

THE PROSOPOGRAPHY OF THE BRAZILIAN EMPIRE

Roderick Barman and Jean Barman
University of British Columbia

An increasing amount of research has been undertaken in the last decade on the power structure of the Empire of Brazil (1822–89). In common with similar work on the rest of Latin America, many of the studies on Brazil have adopted the concept of the elite for their theoretical framework.¹ And, as elite studies must do if they are to be more than speculative, several of the works have drawn on prosopography, or collective biography, to support the interpretations of the power structure that they advance. If the results have not, in general, been entirely satisfactory, this is largely due to the limitations in the prosopographical materials used. Since the materials have not easily lent themselves to a proper quantitative approach, they have been used selectively in the form of examples, the validity and relevance of which must always be suspect.

This deficiency in the prosopographical data on the Empire of Brazil early became a problem in our own research on the power structure of nineteenth-century Brazil. It proved much easier to advance hypotheses than to obtain the data necessary to test their validity. When extant, the prosopographical sources, both printed and manuscript, were usually disparate in nature, uneven in quality, and difficult of access. We therefore began to look for ways in which to acquire and to organize in a permanent form some of the basic prosopographical materials needed by historians for a proper study of the Brazilian Empire.

This essay is concerned with the construction, content, and utility of the resultant database, which contains machine-readable biographies of virtually all individuals of some significance in the Brazilian Empire. The essay is methodological and does not include results from use of the database in our research, which appear elsewhere.² The essay is intended to be of interest both to other historians of the Empire, with whom we would hope to cooperate in the future development and use of the database, and to researchers in other areas who may be considering quantification as a possible supplement to their current research and who may find in our work some applicability to their own.

CONSTRUCTION OF THE DATABASE

Our project differs in two respects from many others concerned with historical quantification. Instead of originating with a set of raw data already in existence,

Department of History, University of British Columbia, Vancouver 8, Canada

our project has entailed the actual creation of a data set, or database, through the collation of a wide variety of disparate source materials. Secondly, these materials are essentially nonnumeric. Thus, while elements from it can be used statistically, the database itself is not a set of historical statistics.

Our project grew out of a box of file cards on individuals holding the major political positions in the Brazilian Empire. By 1972 this supplement to our general research contained some two thousand individuals. The cards could no longer be manipulated effectively and, for conceptual reasons to be discussed in the next section, the scope of our prosopographical research needed to be greatly expanded.

These factors made the creation of a machine-readable database almost inevitable. Extensive discussions with the University of British Columbia Computing Centre led to a successful application for funding to the Canada Council. A research trip was made to Brazil in the summer of 1973 to gather the desired prosopographical materials. During the following two summers, through close cooperation with an extremely able computer programmer, we were able to design the format of our database and to develop the necessary programs, or sets of instructions, for the computer.³

There were two primary concerns in designing the database. First, the format had to be sufficiently flexible to permit not only the quick extraction but the easy addition of information. This was particularly important since the database could only be constructed through the gradual addition of data from different sources. Second, the database had to be designed to retain all relevant biographical information with equal standardized specificity regardless of the varying complexity of separate lives. In other words, an individual attending several universities and occupying a dozen relevant positions would have to have *each* university and *each* position included in the database in the same standardized format as the individual attending a single university and holding a single position. It was therefore impossible to make the format of the database directly compatible with any of the standard statistical programs used in the social sciences, such as SPSS, MIDAS, or OSIRIS, since these are primarily designed for more uniform survey or cross-sectional data.

While the prosopographical materials available on the Empire were, in their original state, very uneven in quality and disparate in form, they did fall into two basic types: (1) lists of individuals sharing some common attribute, often institutional in nature, such as the holding of a particular political position or graduation from a particular university, and (2) collective biographies, whose subjects were selected upon criteria ranging from the author's subjective judgment of importance to common affiliation with a province or institution. Obviously, the same individual might appear in any number of sources, with his biographical information both differing in content and being possibly contradictory in substance.

