテゴリ-2: 手芸 / グループ名: 手作りクラブ / 作品形式: 立体製作物 / キーワード1: 巾着袋 / キーワード2: / キーワード3: / キーワード4: / キーワード5:

ファスナーつきシンプルポーチ

竹早アーカイブ

ロボットAタイプ



作品番号	2023-0001
作品名	ロボットAタイプ
作成者	[不明]
学校種	小学校
学年	6年
指導者	宮寄佳智子
作成年	2023
教科	たけのこタイム
カテゴリ	園工・美術
カテゴリ-2	折り紙
グループ名	シン折り紙クラブ
作品形式	立体製作物
キーワード1	ロボット
キーワード2	創作
キーワード3	折り紙

大江健三郎文庫自筆原稿デジタルアーカイブの概要

- 自筆原稿や校正刷など約18,000枚の紙の自筆資料を撮影した画像データと、大江健三郎氏の著書・作品の情報から構成される
 - 画像は専門業者が撮影、データは大江文学の専門家が整理
- デジタルアーカイブは大江健三郎文庫内(本郷地区キャンパス)からのみアクセスが可能
 - 著書・作品の情報を抜き出したデータベースは一般公開
 - <https://open.l.u-tokyo.ac.jp/oe/>

自筆資料検索 自筆原稿 校正刷 その他

検索ワードを入力

324件の検索結果

自筆原稿

死者の奢り

短篇 資料ID m001 枚数 72

初出情報 「死者の奢り」 文藝界 1957年 8月 1日 11巻 8号

自筆原稿

空の怪物アグイー

短篇 資料ID m002 枚数 80

初出情報 「空の怪物アグイー」 新潮 1964年 1月 1日 61巻

自筆原稿

犬の世界

短篇 資料ID m003 枚数 51

大江健三郎文庫 データベース

まとめて検索 (β版)

凡例

文庫について

②使い方

まとめて検索 (β版)

 自筆資料 初出 書籍

検索ワードを入力

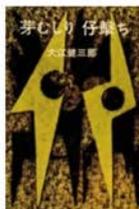
3435件の検索結果



書誌情報 (書籍)

『死者の奢り』

書籍ID b101 著者 大江健三郎 出版社 文藝春秋新社 出版年月日 1958(昭和33)年3月10日 判型 総312頁、B6判



書誌情報 (書籍)

『芽むしり仔撃ち』

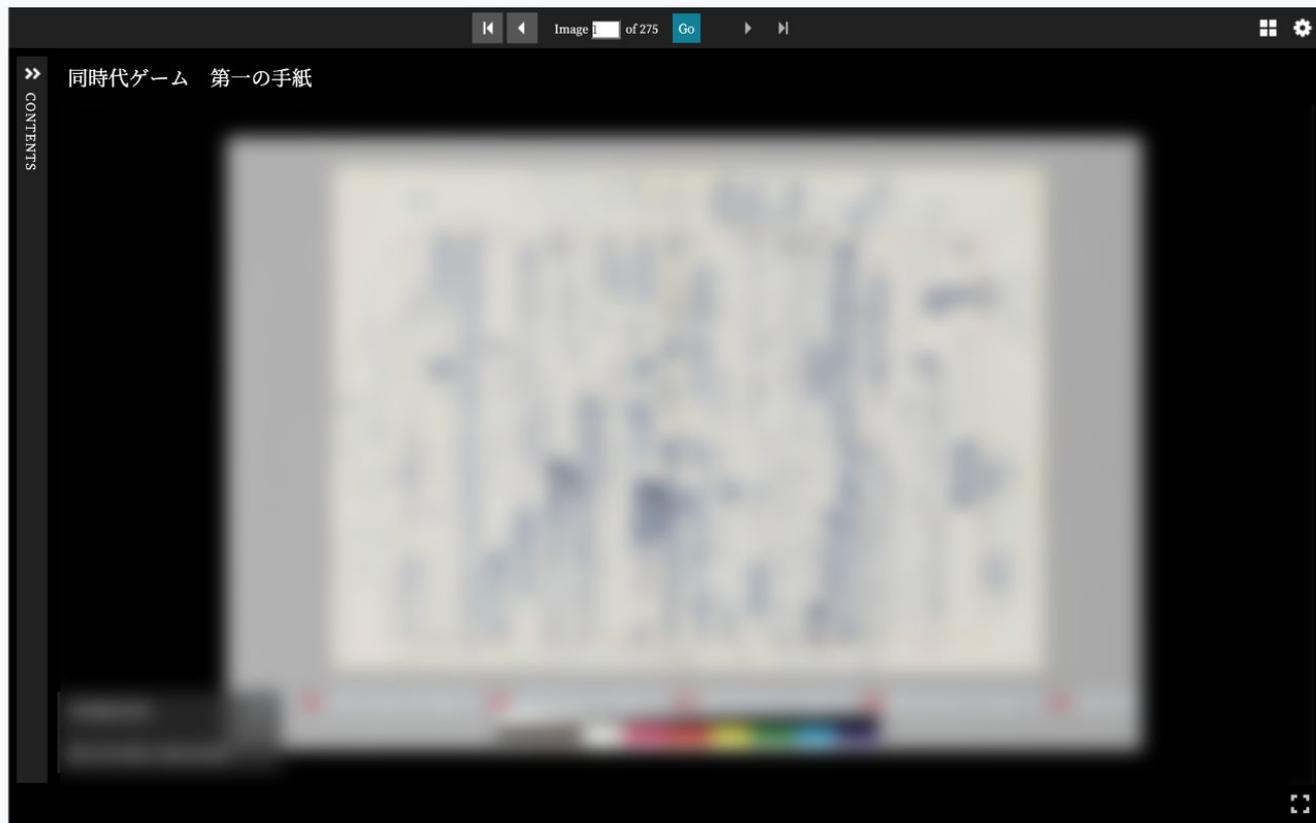
書籍ID b102 著者 大江健三郎 出版社 大日本雄辯會講談社 出版年月日 1958(昭和33)年6月25日 判型 総250頁、B6判



