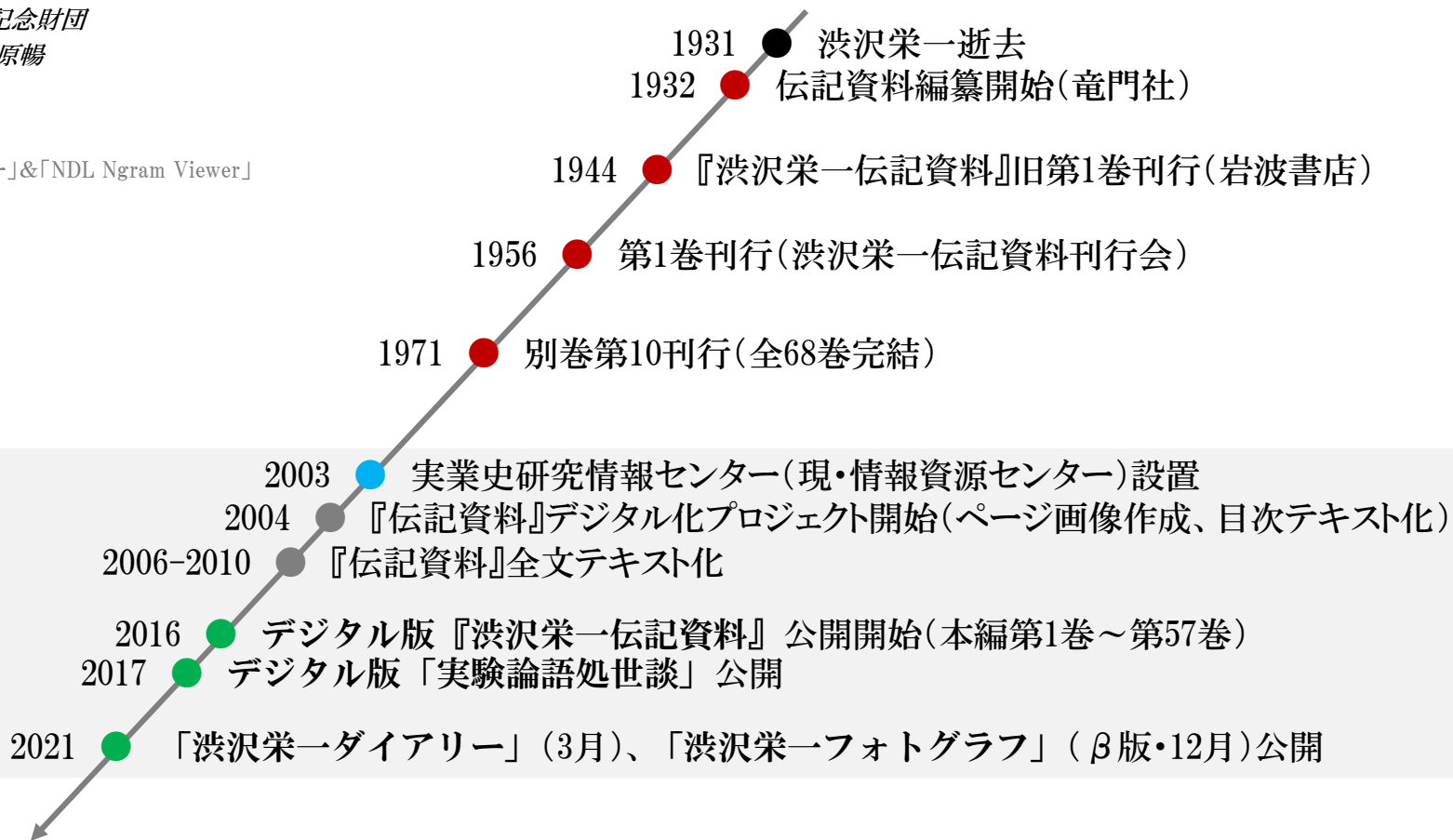


『渋沢栄一伝記資料』デジタル化の軌跡

公益財団法人渋沢栄一記念財団
情報資源センター長 茂原暢

#NDL全文使ってみた
「次世代デジタルライブラリー」&「NDL Ngram Viewer」
2022年11月1日(火)



1-1.『渋沢栄一伝記資料』とは

渋沢栄一(1840-1931):

