

職員録・官員録データベース（國岡DB）の到達点と問題点

國 岡 啓 子

1. 職員録・官員録データベースとは

『官員録』『職員録』は明治政府の官庁の人員構成をはじめ、個人の職歴追跡や確認のために利用されるもっとも基礎的な史料のひとつに位置する。これらは、1980年代には寺岡寿一氏や朝倉治彦氏による複製本¹が発行され、更に近年では国立国会図書館のデジタルアーカイブや国立公文書館のアーカイブで電子画像が公開された事により、広く簡便に参照することが可能になっている。しかし、これらの史料には索引が存在しないため、その利用方法は従来どおり役職に注目して調べていく方法にならざるを得ず、人物名を出発点に調べることは実質的に出来なかった。

この問題を解決しようと試みたのが、筆者が作成した「職員録・官員録データベース」（以降DBとする）である。このDBでは、明治全期、全省庁に亘って個人の官職異動を追える形となっており、全収録役職は42万件越え、収録人数は2万9千人超え、同時配信されている「帝国大卒業生・資格試験及第者データベース」を加えると、3万8千人余を収録し、両者をあわせると現状では明治期官界の最大の人物データベースとなっている。

鹿児島大学司法政策教育研究センターでは、このDBをExcel形式のまま、更には利用者の編集が可能な状態で無料配信を行っている。『官員録』『職員録』の掲載内容をExcel形式のデータベースとして電子化したことにより、人物名からの検索はもとより、更にはその人名の一部からも検索が可能という利便性や、大量な人物との照らし合わせの場合もExcel関数を利用することによって簡便に結果を得られるなど、『官員録』『職員録』の参照方法や利用目的自体を拡張するものとなった。また収録内容には、一次史料の確認の際に必要な情報も合わせて掲載しており、検証可能である点も強調したい。

¹ 寺岡寿一編『明治初期歴史文献資料集 第一集 明治初期の官員録職員録』第一～六巻（寺岡書洞 昭和52～56年）朝倉治彦編『明治初期官員録・職員録集成』第一～四巻（柏書房・昭和56～57年）

そしてなによりデータベース化したことのメリットは、個人の履歴の追跡よりも、むしろ収録データの統計的分析が圧倒的に容易になる点だろう。特定の省庁部局など、各種の組織役職の就任者を割り出し、その官吏集団の履歴動向を調べるなど、基礎的な情報でありながら作業の膨大さから困難であった分析が、明治全期、全省庁レベルで可能となった点は、今後、様々な観点での研究の広がりを期待できる。

明治期の官僚の統計的分析は、従来、『日本帝国統計年鑑』などにある数値が利用されるのが通常であり、また、利用せざるを得ない状況が続いていた。しかしこれらの数値は原典を提示しておらず、如何なる資料や基準から算出されたものかの情報も少なく、検証に耐えなかった²。これに対してこのDBを利用した分析では、どの時点で誰をどの様にカウントしたかを追跡でき、客観的な基準を持った数値を導きだせる。更には、これらの数値を導き出したデータの処理法を論文とともに提示すれば、誰でも同じ結果を追試し、その処理法が妥当であるかの検討を加えるという新しい研究形態も可能となる。

なお、本稿は鹿児島大学司法政策教育研究センターが配信している（明治期官僚・官職データベースVer.3.0）の内容に対応している。同センターで配信しているのは『官員録』『職員録』を収めたDB（以下「職員録・官員録DB」とする）のほか、帝国大学卒業者や各種資格試験及第者を入力して作成したデータベース（「帝大卒・資格試験DB」とする）を合わせたものである。令和4年11月現在、鹿児島大学法文学部附属司法政策教育研究センターがインターネット上に設けている所定Webページ³で申込みを行うと、ダウンロード用の専用サイトの情報が提供され、そこからDBの利用規約や凡例が付属するExcel版のデータセットが提供されるという配布方法になっている。更に同センターのWebサイトでは「職員録・官員録DB」を基とした検索サイト「明治期官僚・官職データベー

² 小林和幸氏は同氏著「近代初期の日本官僚制」（『世界史の中の帝国と官僚』平田雅博・小名康之編 山川出版 2009年）P161で『内閣制度七十年史』掲載の明治22年文官勅任官の員数に誤って、本来入るべきではない武官が含まれている点を指摘している。

³ 鹿児島大学司法政策教育研究センター「明治期官僚官職データベース（國岡DB）Web版」<https://shokuinroku.ls.kagoshima-u.ac.jp/>で「利用登録」を行う。（2022.12.9アクセス済）

ス Web版」⁴を公開しており、人名、役職、組織などからデータが検索できる。このサイトでは検索した人物の、すべてのシートに存在する職歴が一度に表示され、さらには同じ役職の就任経験者もワンクリックで検索できる、という機能を搭載している。

2. 先行研究と問題設定

戦前期の人事に関するデータベース作成の先行業績としては、名古屋大学法学研究科の「日本研究のための歴史情報」プロジェクトが行った『人事興信録』のデータベース⁵や、韓国歴史情報統合システムによる大韓帝国および朝鮮総督府の職員録のデータベース⁶が挙げられる。上記の組織はいずれもこのデータベースを基とした人物検索サイトをWeb上で公開しているが、Webサイトで検索の対象としているデータベース自体を公開しているわけではない。従って利用者が自分の調べている対象者集団を統計処理したい場合には、検索したデータを一から表ソフトなどに入力する必要があり、多大な労力を投じる必要がある。

これらのデータベースとは別に、個々の研究者が収集したデータの統計分析を行うべく、データベースを作成する手法は枚挙に暇がないほど普及していると思われる。たとえば、川手撰氏は「高文官僚の人事秩序の形成に関する試論」⁷において、農商務省系各省の課長以上の職員をデータベース化し、いわゆる高文官僚の人事慣行が、統計数値にも表れるか否かを分析している。論文中には分析結果の各種表が提示されているが、その基となる同氏作成のデータベースは検討することができない。むろんこれには、そもそも個人ではデータ

⁴ 鹿児島大学司法政策教育研究センター「明治期官僚官職データベース（國岡DB）Web版」<https://shokuinroku.ls.kagoshima-u.ac.jp/>（2022.12.9 アクセス済）