The criteria adopted for the inclusion of individuals into the database, a subject to be considered below, meant that the first type of source would be used to construct the database. Once all relevant individuals were included in the database, the collective biographies would be used to add further information

on their lives and careers. Special programs were designed to allow construction of the database in stages; once set up through the collation of two major lists of individuals, further lists could be gradually added.

Central to the collation process is a concept sometimes known as "record linkage."⁴ Information about each individual is organized into a separate record, or "item," which is identified by the name of the individual. The items in a list to be collated contain not only name and common attribute (such as graduation from a particular university) but any additional information given in the list (such as place or date of birth). The items in this input list are compared with those already in the database. Where the names match, the appropriate items are collated into a single item. The remaining input items are merged directly into the database.⁵

During the amalgamation process, conflicts in substance sometimes arise between the content of matched items. The input item may, for instance, give a different province of birth than that in the database item. Before collation takes place, either the database or the input list must be designated as the preferred source, and information from that list is chosen whenever a discrepancy occurs. All conflicts are recorded in a printed list, which can be checked manually to confirm that the right choice was made.

Use of the name as the vehicle for collation drew us into a general problem of prosopographical research on nineteenth-century Brazil. Different individuals sharing the same name were by no means uncommon and, due to the custom of giving male offspring multiple Christian and surnames, the use by an individual of variant forms of his given name at different periods in his life was even more common. To prevent erroneous record linkage, the collation process contains provision for different individuals with identical names to be permanently distinguished and for all information about a single individual with several name forms to be collated together under a preferred form of the name.⁶ In the fall of 1974 we began to construct the database. By mid-1977 it contained some fifteen thousand items including all the lists of individuals given in the left-hand column of the table.

CONTENT OF THE DATABASE

The utility of the database as a research tool depends even more upon a careful selection and organization of its contents than upon good construction. Just as analysis is concerned with the identification of correspondences and relationships that are masked by particularities, so quantification almost by definition implies the organization of material in forms that promote the identification of underlying similarities. When computers, with their capability for analysis and their tremendous potential for misuse, are employed in quantification, great care must be taken to ensure that the selection of materials and their organization is as unbiased, as logical, and as consistent as possible.

In our database the content lies on two very different, if connected, levels. The first is the types of individuals, or items, to be included in the database, the second the kinds, or "fields," of information to be recorded about each item.

Categories of Individuals Included in Data Base

	<i>Categories Now In Data Base</i>	<i>Desirable Categories To Add At A Later Date*</i>
Educational	1. Coimbra University graduates from 1776 (Data taken from items 19, 29, 35, 44 in Sources) 2. Coimbra matriculants from 1771 (19, 29, 35) 3. São Paulo Law School graduates (7, 27, 43) 4. São Paulo matriculants (7) 5. Recife (Olinda) Law School graduates (8, 28) 6. Rio Medical School graduates (6) 7. Bahia Medical School graduates (1) 8. Brazilian graduates and/or matriculants at selected foreign universities from 1771 (32, 38, 39)	1. Rio Politechnical School (earlier Military School, Central School) graduates (5, 36) 2. Ouro Preto School of Mines graduates 3. Bahia Agricultural School graduates 4. Any Brazilian graduates and/or matriculants at foreign universities from 1771 5. Graduates of D. Pedro Segundo Secondary School
Political	9. Councillors of State (22) 10. Ministers (22, 31) 11. Senators (22, 31) 12. Individuals on triple list for Senate (40) 13. National deputies and substitute deputies (15, 22, 31, 41) 14. Provincial presidents and vice-presidents occupying presidency (20, 22) 15. Regents (22)	6. Provincial deputies
Administrative	16. Bureaucrats in ministries of central government, 1845–89 (13, 23, 24)	7. Bureaucrats in ministries of central government, 1822–44