- ・「近代日本経済の父」「論語と算盤」
- ・約500の企業、約600の社会公共事業に関与
- ・膨大な資料が残る

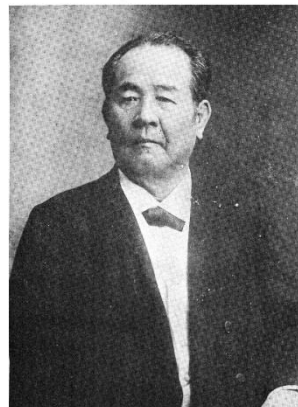
『渋沢栄一伝記資料』(渋沢青淵記念財団竜門社, 1955-1971):

- ・伝記ではなく伝記を書くための資料集
- ・全68巻(本編:57巻+索引巻、別巻:10巻)

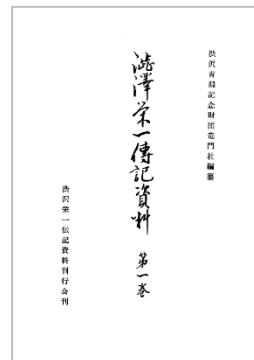
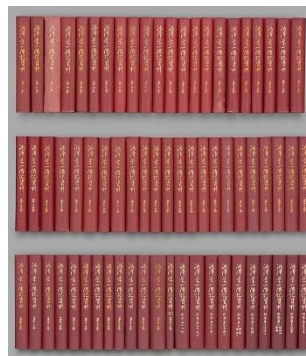
総計 約48,000ページ

本編:年代別／事業別の階層構造(最大7階層)

別巻:日記／書簡／講演録／写真など(種別ごと)



出典: 渋沢栄一フォトグラフ



1-2.『渋沢栄一伝記資料』のデジタル化

目的

- 『伝記資料』を対象とした情報アクセスの改善(全文検索)
- 渋沢栄一および日本近代史に関する一大情報源の創造

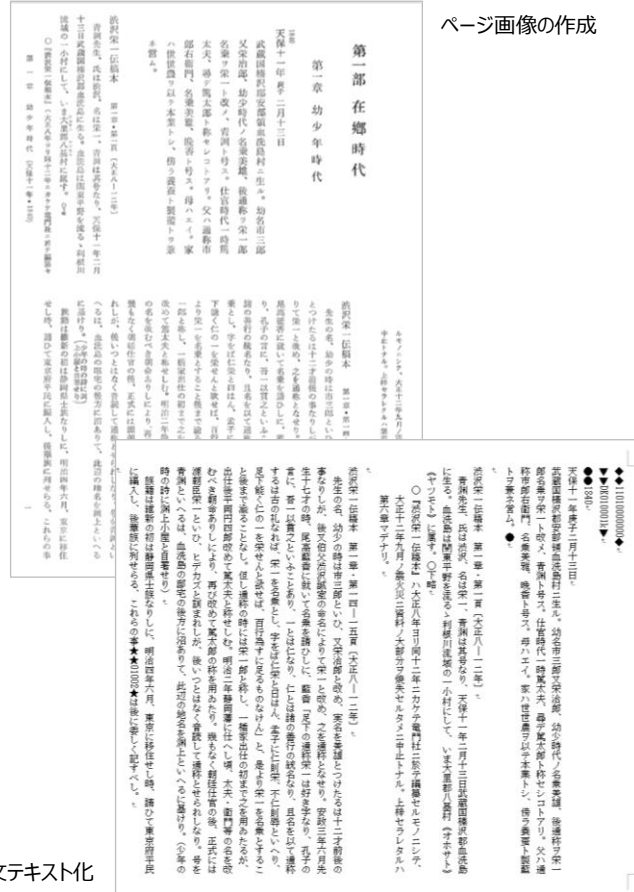
デジタル化の工程(2004年～2010年)

- 1) 2004年：ページ画像の作成(アクセスの拡張、閲覧環境改善)
- 2) 2004-2007年：目次(綱文)のテキスト化(内容へのアクセス)
- 3) 2006-2010年：全文テキスト化(アクセスの充実)

テキスト化の手法

- 手入力(当時のOCRでは対応困難という判断)
- 機械と目視で複数回校正を行うことで精度99.9%を確保

ページ画像の作成



全文テキスト化

2-1. デジタル版『渋沢栄一伝記資料』(2016.11)

