http://shashi.shibusawa.or.jp/about/jsla_bulletin_269_murahashi.pdf
〔渋沢社史データベース｜渋沢社史データベースとは｜社史プロジェクトに関する記事〕

ごぞんじですか？　第 94 回

渋沢社史データベース　―その発想から公開まで　/　村橋勝子

初出：『専門図書館』No.269（専門図書館協議会, 2015 年 1 月）p32-36
[本稿は著者および出版者の了解を得て転載するものです]
掲載：2015 年 10 月 30 日

ごぞんじですか？　第 94 回
渋沢社史データベース　ーその発想から公開まで

1.　はじめに

2014 年 4 月 23 日、社史の内容を探せるわが国初の社史データベース「渋沢社史データベース」（略称・SSD、http://shashi.shibusawa.or.jp）が渋沢栄一記念財団のウェブサイト上で公開となりました。誰でも無料で利用できます。

2.　発意

わが国では近代企業が誕生して間もなくの明治 30 年代に社史の刊行が始まり、これまでに約 6000 社が 1 万 5000 点以上の社史を出しています。

社史には日本経済の成り立ちや企業風土、経営哲学、歴史の教科書では知り得ない産業発展の秘話や製品開発の裏話が盛り込まれていますし、生活密着型の会社の社史は日本の風俗史や生活史まで克明に映しています。めずらしい統計や図も数多く収録されており、総体として社史をみれば、ユニークな産業百科事典でもあるのです。

しかし、なぜか社史は読まれない本の一つ、ほとんど利用もされない本と言われています。百科事典にも載っていないような素晴らしい情報の宝庫であることがほとんど知られておらず、「宝」を探す手段もないからです。

「社史の内容が探せるデータベースがあったら素晴らしい」「企業の壁を越えた社史のデータベースが構築できないか」という話が、2004 年の秋、前年に渋沢栄一記念財団に新設された実業史研究情報センターで持ち上がりました。

3.　データベース構築の準備

とはいえ、データベースの構築、特に、さまざまな資料・情報が多様な形で収録されている社史のデータベースを作るのは容易なことではありません。

そこでセンターでは、まず「社史索引データベース構築検討委員会」という外部委員会を作って、検討することになりました。

委員は経済史・産業史が専門で数多くの社史の執筆もなさっている武田晴人東京大学経済学部教授、索引、データベース、コンピュータに詳しい石坂正男氏（流通経済大学図書館、元・日通総合

研究所）、営業報告書についてはわが国でおそらく一番詳しい横山勝行氏（雄松堂アーカイブズ社長）、社史研究家の村橋勝子（委員長、筆者）の４名で、事業主体と事務局は実業史研究情報センターです。

委員会は2004年12月から、当初は月次開催、その後は随時開催と状況に応じて頻度を変えつつ2009年までに17回開催し、社史を素材にどんなデータベースができるか、著作権の問題はないかなどを徹底的に議論しました。

３．１　データの採録範囲と採録方針

委員会名に“索引”という語が入っていることでもわかるように、このデータベースは社史の全文が見られるものではなく、内容を探すためのものです。社史本文の登載は、著作権の問題があること、作業量や費用が大きすぎることなどから難しいと判断したからです。

委員会で大筋の方向を見定めると、何冊かの社史を使って、実験的な作業を行いました。社史１冊ごとに、「目次」「年表」「資料編」「索引」から細かくデータを採録してみたのです。内容を探すにはこの４つで可能なこと、著作権に関わらないデータであることが理由でした。

また、索引付きの社史を優先的に採録すること、各社史のデータは、記載されているものをそのまま採り、キーワード付与などの知的加工は行わないことなどの方針も決めました。

４．データ入力

おおよそいけそうだということになりましたので、事務局で各編の入力マニュアルを作成、2005年春から本格的にデータ採録・入力作業を開始しました。作業の実働部隊は、１日だけセンターに出向いてマニュアルの説明等を受けた在宅アルバイト（延べ７名）です。社史の現物は宅配便でやりとりし、Excelに入力したデータをUSBに入れて、現物返却に同封してもらいました。