⁵ 名古屋大学法学研究科「『人事興信録』データベース」<https://jahis.law.nagoya-u.ac.jp/who/>（2022.10.18 アクセス済）

これを利用した研究論考としては 増田知子・佐野智也著「近代日本の『人事興信録』（人事興信所）の研究」（1）～（6）（『法制論集』275, 276, 278, 280, 281, 282号）

⁶ 国史編纂委員会「韓国歴史情報統合システムウェブサービス」<https://koreanhistory.or.kr/>（2022.12.9 アクセス済）

⁷ 川手撰著「高文官僚の人事秩序の形成に関する試論」（『都市問題』104(7) P105~113）

ベースを一般に提供する方法自体が存在しない、と考える向きもあろう。しかし現在ではMicrosoft OneDriveなどファイル共有システムの利用によって、こうした個人作成のデータベースでも、一般に提供することは不可能とは言えない。

作成したデータベースを公開するというこの意味は、先に指摘したように論文に提示した数値の検証を可能にする、というだけでなく、同じデータベースを他の論題や視点から再利用することを可能にする。作成したデータベースを公に供しているという点では、藤崎剛氏と東川美和氏が作成し、鹿児島県立図書館へ寄贈された「金録公債証書渡牒データベース」⁸や一橋大学経済研究所の『日本統計年鑑』『帝国統計年鑑』の統計表目次のデータベース⁹なども存在する。こうした研究リソースを活用する上では、その存在、利用方法や作成手続の情報を周知する必要があり、所蔵機関同士の連携の構築が急務となるだろう。

筆者は利用意義の高いデータベースを構築すること自体が知的作業であり、その公開は研究業績として扱われるべき価値があると考え。データベースの作成には多大な資金、時間、労力を必要とする。その構想は単に表を作るということに止まらず、その後の研究の展開を見据え、研究成果の生産可能性や利用可能性を見込んで設計されるものである。さらに利用時の工夫のされ方によっては、データベースの作成者も思い及ばなかった研究を生み出す契機となる可能性さえある。こうした意味で、作成されたデータベースが秘匿され、その一部を覗くことしかできない検索サイトの公開で留められるのは、研究の発展を考えるなら適切な状態ではなく、より広く公開されるべきものであろう。

さて、今回公開しているDBであるが、これを学術的に利用する際には、データベース自体への検討や批判が必要になってくる。しかし、その評価のポイントは何で、どう判断すれば妥当なのか、データベース自体の作成課題やクリアすべき条件を論題とした研究は、管見にして知らない。佐野智也氏はOCR (Optical Character Recognition) で読み取ったテキストデータを、統計分析処理が可能なデータベースに構築する手法として、テキスト処理やテキストマイニ

⁸ 南日本新聞 2021年6月7日 朝刊 P13。

⁹ 一橋大学経済研究所「人文学・社会科学データインフラストラクチャー構築推進事業」<https://d-infra.ier.hit-u.ac.jp/Japanese/> (2022.12.9 アクセス済)

ングなどコンピューター処理を利用された事を発表している¹⁰。しかし一般の利用者を想定したデータベースが、如何なる注意点を基に設計されるべきであるかという視点は存在しない。

一方近年、国立国会図書館の次世代デジタルライブラリー公開¹¹などのおかげで、多くの史料のテキストデータが、個々にOCRで読み取りをしなくとも入手できる状況が整ってきている。これによって今後は公開用データベースの作成も活性化すると見込まれる。本稿では「職員録・官員録DB」の作成過程で明らかになった作成課題やその対処法などを論述し、データベースを作成する際には如何なる注意点、工夫が必要となるのかを明らかにしていく。データベース批判のための評価基準の議論は、まずはその論点の洗い出しから始めなくてはならないが、職員録・官員録DBの作成過程の紹介が、その一助となることを目指す。

3. データベース化に潜む問題

ここで最初に注目すべきは、史料をデータベース化するという問題において、利用者が原典をデータベース化した内容で利用する場合と、原典自体を直接利用した場合との違い、つまり史料を電子化した場合に発生する制限、限界とは如何なるものになるのか、という視点であろう。これはデジタルアーカイブ学会などで、史料をデジタル画像化した場合などにも指摘されている論題である。この議論には多角的な視点からのアプローチを必要とし、軽々には結論できないものと思われるが、本稿ではその論題に対して、原典へのアクセス確保と原典への史料批判の必要を挙げておきたい。

データベースを利用する際、収録された情報が妥当なものであるか、入力者による誤謬やバイアスがかかっていないかなどを判断すべく、入力データの元となる一次史料へアクセスする必要性があることは言をまたないだろう。原典にアクセスする際に必要な情報の提示は、公開・非公開の別なく、データベー

¹⁰ 佐野智也・増田知子「データサイエンス時代の歴史情報基盤の構築」(『デジタルアーカイブ学会誌』 Vol.3 No.2 P147～150)

¹¹ 国立国会図書館「次世代デジタルライブラリー」<https://lab.ndl.go.jp/dl/> (2022.12.9 アクセス済)

ス作成上必須情報と考えるべきである。

職員録・官員録DBのデータ収録に利用した原典・複製本は、明治19年末以降については、官報局・印刷局刊『職員録』（国立国会図書館所蔵本）である。これ以前、太政官制度期の民間刊行の『官員録』『職員録』については、複製本で寺岡寿一編『明治初期歴史文獻資料集 第一集 明治初期の官員録職員録』第一巻、第二巻、改訂第二版¹² 第三～五巻、第一版（寺岡書洞 昭和52～56年）を利用している。（以降、官報局・印刷局発行本を官刊『職員録』、寺岡氏による前掲本を「寺岡複製本」とする）。