<https://eiichi.shibusawa.or.jp/denkishiryō/digital/main/>

『渋沢栄一伝記資料』本編から 第1～57巻をテキスト、画像で公開

資料(抜粋)

網文

見出し

渋沢栄一伝稿本 第一章・第一頁 (大正八―二年)

青洲先生、氏は渋沢、名は栄一、青洲は其号なり、天保十一年二月十三日武蔵国榛沢郡血洗島に生る。血洗島は関東平野を流る、利根川流域の一小村にして、いま大里郡八基村に属す。

○『渋沢栄一伝稿本』ハ大正八年ヨリ同十二年ニカケテ竜門社ニ於テ編纂セラル。家ハ世世農ヲ以テ本業トシ、傍ラ養蚕ト製藍トヲ兼ネ営ム。

第一章 幼少年時代 (天保十一年・1840)

天保十一年 庚子 二月十三日

武蔵国榛沢郡安部領血洗島村ニ生ル。幼名市三郎又栄治郎、幼少時代ノ名乗美雄、後通称ヲ栄一郎名乗ヲ栄一ト改メ、青洲ト号ス。仕官時代一時篤太夫、尋テ篤太郎ト称セシコトアリ。父ハ通称市郎右衛門、名乗美雅、晩香ト号ス。母ハエイ。家ハ世世農ヲ以テ本業トシ、傍ラ養蚕ト製藍トヲ兼ネ営ム。

第一部 在郷時代

第一章 幼少年時代

『渋沢栄一伝記資料』第1巻より

デジタル版『渋沢栄一伝記資料』

伝記資料を検索 AND OR 検索 詳細検索へ >

TOP > 各巻リンク > 第1巻 目次【本文】 > 第1巻(DK010001k) 本文

凡例 更新履歴 このサイトについて

第1巻 (DK010001k) 資料リスト	第1巻 (DK010001k) 本文	次へ (DK010002k) >
-----------------------	--------------------	------------------

公開日: 2016.11.11 / 最終更新日: 2017.12.12

1編 在郷及び仕官時代

1郎 在郷時代

1章 幼少年時代

見出し

■ 網文

第1巻 p.1-61 (DK010001k) [ページ画像](#)

天保十一年庚子二月十三日 (1840年)

武蔵国榛沢郡安部領血洗島村ニ生ル。幼名市三郎又栄治郎、幼少時代ノ名乗美雄、後通称ヲ栄一郎名乗ヲ栄一ト改メ、青洲ト号ス。仕官時代一時篤太夫、尋テ篤太郎ト称セシコトアリ。父ハ通称市郎右衛門、名乗美雅、晩香ト号ス。母ハエイ。家ハ世世農ヲ以テ本業トシ、傍ラ養蚕ト製藍トヲ兼ネ営ム。

網文

■ 資料

渋沢栄一伝稿本 第一章・第一頁 [大正八―二年] (DK010001k-0001) [ページ画像](#)

第1巻 p.1

渋沢栄一伝稿本 第一章・第一頁 [大正八―二年]

青洲先生、氏は渋沢、名は栄一、青洲は其号なり、天保十一年二月十三日武蔵国榛沢郡血洗島に生る。血洗島は関東平野を流る、利根川流域の一小村にして、いま大里郡八基村《オホサト》《ヤツモト》に属す。○下略

○『渋沢栄一伝稿本』ハ大正八年ヨリ同十二年ニカケテ竜門社ニ於テ編纂セルモノニシテ、大正十二年九月ノ震災火災ニ資料ノ大部分ヲ焼失セルタメニ中止トナル。上梓セラレタルハ第六章マデナリ。

資料(抜粋)