採録対象の社史は、当初は日本の既刊社史すべてを考えていましたが、あまりに膨大で、費用とマンパワーを考えるととても無理ということがわかり、また、渋沢栄一記念財団の事業であること、同財団所蔵の社史を使ってデータ入力を行うことなどから、渋沢栄一が関連した会社の社史に絞り込むことにしました。データベース名に「渋沢」と入っているのは、そのような理由からです。

データベースの方向性と範囲について、揺れ動きながら収束していった感があるこのあたりの議論は、社史へのアクセシビリティを向上させ、ユーザーの使い勝手をよくしたいという作成者達の共通認識がもとになっています。

４．１　データ入力の具体的方法

４つのどのデータもExcelに入力することにしました。すべてのファイルは、１行目A1セルに「書名典拠データ」（『書名』、社名、発行年）を入力、

例：『東レ70年史』（東レ、1976）

あとは、ひたすら「そのまま、細かく」入れてゆきます。

各データの採録方針・方法は次の通りです。

●目次
- 目次データすべてを採録する。
- 本文との相違があっても、目次データをそのまま採る。
- 各章の扉等に出現する詳細目次は渋沢関係会社の社史に限り入力する。
- 目次のない社史は本文から簡略に入力する。
- 目次の階層は社史によって異なるが、そのまま採る。階層は10までとする。
- ページは初ページのみ入力する。

●年表
- 各社史の年表のすべての行を採録対象とする。
- 社史からスキャンしてOCRで読み込み（作業は業者に発注）、その後、校正・編集する。
- スキャンは列（年表の欄）単位とする。

＊年表付社史は736冊で、今回登載した社史の71％でした。

●索引
- OCRで読み込み、その後、校正・編集する。
- 社史により採録の精粗・表記の違いがあるが、

そのまま採る。
・同じ項目が複数ページに出てくる場合もそのまま採る。
＊ちなみに、索引のある社史は140冊で、今回登載した社史の13.5％でした。既刊社史全体で索引のあるのはせいぜい２％程度ですが、今回は索引付き社史から手掛けたため、割合が多くなっています。

●資料編
・社史の巻末等または別冊の「資料編」を採録対象とする。
・Excelに①～⑦と、列を分けて記入する。
①資料ID
②カテゴリー
＊カテゴリーとは「財務・業績」「役員」「売上」などザッとした類に仕分けするためのものですが、固定的なものではなく、随時追加しています。資料編には思いがけないものが収録されていることが少なくないからです。
③内容１（各資料・データ・表などのタイトル）
④内容２（内容１が示す時点・期間）
⑤内容の詳細（表、図、写真など表現方法、内容）
⑥初ページ
⑦備考

４．２　渋沢栄一関連会社社名変遷図作成

前記の通り、データベースには渋沢栄一が関わった会社の社史を中心に登載することにしました。栄一が関わった会社は、以前から「約500社」と言われていますが、実は渋沢栄一記念財団でも正確には把握されていませんでした。

社史は会社名で整理・探索することがほとんどですが、合併・分離や単なる社名変更で改称した会社が少なくありません。「いつ、どんな社名であったかがわからなくてはならない」ことに気が付き、社名の変遷を把握することになりました。

『渋沢栄一伝記資料』（本編全58巻、別巻全10巻）に載っている会社の変遷を一つ一つ調べ、さらに正確を期すために、有価証券報告書、営業報告書、社史、『主要企業の系譜図』（神戸大学編）を典拠資料として、2005年末から４年がかりで作ったのが「渋沢栄一関連会社社名変遷図」（1524社・122図）です。作業は、当初、アルバイトの東大経済学部大学院生が120社分のサンプル調査を手掛け、全体の調査と作図は、このプロジェクトの主力となった門倉百合子事務局員が行いました。