つまり、内閣制度期に関しては、当時に刊行された印刷物を直接原典として利用のに対して、それ以前は複製本を利用しているわけである。これは筆者がDBの作成を開始した当時では、寺岡複製本がもっとも利用しやすかった為である。しかし現在では国立国会図書館のデジタルコレクションや国立公文書館のデジタルアーカイブなどで、民間刊行の『官員録』『職員録』が手軽に多数閲覧できる一方、寺岡複製本は手に入りにくくなっている。図書館所蔵の『官員録』『職員録』は、寺岡複製本の底本とは同版ではあっても、修正等で内容に相違が生じているものが多い。とはいえ奏任官以上のみを採録したDBのデータと比べた場合は、同内容の記載が確認できるものが多い。DBの内容の確認手段として、Webでの閲覧が可能な本はどの程度利用できるのか、を比較してみた。その結果は「寺岡複製本とWeb閲覧の官員録・職員録」として鹿児島大学司法政策教育研究センターで配信しているので参照されたい¹³。なお、Web閲覧可能な『官員録』『職員録』でDB内容を確認する際には、寺岡複製本の底本自体に振られている頁数、つまり丁番も必要となるので、これもDBのVer3.0から収録し提示していることを付言しておく。

なおここで当然のことながら、DBとの関係では、明治期に発行された印刷物を原典（元データ）とすることと、複製本を原典とすることの違いに注意する必要がある。複製本では一度編者の編纂作業を経るわけだが、DBで利用し

¹² 寺岡複製本の第一～二巻の改訂第二版は第一版とでは収録底本に変動があるので、注意されたい。改訂版での収録底本の変更については、第一、二巻の（序）（官員録・職員録 第一巻の改訂再販にあたって）に詳しい。

¹³ 前掲注13参照。

た寺岡複製本の場合、その刊行時期の推定に疑問が持たれる、或いは底本提示が不完全な可能性がある、などの問題が発生している。特に慶応4年8月刊行としている村上勘兵衛・井上治兵衛刊『太政御職明鑑』については、刊行時期が7月の疑いがあるが、その判断にはDBでは情報を収録していない原典の刊記の情報が必要となっている。つまり、史料をデータベース化したことによって、収録から漏れた情報への気配りを怠ると、原典への史料批判が不十分になる可能性が懸念されるわけである。

4. 原典への史料批判

本節ではDBでの利用だけでは気づきにくい、原典の問題点について論述していく。寺岡複製本に対しては、底本の明示の問題や原典の発行年月の推定の問題を、官刊『職員録』を含めた原典全体に対しては、掲載の欠落する組織の存在、掲載内容の正確性の問題などを指摘していく。

1) 複製版の底本について

この時期の印刷物には、書名や編者、発行元など奥付内容が同一でありながら、内容には変化が生じているものが多く存在する。同一の本として扱われる印刷物の印刷の最中にも修正を加え続けた結果であるが、民間刊行の『官員録』『職員録』にも、こうした内容の違う同一書名、同一奥付のものが多く存在する。この為、複製本の作成の場合には底本が確定できるように、その所在の明示が必要となる¹⁴。寺岡氏による複製本の場合は、多くは寺岡氏自身の所蔵本を底本としており、他図書館などの所蔵本を利用した場合はそれが明記されている¹⁵。しかし明示されたもの以外にも、例えば明治15年5月版の『改正官員録』（彦根正三編 博公書院発行）は国立国会図書館所蔵本と思われ、すべて

¹⁴ 藤實久美子著「本報告書で使用する書誌用語について」『「太政官日誌」を対象にした史料学の構築と戊辰戦争期の社会文化論に関する学際的研究』参照。 https://drive.google.com/file/d/0B_r7GuKRp-HCdnYxME85Q1RZUGc/edit?resourcekey=0-vbIzpMAhJXwWQyjeKp8GA（2022.3.17アクセス済）

¹⁵ 寺岡氏所蔵本以外の底本としては、明治5年2月、6年1月「袖珍官員録」（須原屋茂兵衛・和泉屋市兵衛刊）の2冊が国立公文書館所蔵本（寺岡前掲本二巻P18参照）、明治16年12月「改正官員録」（彦根正三編 博公書院発行）が東京大学明治新聞雑誌文庫所蔵本、明治17年5月「改正官員録」（彦根正三編 博公書院発行）が国立公文書館所蔵本（寺岡前掲本五巻P8参照）、と明記されている。

の底本が明示されているのかには疑問が残る¹⁶。

2) 発行年月の推定への疑問

寺岡複製本で原典の発行年月の推定には疑問のある部分があり、ここでは3つの例を挙げておきたい。

一つ目の例は、前節でふれた村上勘兵衛・井上治兵衛刊『太政御職明鑑』である。寺岡氏はこれを慶応4年8月刊としているが¹⁷、これは「7月改」の可能性が高い。寺岡氏の八月刊行の推定は、収録冊子の冒頭の序に「慶応四年八月」とある点を根拠としたと思われるが、この冒頭の序は、『京都府役鑑』のもので、京都府出仕者を掲載した別の冊子のものである。なぜ別の冊子の序がここに混在するのかわからないが、寺岡複製本収録の底本である版の特徴として、24丁が重丁となっている、つまり24丁目の次丁、本来25丁となるべきところが（又24丁）という丁番号で、丁数24が二度使用されている点が挙げられる¹⁸。朝倉治彦著「太政御職名鑑・京都御役鑑の板次」¹⁹によれば、この特徴が確認されている版は、『太政御職明鑑』三十丁本、村上勘兵衛・井上治兵衛刊に特有のもので、同氏は同版本である福井県立図書館松平文庫所蔵本に付属していた袋から、「慶応四戊辰七月改」と推定している。

二つ目の例は、明治2年5月『官員録』（和泉屋市兵衛・須原屋茂兵衛刊）である。これは内容が5月刊行のものよりも、3月刊行のものの方が近い。国立公文書館には明治2年の和泉屋・須原屋版『官員録』として、十一月を除いて三月版から十二月版までの各月の版が揃っており、そのうち、五月版は二種類が所蔵されている²⁰。これらを比較してみると、三月版の版木が修正を加え

¹⁶ 同一本と判断した根拠としては、蔵書印の比較を行った。寺岡複製本第四巻P384の蔵書印と『改正官員録』（彦根正三編 博公書院発行）国立国会図書館所蔵 請求番号（14.1-8）総目録一頁目の蔵書印である。