独自仕様のxml<kobunset>

```
<kobunset xmlns="http://">
  <info>
    <title>渋沢栄一伝記資料 全文公開システム</title>
    <creator>公益財団法人 渋沢栄一記念財団</creator>
    <wordfilename>DK010001k.doc</wordfilename>
    <modificationhistory>
      <data>
        <date>2016/08/17 ; 2018/04/20 ; 2019/04/10 ; 2019/11/14 ; 2020/02/15</date>
        <modifier>saya ; take ; take ; take ; saya</modifier>
        <comment>修正81 ; 修正351 ; 図表1,5修正 ; 図表1再修正 : 参照429,430,431</comment>
      </data>
    </modificationhistory>
  </info>
  <midashiid>11010000000</midashiid>
  <kobunid>DK010001k</kobunid>
  <kobun>
    <ad>1840</ad>
    <jcal>天保十一年庚子二月十三日</jcal>
    <kobuntext>武蔵国樺沢郡安部領血洗島村二生ル。</kobuntext>
  </kobun>
  <other>
    <othertype>0</othertype>
    <otherText></otherText>
  </other>
  <honbun>
    <shiryo>
      <shiryoId>DK010001k-0001</shiryoId>
      <shiryoText>渋沢栄一伝稿本 第一章・第一頁【大正八——二年】<br />
        青洲先生、氏は渋沢、名は栄一、青洲は其号なり、天保十一年二月十三日武蔵
        ○『渋沢栄一伝稿本』ハ大正八年ヨリ同十二年ニカケテ専門社ニ於テ編輯セルモノニシテ、大正十二年九月ノ震災災
        ニ資料ノ大部分ヲ焼失セルタメニ中止トナル。上梓セラレタルハ第六章マデナリ。
      </shiryoText>
    </shiryo>
  </honbun>
  <otherfile>
    <otherData>
      <filename>DK010001f</filename>
      <text>◆◆11010000000◆◆
        ▼▼DK010001f▼▼
        第一編 在郷及比仕官時代 (一)
        天保十一年—明治六年
        ★★01001★
        第一部 在郷時代
        第一章 幼少年時代
        DK010001f 1/1
      </text>
    </otherData>
  </otherfile>
</kobunset>
```

<midashi>

<kobun>

<shiryo>

デジタル版『渋沢栄一伝記資料』

1編 在郷及比仕官時代
1部 在郷時代
1章 幼少年時代

見出し

■ 綱文

第1巻 p.1-61 (DK010001k) ページ画像

天保十一年庚子二月十三日 (1840年)

武蔵国樺沢郡安部領血洗島村二生ル。幼名市郎又栄治郎、幼少時代ノ名乗美雄、後通称ヲ栄一郎名乗ヲ栄一ト改メ、青洲ト号ス。仕官時代一時篤太夫、尋テ篤太郎ト称セシコトアリ。父ハ通称市郎右衛門、名乗美雄、晩香ト号ス。母ハエイ。家ハ世世農ヲ以テ本業トシ、傍ウ養蚕ト製藍トヲ兼ネガム。

■ 資料

渋沢栄一伝稿本 第一章・第一頁【大正八——二年】 (DK010001k-0001)
第1巻 p.1 ページ画像

渋沢栄一伝稿本 第一章・第一頁【大正八——二年】

青洲先生、氏は渋沢、名は栄一、青洲は其号なり、天保十一年二月十三日武蔵国樺沢郡血洗島に生る。血洗島は関東平野を流るゝ利根川流域の一小村にして、いま大里《オホサト》郡八基《ヤツモト》村に属す。○下略

○『渋沢栄一伝稿本』ハ大正八年ヨリ同十二年ニカケテ専門社ニ於テ編輯セルモノニシテ、大正十二年九月ノ震災災ニ資料ノ大部分ヲ焼失セルタメニ中止トナル。上梓セラレタルハ第六章マデナリ。

各巻リンク | 第1巻 目次【綱文】 | 第1巻 (DK010001k) 資料リスト | ▲ページTOP

渋沢栄一伝稿本 第一章・第一四——五頁【大正八——二年】 (DK010001k-0002)
第1巻 p.1-2 ページ画像

渋沢栄一伝稿本 第一章・第一四——五頁【大正八——二年】

先生の名、幼少の時は市郎といひ、又栄治郎と改め、実名を美雄とつけたるは十二才前後の事なりしが、後又伯父渋沢誠室の命名によりて栄一と改め、之を通称となせり。安政三年六月先生十七才の時、尾高監當に就いて名乗を請ひしに、藍香「足下の通称栄一は好き字なり、孔子の言に、吾一以貫之といふことあり、一とは仁なり、仁とは諸の善行の統名なり、且名を以て通称するは古の礼なれば、栄一を名とし、字をば仁栄と日はん。孟子に仁則栄、不仁則辱といへり、足下能く仁の一を榮せんと欲せば、百行爲すに足るものなけん」と、是より栄一名乗とすること後まで渝ることなし。但し通称の時には栄一郎と称し、一橋家出仕の初まて之を用ゐたるが、出仕後平岡四郎邸改めて篤太夫と称せしむ。明治二年静岡藩に仕へし頃、太夫・衛門等の名を改むべき朝命ありしにより、再び改めて篤太郎の称を用ゐたり。幾もなく朝廷仕官の後、正式には源朝臣栄一といひ、ヒデカズと訓まれましたが、後いつとはなく音読して通称とせられしなり。号を青洲といへるは、血洗島の邸宅の後方に沼ありて、此辺の地名を洲上といへるに基けり。(少年の時の時に洲上小屋と自署せり)