23の業種ごとに作ったこの変遷図は、2008年３月10日から2012年７月３日まで16回にわたり渋沢栄一記念財団のウェブサイトで順次公開しました。

図は、時代の流れに沿って上から下に流れるようになっており、①渋沢栄一が役員を務めたり、株主であったり、助言・援助をするなどの関わりをもった「渋沢栄一関連会社」（521社）は赤、②上記の事業を継承あるいは社名が変遷した「後身会社」（284社）はオレンジ、③「その他の会社」（719社）は青と、社名を記した枠を色分けしています。さらにモノクロでもわかるように、枠にも工夫を施してあり、とても見やすいものです。

＊　＊　＊

栄一が関わった会社が明らかになりましたので、それらの会社の社史をピックアップして、さらに作業を続けました。

入力作業で最も面倒かつ手間と時間がかかるのが資料編のデータです。会社により、社史により、内容も手法も実にさまざまで、包括的に処理できない、いわば「多品種少量生産」で、コンピュータ処理しにくい世界でもあります。財務データにしても、財務諸表の費目名にとても珍しい言葉が登場したりしています。データベースを作る意図が十分にわかっていないと、採り方が粗くなったり、ユニークなデータを取りこぼしてしまいがちなのです。

また、途中からですが、社史に載っている役員の顔写真も採録対象にすることにしました。松下幸之助とか本田宗一郎など著名な経営者の顔は皆よく知っていますが、大企業の役員でも、顔や名前はほとんど知られておらず、写真の入手も難しいからです。

4．3　公開前PR

かなりの量のデータが集積され、データベース化も視野に入った頃、事務局では事前PR的に二つのことをやりました。一つは、作業の終わった社史について、創業の頃に焦点を当てて「社史紹介」を書き、財団のウェブサイトに逐次連載したこと、もう一つは、集積したデータを使って、2012年12月から「インフルエンザ」「ノーベル賞」「ビートルズ」「紅白歌合戦」など普通の言葉で検索した結果にコメントを加えた記事を「おもしろ社史検索」としてセンターのブログに載せたことです。

また、検索事例として、2011年の東日本大震災の直後には、構築したデータを「関東大震災」で検索し、その結果を「社史に見る災害と復興」というカテゴリーでブログに連載しました。

5．「データ」から「データベース」へ

社史データベースプロジェクトの開始から10年、気が付けばデータ数は200万件にもなっていました。しかし、データベースの完成形は見えていません。これを「玉葱のみじん切りが山ほどできたけれど、料理が決まらない」と言っていたものでした。

「データを採録・集積する」ことと「データベースを作る」ことは別のことです。ここで、二つの流れがあり、二つの設計図が必要になります。

委員会では「社史を素材にどのようなデータベースができるか、大筋の方向を見定め、煮詰める」、つまり、大まかな設計図を描くこと。事務局である実業史研究情報センターでは「素材として集積したデータを使って、具体的にどんなデータベースに仕上げるか」、つまり、具体的かつ詳細な設計図を提示することです。

データベース化は専門家の手を借りなくてはなりません。そこで、事務局では国立情報学研究所（NII）高野明彦研究室に相談し、研究室の成果を発信するためのNPO法人連想出版に制作を発注しました。

どういう検索をしたいか、全体の構成や各サイトのデザイン、検索結果の表示方法や並び順など、NII、連想出版との数回にわたる話し合いを基に、連想出版制作のテスト版で推敲を重ね、約２ヵ月半で仕上がりました（サンプルデータを送ってからだと10ヵ月）。

フリーワードの検索結果の業種別割合を示す円グラフや年表項目の検索結果を年別に示す棒グラフなど、渋沢側が意図しなかった付加価値も付いたもので、内容の厚さを伴いつつもスッキリと見やすいものに仕上がっています。

6．公開

発想から企画・準備、そしてデータ入力…と10年、多くの人のアイディアと汗によって社史の内容を検索できるデータベースがようやくできました。公開時の登載社史1035冊（わが国既刊社史の７％）、データ数は140万件強です。