¹⁷ 寺岡前掲本第一巻P69参照。

¹⁸ 更に本来26丁である筈の次々丁の番号は25となり、これによって最終丁番号は29丁でありながら、総枚数は30枚となっている。

¹⁹ 朝倉治彦著「太政御職名鑑・京都御役鑑の板次」（『図書館と出版文化 弥吉光長先生喜寿記念論文集』弥吉光長先生喜寿記念会編）1977年P3～12）P6参照。

²⁰ 国立公文書館所蔵 和泉屋・須原屋版「改正新刊 官員録 三月改」（明治2年3月）請求番号（職A00007100）和泉屋・須原屋版「改正新刊 官員録 全」（同年4月）（職A00008100）和泉屋・須原屋版「官員録 五月改 上」（同年5月）（職A00010100）和泉屋・須原屋版「官員録 五月改 下」（同年5月）（職A00011100）和泉屋・須原屋版「官員録 五月改 全」（職A00009100）を参照。

ながら、四月版から五月版の『職員録・明治二年五月・官員録改（上）（下）』まで利用されたと推測できる。その後発行された『職員録・明治二年五月・官員録全改』では版木を新しくし、同時に大幅な情報更新として、同年3月設置の待詔院、教導局や明治2年4月設置の民部官などが新しく掲載されている。これら国立公文書館所蔵本には寺岡複製本と同一内容のものは存在しないが、内容的には三月版が最も近い。なお、同年3月6日に議定就任の徳川慶勝が掲載されているので、発行推定が3月より早まる可能性はない²¹。

三つ目の例は、原本の発行年月の推定の相違ではなく、寺岡複製本の目次の記載漏れと思われるものである。同書に明治18年7月『改正官員録』（彦根正三編 博公書院発行）とされるものがあるが、7月であるのは地方編の『改正官員録 下』（明治18年6月15日届 同年7月2日刊）だけと思われる。中央官職に関しては、寺岡複製本の原典には外務省の冒頭に「外務省四月二十日調」、また海軍省冒頭には「海軍省四月十日調正」の記載があるが²²、上記七月版の上巻『改正官員録 上』（明治18年6月15日届 同年7月2日刊）では「外務省六月十九日調」「海軍省六月十一日調正」と日付が違っている。寺岡複製本と同じ記述を国立国会図書館所蔵本で調べてみると、彦根正三編『改正官員録 上』明治18年4月15日届 同年5月2日刊 国立国会図書館所蔵 請求番号（14.1-8）と一致した。以上より、筆者は、中央官職の部分は同年5月刊である点を、目次に記載漏らしたのではないかと推測している。

さらに寺岡複製本には、作成時に発生したと思われる落丁、乱丁が多く含まれる。これらの存在は同書の閲覧だけでは判定できず、人事内容さえも読み間違える可能性を含んでいる。この問題については訂正一覧表を作成し、「寺岡複製本の落丁・乱丁」として、鹿児島大学司法政策教育研究センターから配信しているので参照されたい²³。

3) 掲載の欠落する組織

掲載欠落として、寺岡複製本には明治2年12月版から同6年1月版までは、各府県の地方官の部がない。これについては、全県をまとめて記載した原本

²¹ 寺岡前掲本第一巻P92参照。

²² 外務省の冒頭は寺岡本第六巻P32、海軍省冒頭は同書P101参照。

²³ 前掲注13参照。

が発行されていないためと寺岡複製本の「序」に記されている²⁴。従って2022年現在の現状では、当DBでもこの時期の地方官が欠落している。この部分の情報を補う手段としては、国立公文書館所蔵の「任解日録」（請求番号 187-0130）を利用することがひとつの方法であり、現在これをデータベース化する作業を進めている。

また、明治4年版に本来存在するはずの土木寮が現れないという問題が存在することがわかっている。これは土木寮が同年10月8日に工部省より大蔵省に事務移管された事と、明治4年から明治6年には、版元が各省ごとの『袖珍官員録』しか発行しなかった影響である。つまり寺岡複製本で明治4年には各省ごとの冊子が収集されているのだが、各省での調査時期が異なっている結果、工部省（11/20調べ）、大蔵省（10/28調べ）の両省に土木寮が存在しない、という事態となっているのである²⁵。

逆に欠落しているかに見えるが、実態は違う例として、官刊『職員録』では明治19年版の乙巻²⁶や24年版の甲巻、31年版甲乙巻など、発行年次自体が抜けている巻の場合がある。しかしこれらの時期は、調査時期を年末から年頭へ移動させる、或いは27年版のように年頭と年末の調査結果、二種類を発行させることなどによって間隔を補っており、実質的に一年おきの官員調査、発行という原則はほぼ守られたと言える。ただ、一年おきの調査では細かな人事異動や政局変動の様子は追えず、例えば隈板内閣期の人事変動の状況に関して言えば、明治30年11月1日調査の次が、同32年2月1日調査に飛んでしまい、この間の異動がDBには表れにくくなっている。

さらに官刊『職員録』の記載は、該当する役職の職制が制定された時点から開始となる例が多い。例えば官国幣神社の官司の様に、職制制定（明治35年2月8日の勅令27号官国幣社職制制定）を受け、同年の『職員録』から記載されるのが一般的である。これ以前の官司達の任免に関しては『職員録』には

²⁴ 寺岡前掲本第二巻P18参照。

²⁵ 寺岡前掲本第一巻P366～423参照。本来であれば、十月二十八日調べの大蔵省に土木寮が編入されるべきであろうが、情報更新が遅延した結果と思われる。

²⁶ 明治19年12月31日調査、同20年4月21日出版の内閣官報局刊『職員録』乙巻を国立国会図書館は『職員録 明治19年（乙）』と標題をつけている。これはCiNii Booksで『職員録』明治20年 乙 としている書誌と同一の本である。

表れてこない。これはつまり、ある個人の職歴で、奏任官以前の職歴がDBに収録されていないのは当然としても、同一の役職に就任していながらも、職制の有無によって、『職員録』への記載が変化した例が存在する、ということの意味する。『職員録』の史料的特徴として注意が必要だろう。

4) 掲載内容の正確性

原典についての最後の問題点として、掲載内容の正確性が挙げられる。民間刊行の『官員録』『職員録』は誤字誤植が非常に激しく、奏任官以上に範囲を限定しても、その人名表記を始め、役職名、本籍地などの記述の信頼性には疑問が残る。これに比べれば官刊『職員録』となってからの精度はある程度上がったと言えるが、情報更新の精度に関しては明治期全体を通じて十分とは言えない。