書誌情報 (書籍)

『見るまゑに跳べ』

書籍ID b103 著者 大江健三郎 出版社 新潮社 出版年月日 1958(昭和33)年10月31日 判型 総258



同時代ゲーム 第一の手紙

長篇 資料ID m010 全小説 8 枚数 275

[備考 (注記) ※手書きノンブル: 224, 225が逆 全小説: 第一の手紙 メキシコから、時のはじまりにむかって]

大江文庫DAにおける「ミニマムな」工夫 (1/2)

- **元データと画像から静的サイトを生成する構成を採用** (画像配信を除く)
 - 静的サイトジェネレータとしてEleventyを採用
 - 元データ(Excelブック)が更新されるたびに、サイト全体を構築し直す。サーバ上にデータベースなし
 - データの投入～アップロードは概ね1分以内に終了

阿達 藍留, 菊間 晴子, 阿部 賢一, 大向 一輝. グラフ構造を有する資料群を対象とした静的デジタルアーカイブの構築—「大江健三郎文庫自筆原稿デジタルアーカイブ」を事例として. 研究報告人文科学とコンピュータ(CH). 2024, vol. 2024-CH-135, no. 7, p. 1-5. <http://id.nii.ac.jp/1001/00234083/>

大江文庫DAにおける「ミニマムな」工夫 (2/2)

- 「文庫内限定のDA」と「一般公開のDB」でサーバ/サブドメインを分割
 - 画像データの「文庫内限定」アクセスを実現するための組み合わせ
 - 学内のIPアドレスからのアクセスのみ許可(クラウド)
 - ID/パスワードによるBASIC認証(Webサーバ)
 - 認証の独自実装を排することで開発・運用コストを削減

まとめ (1/2)

- DAKitのコードはGitHub上で公開
 - <https://github.com/utokyodh/dakit>
- ⚠ 現バージョンの外部連携用のメタデータの出力は、今後再実装予定
 - 今後は、生成されるWebページのブラッシュアップ、カスタマイズ項目の追加などを行う予定

まとめ (2/2)

- U-PARL(東京大学附属図書館アジア研究図書館上廣倫理財団寄付研究部門)のプロジェクトとして、アラビア文字写本集成「ダイバーコレクション」データベースの再構築を実施中
- 「ミニマムな」デジタルアーカイブのアプローチを横展開していく
- ご興味があれば、一緒にやってみましょう

ご清聴ありがとうございました