「社名変遷図」はもちろん、実業史研究情報センターが2004年から推進してきた「社史プロジェクト」から生まれた「社史紹介」その他の関連情報や、社史現物を実際に見るため、社史を所蔵・公開している全国の図書館一覧なども載せ、社史を広く立体的に利用できるようになっています。

既刊社史の１割にも満たない社史のデータでどの程度のものができるのか、また、社史の本文が入っていないので、「物足りない中途半端なデータベースになるのでは？」と不安な思いもありましたが、思いついた言葉で検索してみると、意外なことがわかったりして興味深いですし、結構使えます。

6．1　利用のヒント

詳しくは、サイト上の「渋沢社史データベースの使い方」を見ていただきたいのですが、利用について少し説明しましょう。

●特定の会社の社史を見る

会社名で社史を検索すると、目次、索引（索引がある社史のみ）、年表、資料編のすべてを見ることができます（資料編は各資料・データ・表のタイトルなど概要のみ）。

●横断的に探す

なんといっても、これがこのデータベースの特色と強みです。「ワード検索」では自由な言葉で、社史から抽出したデータ全体を網羅的に検索できます。経営用語や生産関係の言葉だけでなく、自然災害、事件などごく普通の言葉や地名でも、その言葉がある社史を探すことができ、書誌情報（基本情報）、目次、索引、年表、資料編それぞれの件数が表示されます。

また、人名は、索引のある社史に収録されている索引データの中から人名を自動抽出した「人名一覧」によって探すことができますが、索引になくても、目次、年表、資料編などにあれば、フリーワードで検索できます。

●年月日で探す

例えば、自分が生まれた日、あるいは100年前の今日はどんな日だったかなど、年月日で見てみると、時代の様相を読み取ることができます。年表付社史736冊のすべての行のデータを採録していますから、経済・経営や企業の本務でないこともわかって、とても面白く興味深い結果が得られます。

7．反響

データベース公開後、企業の社史担当者、学者・研究者、図書館職員、マスコミなどに案内したところ、下記のようなさまざまな感想が寄せられました。

・使い方が分かりやすく書かれていて、難しくて堅い社史が、DBを利用することで身近な資料、とても使いやすいものに感じられる。
・検索方法が容易かつ多様で、便利で楽しい。まだ途中段階とはいえ、充実した内容だと思う。
・分量の多さに驚いた。
・面白い！　何が書いてあるか見出しも見ることができ、社史を読む気になる。
・Googleがやろうとしている全文検索よりもよほど手がかかっている。日本企業の様相をキーワードで串刺しでき、横断検索がかけられるというのがすごい。
・年表、目次等が検索でき、同業種の出来事を横断的に調べたり、特定の会社の歴史を知りたい時など、色々なケースでとても便利。
・会社の沿革だけでなく、切り口を変えれば製品や事業の歴史など、いろいろな資料を引き出すのにも役立ちそう。
・どの本の何ページに出ているか、また、典拠がページまで細かく記載されリンクしており、社史の所蔵先情報なども充実していて、とても親切で有難い。
・社名変遷図はかなりの労作。歴史のある会社を調べる際はとても参考になる。
・自館所蔵の社史の情報チェックに利用できる。

8．今後の計画・課題

一応の形になったとはいえ、このデータベースはまだまだ未完成です。2015年には更に500冊分のデータを追加する予定ですが、完成といえるまでには、もうしばらくはかかりそうです。

それにしても、社史のデータベースを作るのは本当に大変です。年表と索引はある程度機械的な入力ができましたが、ほかは本を見て逐一、手入力ですから、とにかく時間と手間がかかります。

「英文版も欲しい」「すべての社史を採録対象にしてほしい」「索引データベースではなく、社史本文をも登載した全文データベースにしてほしい」など、いろいろな希望も寄せられていますが、課題は１日も早く現在のデータベースを完成させることです。

このデータベースによって、社史の世界が大きく広がることを期待しています。

社史研究家、
社史索引データベース構築検討委員会委員長
村橋　勝子（むらはし　かつこ）