例えば、DBでは「安立綱之」は明治32年に佐賀県内務部長と長崎県警部長を兼任しているが²⁷、同一人物が二つの県の職員を同時に兼ねることは考えにくい。おそらく長崎県の方の免官情報が反映されなかった結果と思われる。毎月発行された民間刊『職員録』や『官員録』では、情報更新が正確性に欠けている例は枚挙に暇がないが、官刊『職員録』でも一年という発行期間中の人事異動が、本来そうであるべき翌号の内容に反映されず、次々号以下に持ち越されてしまうという例が散見される²⁸。

5. 人名の取り扱い

『職員録』『官員録』などDBで取り扱っている史料は、人名に対応する情報を収録したもので、人名が収録データの大きな柱となる。しかしデータベースにおける人名の取り扱いには、改姓改名、漢字表記、同姓同名の別人を如何に処理するかなど問題がある。本節ではこれらの問題にDBではいかなる対処を

²⁷ 地方官ID [200252] 安立 綱之 『職員録』明治32年乙巻P74, 363参照。

²⁸ 同様の例として地方官ID [201174] 岩崎 奇一が明治24年に愛知県収税長と和歌山県収税長を兼ねているように記載されている。(『職員録』明治24年乙巻P134, 287) また地方官ID [206057] 戸田 恒太郎が、明治33年に茨城県内務部長と福井県内務部長を兼任しているよう記載されている。(『職員録』明治33年乙巻P127, 269) これは岩崎の場合は和歌山県側の免官情報が、戸田の場合は福井県側の免官情報が反映されていない結果と思われる。

施したかを明らかにする。以下、下線つきは列の項目名を、波線付きはシート名、セル内のデータは[]付で示す。

1) 改姓改名、読み仮名について

明治期の氏名は改姓改名が頻繁で、一個人に複数の氏名が存在する。更に各個人の正確な読み仮名、漢字の表記を割り出す事はかなり困難という問題が存在する。漢字表記に関しては戸籍に登録されたものが正式であるわけだが、現代でも新字や旧字を利用して、本人自身が自分の氏名を複数の表記で書き表し、これを誤記とは捉えない例は多いだろう。明治前半期にはその振れ幅が大きく、当時の編集者は、例えば(二郎)(次郎)(治郎)などの違いは誤植とは捉えていない節もある。

こうした原典での人名の取り扱い状況に対して、DBでは利用者が必要な人名を確実にヒットできることが最優先であると考えた。まず原典記載の表記の中で、なるべく改姓改名後の氏名を、人名検索の中心となる文官・地方官・武官人名(以下人名と略す)の列に採録した²⁹。原典に現れる他の表記は、例えば[徳大寺大納言]とか[伊藤五位]などの官職名を含めて、すべて別名1の列に羅列した。この中には誤植と思われるものも多く含まれる。また原典には表れないが、他の人名辞典などから同一人物の別名と判別できた氏名は、別名2に入力している。

これに比して「帝大卒・資格試験DB」では原典での表記を採録した列である、人名(原文)を設けた。これは一つの原典に同一人物が何度も記載されないため、一人の収録者の原文での表記が一つに限定できた結果である。検索用にはもう一列DB用氏名を作成し「職員録・官員録DB」と同一人物が同一表記になるように処理した氏名を入力し、「職員録・官員録DB」と連携ができる形をとっている。どちらのDBも複数のシートを擁する為、どのシートに求める人物が存在するかの情報が必要になる。こうした検索を容易にすべく人名検索シートを作成し、両DBの各シートの人名、別名1、別名2を網羅した。

²⁹ 高村光雲など各種の雅号は、基本的に本名・諱を人名の列に入力し、雅号は別名2へ入力した。しかし高村光雲など、雅号で原典に記載があった場合は、雅号を氏名に採録している。

また、DB内の氏名の読み仮名は本来のものとは全く違う。例えば（河野）は全員（コウノ）、（上村）は全員（ウエムラ）となっている。これは同一人物を違う読み仮名で入力した結果、修正時にチェックミスをしてしまって泣き別れを発生させてしまい、別人として処理されてしまう事を避けようとした結果である。この事情から、DBでは読み仮名をかなり恣意的に変更している。この結果、シートを人名の五十音順に並べて、該当者を探すとミスを誘発するという問題が生じている。

2) 漢字の表記問題

漢字表記の問題とは旧字新字、異体字などをどう処理するかという問題である。鹿児島大学司法政策教育研究センターのWebから提供している検索サイト「明治期官僚・官職データベースWeb版」では漢字統合などの処理により、旧字新字、異体字で検索をかけても問題ない。しかしExcel上の検索エンジンでは、漢字の旧字、新字、異体字などをすべて別文字として扱う。そのため、DB作成側と利用者側双方が使用する漢字を、親字（漢和辞典に見出しとして採録されている漢字）にするというルールを設けた。この場合、斎藤と齋藤、脩と修、岡と丘は別文字となり、灑、熙は熙、昂は昂、顚は顚、すべて後者が親字となる。阪と坂は本来で言えば別文字であるが、原典での混用があまりに多い為（坂）に統一している。

PC上に存在しない漢字、つまり外字の問題はデータベース作成に留まらず、日本語をベースとしたすべての記録に必要な問題で、広く一般に利用できる外字データの公開が強く望まれている。既に作成されている外字データとしては、例えば国立国会図書館や官報を発行する国立印刷局などは、それぞれが独自に大量のデータを作成しているが、これを一般にできるように公開はしていない。この問題は日本語利用者全体への利益となる課題であり早期の解決を願う。DBでは外字が必要な場合には、なるべく旁が近い漢字に（'）を付けて代用した³⁰。この処置は検索においては意味をなさないが、利用者に原文が外字であることを伝えているのである。

アルファベット表記の判明した外国人はこれを人名の列に収録したが、原典

³⁰（'）付の漢字、親文字の例などはDB付属の凡例シートを参照。

のカタカナ表記をそのまま収録している例もある。『職員録』に現れるカタカナ表記は、例えば [P. E. J. Hemelryk] が [ピー・イー・ゼー・ヘソルリック]³¹ になっているなど、アルファベット表記からは連想しにくい形になっているものが多い。カタカナ表記では利用者の検索にヒットするか疑問があるが、逆に任免辞令などはカタカナ表記であり³²、これらの史料から検索するにはカタカナ表記が必要となる。人名の項目をアルファベット表記にした場合、カタカナ表記は項目別名 1 に残した。