族籍は維新の初は静岡県士族なりしに、明治四年六月、東京に移住せし時、請ひて東京府平民に編入し、後華族に列せらる、これらの事

第1巻 p.2 - ページ画像

は後に委しく記すべし。

主な問題点：

- 個別の特徴を再現することに主眼を置いたことで、マスター・テキストに「複数の仕様」
- 将来、さらに新しい仕様が生まれる可能性
- あるいは、既存の仕様の中へ「無理矢理落とし込む」という不合理
- マスター・テキストの長期保存(管理)に影響がでる可能性

→ 『伝記資料』別巻にTEIの適用を検討開始(2018.11)

TEI適用の目的(2019.11)：

- 将来も恒常的に利用できる状態を維持するため。つまり、デジタルデータに汎用性を持たせることで、データの永続性を図る。
- デジタル・テキストについて活用の道を開くことで、渋沢栄一関連情報資源の利用価値向上を目指すとともに、人文学、情報科学、統計学等の分野における貢献を行うため。
- これまで明らかになっていなかった情報の抽出や集約、汎用ソフトウェアによる機械的な解析や可視化を可能にするため。

2-3. 渋谷栄一ダイアリー(2021.04)



<https://shibusawa-dlab.github.io/app1/>

- 令和2年度国立歴史民俗博物館総合資料学奨励研究の研究成果
- 『伝記資料』別巻第1・2（日記、予定表）の本文テキストとページ画像をTEI、IIIFベースで展開
- 本文をまるごと読めるだけでなく、書かれている情報を可視化
- 外部リソースと連携（国文研、Wikipedia）
- 研究成果（プログラム、テキストデータ）は財団のGitHubアカウントからダウンロード可

『伝記資料』別巻第2「日記」

カレンダー表示：月別分布

日別時系列表示

一月一日 日 快晴 風ナクシテ寒カラス

一月一日 日 快晴 風ナクシテ寒カラス

午前八時 起床、直二洗面シテ入浴セ。衣服ヲ整頓シテ 兼子 ト共ニ唐簪ヲ祝賀シ揚ク。次テ事務所員一同及倉庫部其他ノ關係諸員相會シ交年シ始メ祝賀アリ。同族 武、助、正雄、秀雄、敬三 等來リ祝詞ヲ述フ。午前十時 敬三 ト同車 第一銀行ニ二抵リ、先少重徳諸氏ヲ祝賀ス又交換シ尋テ本店及市内支店ノ行員來セ祝賀シ揚ク。數百人ノ會同シ祝詞ヲ讀ミ奉ル。畢テ會堂ニ参集シ一壇ヲ祝詞の訓示ヲ爲ス。銀行長兼子トテ交年シ終ニ意ヲ示キ要旨ヲ指示シ。切實ニ現下ノ世相ヲ指摘シテ深厚ノ訓誨ヲ爲ス。畢テ 番街街ノ家ヲ訪ヒ年賀ヲ祝シ、四時 過帰宅ス。夜、同族各家ノ主婦及子女相集リ新年ヲ賀ス。夜、ヲチオノ放送ヲ聴取ス。又、新聞紙其他ノ書類ヲ閱覽ス。十時 過歸寢ス。
(續外)

■ ■ [img] ■ ■ 此日天気朗晴、風ナクシテ寒氣甚シカラス、顔ル好 天美 日タリ、去年ノ日記ヲ檢スルニ同シク快晴ナリシヨリ本年ノ平寧ヲ予想セシヲ揮々ノ不祥事ニ遭遇シタルニヨリ、本年ハ例ニ拠ラサル事ヲ談話セリ、銀行年賀ノ先後ニ於テ數名ノ年賀客アリ