名前の収録にあたって、原典の情報を忠実に採取するという観点にたてば、原文通りの表記を漢字の表記もそのまま収録した列が存在するべきであろう。「職員録・官員録DB」の場合の様に一人に対して多数の表記が現れるのであれば、名前の表記とその出典だけを一覧できるシートを作成するのが望ましい。原文通りの収録名と検索用の氏名は別の列として作成し、検索用氏名だけ表記漢字を親字にするなどの限定をするべきだ。しかし旧字・新字・異体字を取り混ぜた表記、更に誤植を取り混ぜた違い、例えば（次郎）と（二郎）の違いなどをすべて網羅するとなると、かなり大規模なデータとなる。その一方でこれを如何に利用するのか、各種人名事典など他の文献においても人名の新字、旧字の表記は様々であり、上記のような原典の氏名表記一覧シートを作成して、現状以上に利用者の利便性が向上するとは思えないという観点から、この作業は実行しなかった。

3) 同一人物性の判定

同姓同名の別人、混同の多い人名（例えば吉田次郎と吉田二郎）には、名前の後に番号を振って区別した。この番号は「官員録・職員録DB」と「帝国大卒、資格試験及第者DB」の全シートに共通しており、同一番号は同一人物を意味する。DBには収録人物に関して各種人名事典の記載頁をはじめ、学歴、出身地、没年月などを提示しているが、これらの情報はいずれもDB収録者と参考文献の記載者を同一人物であると、筆者が判断した結果で誤謬に対しては筆者に責がある。さらに言えばDBで一行にまとめられた職歴も、筆者がこれらを同一

³¹ 文官ID [100035] 『職員録』明治36年版P51参照。

³² 例えば文官人物ID102779 [エドアール・フハン・デル・ストレーテン] は官報の明治30年6月19日P266にカタカナ名で辞令が記載されている。

人物のものとして判断したことを意味する。

この同一人物であるか否かを検討する作業は、「職員録・官員録DB」では非常に重要な問題で、本稿では「同一人物性の判定」と表現している。原典の記載には、誤植や同姓同名の別人も混在する状況で、どこまでが誤植でどこからは別人であるのか、『職員録』の役職だけでなく、帝国大卒業者や資格試験合格者とも同一人物であるかの検討が必要となる。

判断の多くは『明治過去帳』³³など各種人名辞典の記載情報によるもので、例えば地方官ID [201535] の[遠藤 達1]は『明治過去帳』P356記載の通り、大阪府少書記官から福井県書記官や広島県書記官、北海道庁書記官に就任しており、同一人物と考えられる。しかしこの人物は明治25年7月に死亡しており、明治34年以降に『職員録』の記載が現れる文官ID [102857] の[遠藤 達2]とは別人と判断した³⁴。

帝大卒業者との検討の場合は『学士會會員氏名録』³⁵などを利用し、同書の職業がDBと一致する場合は同一人物と判断した。これと比較して各人名事典の情報でも判断しきれなかった者の学歴は、参考学歴の諸列に提示しているが、判断しきれないとは具体的にどの様な場合であるかを例示しよう。

文官ID [100077] [青 篤世]は『学士會會員氏名録 明治44年調』P8に職業は載っていないが、住所が「全州区裁判所内」とあり、DBの同年役職である[朝鮮総督府全州区裁判所判事]³⁶と内容が一致する。つまり職業欄は空欄でも、住所から職業が割り出せる場合は同一人物と判断した。これに比して文官ID [109015] [杉山 直治郎]は『学士會會員氏名録 明治44年調』P254に職業は空欄で、住所は「長崎県西彼杵郡長崎村」とある。DBでは明治42年から[長崎高等商業学校教授]³⁷となっており長崎県の住所が任地と一致するが、この場合の住所のみでは役職まで確定できないと判断し、同氏の学歴は参考学歴と

³³ 大植二郎編『明治過去帳』（東京美術刊・昭和46年）

³⁴ 『職員録』明治34年甲巻P683,698参照。また各人の死亡年月は個人特定に大いに役立つ。DBのVer.3.0にはこの没年月の項目を追加している。

³⁵ 『学士會會員氏名録 明治43年調』（学士會事務所編 1910年）「学士會月報」第273号付録 『学士會會員氏名録 明治44年調』（学士會事務所編 1911年）「学士會月報」第285号付録

³⁶ 『職員録』明治44年版P885。

³⁷ 『職員録』明治42年甲巻P610 同43年甲巻P601 同44年甲巻P608参照。

して扱った。

DBの役職情報のみから、別人の可能性ありとした例も多い。地方官ID [204817] [鈴木 貞太郎 1] と文官ID [109083] [鈴木 貞太郎 2] は、明治43年版で前者は新潟県技師、後者は学習院教授と記載があり³⁸、この二つの役職を一人が兼任するのは不可能と考え、別人と判断した。また、対して判断しきれなかった場合の例、文官ID [113662] [林 省三1] と地方官ID [207242] [林 省三 2] も挙げておきたい。前者は明治20年に[大津始審裁判所判事]³⁹に就任し、後者は同42～44年に[神奈川県警視]⁴⁰となっている。判事と警察官吏という任用資格の大きく違う職種である点、二十年間の空白などを考慮すると別人とも思えるが、(別人の可能性あり)で断定は留保すべきと判断した⁴¹。この場合の様に同一人物と断定できない場合は、原則別人として処理している。

名前は同一でないが同じ役職を続けているという情報から、同一人物の可能性があると判断した場合もある。例えば文官ID [109018] [杉山 平治郎] と文官ID [109014] [杉山 高行] は前者が明治25年まで[岐阜区裁判所判事]に就任しており⁴²、同26、27年の同職に[杉山 高行]として記載されている。こうした場合や先の同一人物と断定しきれなかったケースは両者の役職を参照してはならないので、他シートと共通IDに?付の同じ番号[文文-6?]が振っており、人名備考の項にも「○○と同一人物か」等の記載を施した。

次に人物IDにおける課題について考察してみる。データベースには各個人に人物を特定するIDコードが必要である事は当然として、DBでは同一人物の役職が複数のシートに分散している場合、他シートと共通IDに全シートで同一のコード番号を付けている。こうした人物のIDコードは本来であれば、この他シートと共通IDに統一するべきであるが、現在DBVer.3.0では各シートご

³⁸ 『職員録』明治43年乙巻P147 同年甲巻P91参照。

³⁹ 『職員録』明治20年甲巻P218参照。

⁴⁰ 『職員録』明治42年乙巻P93、同43年乙巻P97、同44年乙巻P103参照。

⁴¹ 林 省三 1 は明治20年に彦根支庁判事補から判事へ登用されているが、これは翌年の文官試験試験補及見習規則が施行される直前に、駆け込み的に司法省部内者を判事へ大量登用された時期である。この時期の司法省部内から登用された判事には法曹界を早くに退職し、他省庁地方官などへ転出している例も存在する。従って林 省三 1 と林 省三 2 は同一人物の可能性も高いとも言える。

⁴² 『職員録』明治25年甲巻P275参照。同26年甲巻P287、同27年甲巻P274参照。

とに別の人物IDが存在している。実例をあげれば、[木梨 精一郎]は文官人物IDでは[105668]、地方官人物IDでは[203124]、武官人物IDでは[300686]、更に他シートと共通IDは三シート共通で[文地武一1]となっている。この一人に対して、複数のID番号が存在する状態は誤操作を招きやすいが、これは編集上の都合から生じた結果である。つまり新しい文献や史料の調査を追加する度、別人で処理していた者同士が同一人物と判明したり、またはこの逆のパターンなどが多々生じる。一人の人物コードを一つに統一してしまうと、その都度に必要となるコード番号の変更方法が複雑になり、間違いを誘発しやすかった。同一人物の人物IDを一つに統一するのは同一シート内に限り、他シートとの照合用には他シートと共通IDを利用する方式が、簡便でミスを抑制の効果が高かった。Excel版の公開時に人物IDコードを一人に一つに付け替えることも考慮したが、バージョンを重ねる過程で人物を特定するIDコードが度々変更になるよりは、現状の一人に複数のIDコードが存在する方を採択した。

本節で述べてきた同一人物性の判定は、人名を情報基盤とするデータベース作成の場合には必須となる作業である。例えば今後、官報の任免情報のテキストデータをプログラミング処理で統計処理の可能なデータベースに編成できたと仮定してみよう。任免データを人名で名寄せした後に必要となる作業は、一つの名前に寄せられた任免データがどこまで同一人物のものか、の判定である。逆に改名改姓や誤植によって、本来は同一人物の任免データが別の名前の項に表示されるケースも出てくる。誤植であるかの判断などには、AIの進化などが期待できるという見通しもあるが、同一人物性の判定は最終的には専門家のチェックが外せないと思われる。

6. データベースの構成

データベースの構成はデータの利用目的に合わせなくてはならない。また利用者がその構成や収録、入力など編集のルールを理解しやすい、という必要もあるだろう。特に一般の利用の為に公開しているDBの場合、利用者が必要に応じてデータを編成することを前提としているが、その際に必要なデータの遺漏などを発生させない様、収録や入力時のルールの明示は必須となる。

DBでは、利用目的の第一は人物レファレンスとした。編集のイメージは『職

員録』の人名索引である。従って該当者の一つの役職が重複して記載されている場合や、奏任官が判任官の役職を兼任した場合も収録対象に含めた。収録や編集のルールに関してはこのように既存のもので表現し、作成者と利用者が共通の認識が持てるものが望ましいと考えた。利用者が重複した役職を不要と判断した場合は、役職備考の列にある[二重に記載あり]などの記載から、該当項目を検出し削除することができる。

役職の配列に関しては「個人の履歴を一覧で理解できる形に配置する」という目的で、同一役職は同一行に連続させ、行数の節約から一行に複数の省庁の役職をまとめることも行っている。こうした構成の結果、役職履歴が複数行に亘る対象者の場合、その複数行の中には多数の空欄も混じるが、その空欄がどの行に生じるかは個々別々になってしまった。つまり原典に収録すべきデータが存在する場合、必ず入力されるメインの行というものはなく、該当者のすべての行を調べなくては、その年の『職員録』に記載があるか否かは判断できない構成となっている。

これはDBの第二の利用目的である、奏任官以上の統計分析の際に障害となる。例えば必要な役職の就任経験者を検出して、該当する行を他のシートに抽出した場合、残された行に兼任した役職や他の組織に転出したデータなどが取り残される可能性がある。逆に一つの行に複数の省での役職が混在しているので、必要な組織以外の役職が混ざりこむ処理となる可能性も存在する。

ここで明らかになるのは、個人の役職履歴を見やすくするという目的と、奏任官以上の記載を統計分析するという目的では、必要とされる構成が違ったものになるという問題である。双方を満たす解決方法は見いだせなかったため、当DBでは人物レファレンス目的を優先した。そのため、本稿では、複数行に亘る役職履歴などに対しては、各個人役職欄のデータ数を数え上げた別シートを作成する、などの処理で対応することを提案しておきたい。この数値は自分の研究対象以外のものに限るなどの処理も可能であるし、数値ではなく「〇〇省」など文字列で役職の概要が示すことも、関数の処理やVBAなどの利用で可能であろう⁴³。

⁴³ 収録した役職を数値の1として数え、抽出した人物が各年に幾つの役職でDBに

DBにおいてはVLOOKUPやCOUNTIFなどの関数を利用して、多数の人名の照合をかけることや、抽出した集団の実人数を算出することも可能であり、その際の注意事項やポイントなどについては補論を参照してほしい。

おわりに

以上、DBの作成過程で生じた課題とその対処法について述べてきた。史料をデータベース化することには、史料批判が行き届かない可能性が生じる点や原典へのアクセスのための情報が不可欠であることを指摘した。さらに作成上の人名の取り扱いにおいて、改姓改名や読み仮名の問題、漢字表記の問題、同一人物性の検討の問題などの課題があり、それぞれDBでの対処法を示した。最後にデータベースの構成と利用目的の関係について、DBの場合に生じた問題点を明らかにした。