日記本文

 人物情報



浅野総一郎

浅野 総一郎（浅野總一郎 あさの そういちろう、1848年4月13日（嘉永元年3月10日） - 1930年（昭和5年）11月9日）は日本の実業家。一代で浅野財閥を築いた。（Wikipediaより）

 フォーカス

ネットワーク

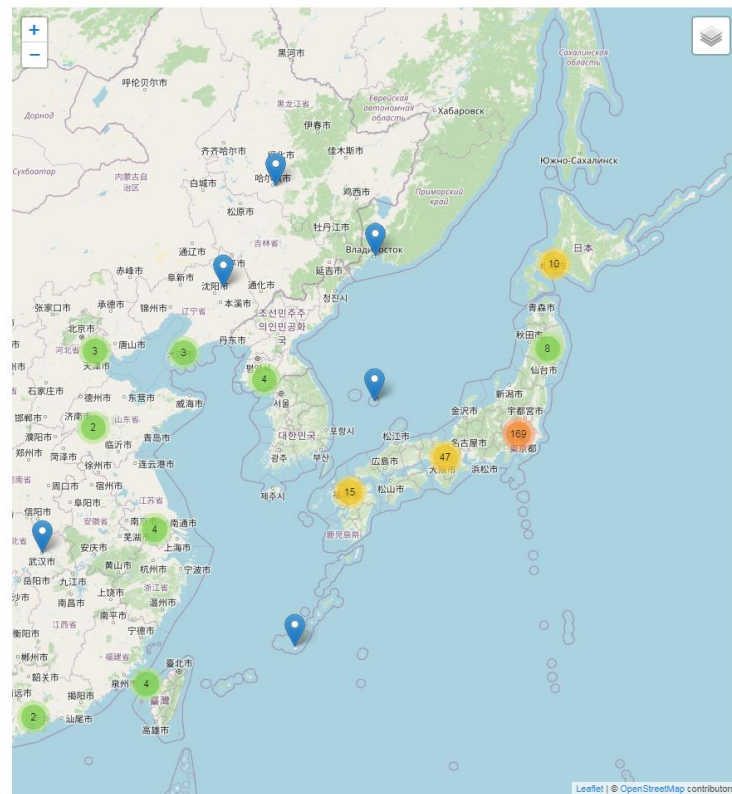
つながりを表すアイテム

■ つながり (13)

- 佐々木
- 浅野
- 井上
- 佐々木典之助
- 安達審忠
- 徳川
- 尾高
- 石井
- 大隈
- 阪谷
- 伊藤
- 篤二
- 竹田政賢

地图

Wikipediaで位置情報が取得できた場所のみを表示しています。



全文
『伝記資料』別巻第1と第2の全文をテキスト＋ページ画像で通覧。

人名・地名
人名・地名を登場頻度でランキングし、ウィキペディア等の情報を参照。

カレンダー
各日付ごとの記事をカレンダー形式で表示、全体/年/月/週…と表示を切替可。

データセット (TEI/XML)
『伝記資料』のデータを一括ダウンロード。再利用可能。

Ngram Viewer
検索キーワードの出現頻度を年代ごとに可視化。正規表現の利用も可。



本文検索
「日記」「集会日時通知表」の記事検索。それぞれを日付順に表示。

原本概要
「日記」原資料の情報（国文研のデジタルアーカイブにリンク）。

ネットワーク
人名の登場頻度などから関係図を自動表示。人物間の新たな繋がりを可視化するためのツール。

地図
登場する地名から地図情報を自動処理、オープン・ストリート・マップに表示。

『渋沢栄一伝記資料』デジタル化の軌跡

こちらをご覧ください：

- 渋沢栄一記念財団情報資源センターが公開するデジタルアーカイブについて(『びぶろす』85・86号 p. 25-29.)
<https://doi.org/10.11501/11456963>
 - デジタル版「実験論語処世談」へ至る過程
- 『渋沢栄一伝記資料』とデジタル化の現在(カレントアウェアネス-E No.427)
<https://current.ndl.go.jp/e2456>
 - 最新の「渋沢栄一フォトグラフ」に至る『伝記資料』デジタル化プロジェクトの経緯
- TEIを用いた『渋沢栄一伝記資料』テキストデータの再構築：「渋沢栄一ダイアリー」公開まで(『人文学のためのテキストデータ構築入門』(文学通信, 2022.07)p.242-259)
 - 「渋沢栄一ダイアリー」で行ったTEIによる構造化