データベース作成上の問題はこれだけに留まらず、作成するソフトの選択、セキュリティの設定、公開方法やデータの更新問題、更には完成後であっても、永続的な利用を可能にする保存法など、多方面にクリアすべき条件や課題が山積している。しかし共用データの公開は今後、確実に拡大していく潮流である。先々に控えている課題がいかなるものであろうが、本稿で示したような試行錯誤で、一歩ずつ改良の道を探っていく他に方法はない。

現在公開しているDBについても、利用アイデアや問題点を見出すには、広く各界の知見を寄せていただくことが必要である。利用してみても問題解決や分析処理の妥当性についての問い合わせに対しては、製作者として誠実に応じていきたいと考えている。こうした利用者間の情報交換こそが、DBを駆使した新しい歴史研究の道筋を作り出すことは間違いないことであり、そのための一助となることが本稿の目的であることは、冒頭に述べたとおりである。

補 論

利用者間の情報交換が重要であるというのが本稿の主張であることから、ここにDBに入力されているデータの特質に関連して、1) 多数の人名との照合

記載があるのか、という表は関数SUMIFを利用して作成できる。

2) 集団内の実人数の割り出し 3) 各人各年の収録役職数一覧の3点について、Excelの関数式を利用する際に推奨する手順方針と注意点を補論として記して、公にすることをお許しいただきたい。

1) 多数の人名との照合

ここではDB側の人名を参照側とし、利用者が調べたい人名を検索側とする。手順の方針として、まず両者の氏名を検索用に加工したものを用意する。関数VLOOKUPなどで加工した氏名を一致させ、加工以前の氏名を抽出させる。次に両方の加工以前の氏名が一致するかをEXACTで検索し、一致しない場合のみ個別に同一人物かの確認をする方法を提案する。

加工内容として、DB側には数字付や（'）付の氏名が存在するので、数字は削除し（'）付は苗字のみにするなどの作業が必要である。但しこの作業の結果、検索側（DB側）の氏名には同姓同名者が含まれる。更に以下の作業を実行すると、検索側の人名リストの中で同姓同名者がヒットした場合は、その中で行数が最小であるデータだけが戻り値として表示される。さきに提案した方法では、戻り値が加工以前の氏名なので、ここに数値が入っている場合は、検索リスト側に他の同姓同名者が存在すると判断できる。

参照側と検索側の氏名の加工内容の続きとして、両者で統一が必要なのは、苗字と名前の間のスペース有無、新旧字、異体字、本字、俗字など利用漢字が挙げられる。さらに混用の多い（菊池と菊地）（次郎と二郎）なども一方に統一する方が、漏れのない結果を得ることができる。筆者が作業している場合に統一している文字列を挙げる。

斉藤→斎藤 菊地→菊池 大田→太田 大築→大槻 渡部→渡辺 度会→渡会
灘波→難波 丘、邱→岡 阪→坂 嶺→峰 庄→荘 己→巳 蔭→陰 河→川
脩→修 虎→寅 麿→鹿兒 麻呂→麿 造→蔵 次郎、治郎→二郎
輔、丞、介、弼、亮→助

さらに新旧字や親字への統合として以下の字を統一している。

瀧→滝 龍→竜 澤→沢 壽→寿 豫→予 與→与 嶽→岳 兒→児 峯→峰
曾→曾 嶋→島 條→条 傳→伝 卯→卯 富→富 祢→禰 蘆→芦 繩→縄
縣→県 惠→恵 禮→礼 萬→万 埜→野 堯→堯 鍊→鍊 哲→哲 勵→励

諫→諫 楳→梅 晉→晋 亘→宜 苜→刈 踐→踐 鉛→鉛 菟→菟 仞→仞
 鑿→鑿 惠→德 僊、僊→仙 穎→穎 昂→昂 迪→迪 勗→勗 桿→杆
 鋪→鋪 淵→淵 彝、彝→彝 儁→俊 雞→鷄 灑、熙→熙

上記の漢字は、人名の検索エンジンを改良して、新字旧字、異体字でも検索可能にする際にも統合すべき文字として利用でき、実際に「明治期官僚・官職データベース Web版」では、上記の文字をすべて統合している。

以下、実行すべき手順を模式的に述べる。(A列) 調べたい人名 (B列) A列の加工済み人名 (C列) D列の加工済み人名 (D列) 官員録・職員録人名と並べる。(E列) VLOOKUP関数でB列と一致する人名をC列から検索し、抽出データとしてD列の人名を表記させるよう設定する。この場合、セルE2の式は=VLOOKUP(\$B2,\$C:D,2,0)となる。(F列) EXACT関数でE列とA列を一致させる。セルF2の式は=EXACT(E2,A2)。これによって調べたい人名 (A列) が官員録・職員録氏名 (E列) に存在しない場合は、E列F列に (#N/A) が表示される。F列が (FALSE) の場合は、名前の加工の結果、氏名が一致しているだけなので、本当に求める人物かのチェックが個別に必要となる事例を意味する。

2) 集団内の実人数の割り出し

手順の方針は、関数のIFとCOUNTIFを入れ子状態の数式を利用する。まずCOUNTIFで同一の人物IDが対象の列に幾つ存在させるかを検索させる。IFを利用して、対象列に一人しかいない場合は戻り値1を表示させ、複数存在する場合は、その初出の場合のみに戻り値1を表示させる。最後に関数式をいれた列をΣで合算する。

実際の手順を模式化する。

(A列) 対象人物の人名 (B列) 人物ID (C列) =IF(COUNTIF(\$B\$2:B2,B2)=1,1,"")と配置する。C列をΣで合算する。

3) 各人各年の収録役職数一覧

手順の方針は、対象者の該当役職データを数値1に変換させた列を作成(このシートは仮に“参照”と名付ける)し、別シート (仮に“役職数”と名付ける)

に各年ごとの役職数を合算させる。

参照シート（A列）対象者氏名（B列）人物ID（C列）以下、計測する役職が年別に表示。

役職数シートも（A列）対象者氏名（B列）人物ID（C列）以下、各年を配置。C列以下に入る数式は=SUMIF(参照!\$B:\$B,\$B”，参照E:E)となる。

上記の中には、筆者が便宜的に用いているにすぎない方法も含まれており、より適切な方法のための議論や情報交換、さらに既存データへの対処方法の提案がなされるなど、史料の電子化の推進と利活用の活性化のための建設的な作業がなされることを期待したい。