Categories of Individuals Included in Data Base

	Categories Now In Data Base	Desirable Categories To Add At A Later Date*
	17. Envoys and chargés d'affaires (16)	8. Provincial secretaries
	18. Professors at Brazilian institutions of higher education (8, 18, 26, 28, 36, 42)	9. Lower-level diplomatic corps 10. Heads of Brazilian secondary schools
Judicial	19. Judicial officials from <i>juiz de direito</i> up, 1832–89 (2, 3, 10, 14)	11. Judicial officials from <i>juiz da fora</i> up, 1822–32 12. Lower-level judiciary
Military	20. Army officers from <i>brigadeiro efectivo</i> up (37)	13. Councillors of War, <i>vogais</i> , members of Supreme Military Council (25)
	21. Naval officers from <i>chefe de divisão efectivo</i> up (9, 11, 12, 17; some early data incomplete)	14. Upper-level National Guard officers, selected years (10)
Religious	22. Bishops (30)	
Social	23. Holders of Brazilian titles (33)	15. Brazilian-connected holders of non-Portuguese foreign titles
	24. Brazilian-connected holders of Portuguese titles (45)	16. Members of Brazilian Historical and Geographical Institute (21)
	25. All members <i>do conselho</i> (4)	
Economic	26. Members of Commercial Tribunals (after 1876, <i>Juntas do Comércio</i>) (10)	17. Directors of D. Pedro II Railway (23) 18. All other railway directors 19. Directors of Bank of Brazil (23) 20. All other bank directors 21. Officials of Rio Commercial Association (34; some data incomplete) 22. Officials of local commercial associations

Categories of Individuals Included in Data Base

<i>Categories Now In Data Base</i>	<i>Desirable Categories To Add At A Later Date*</i>
	23. Any systematic lists of property owners or other holders of large-scale economic wealth or power

*For most of these categories, data is lacking at present.

Note: Unless otherwise specified, the time period for each category covers the Empire, 1822–89, or that part of Empire during which category existed. Numbers after each category refer to items listed in Sources.

While the particular criteria for inclusion will, on either level, be determined in their specifics mainly by qualitative considerations, they must in their essentials be as universal, precise, and uniform as possible if the resultant database is to be viable as a research tool.

Criteria for Inclusion of Individuals in the Database

The criteria for the inclusion of individuals in the database are ultimately determined by the perception held of the power structure of Imperial Brazil. If that structure is seen in terms of a narrow elite, the result will be a small database that can be constructed fairly rapidly and is not difficult to manipulate. On the other hand, such a database can only with difficulty be used to test other perceptions of the power structure. If the structure is seen as involving a large or multiple elite, then the database will not only be much larger in size but can be employed to test a number of hypotheses.

Also in favor of a broad database is the limited analytical value of an elite study restricted in scope to the elite itself. The functioning of elites can be seen most clearly against the background of their sociopolitical milieu. This milieu is basically formed by the groups standing next to the elite in influence and status. Through these groups the elite has contacts with the larger society and from these groups most of the elite's members are drawn. Only comparison with this broader "elite pool" can establish characteristics unique to the elite or make clear the process of elite recruitment.

The adoption of a broad framework for the database is also encouraged by the difficulties in establishing in quantitative terms what constitutes elite membership. It is easy to assign elite membership to specific individuals on qualitative grounds, but the criteria adopted in each individual case will not be uniform between cases. The usual, and most practical, remedy for this difficulty is to equate the sharing of common attributes, or membership in designated institutions, with membership in the elite. The drawbacks to this method are

obvious. Elite membership may well be much broader than institutional membership, especially if the designated institutions are small in size and the turnover in membership is slow. Through special circumstances nonelite elements may gain entrance to these institutions, and the importance of the institutions may vary over time.

While the adoption of a broadly framed database does not resolve these problems, it does decrease their importance. Elite members without institutional membership are more likely to be included within the database, and characteristics found to be typical of institutionally-connected elite members (perhaps paternity or rapid advance in lower-level positions) can be searched for in the entire database to locate possible elite members not institutionally connected.

A practical constraint that must be taken into account when the database is planned, whatever its conceptual framework, is the nature and availability of prosopographical material. The best and most direct evidence is usually in the form of lists of members of institutions, the entries of which may contain some biographical information. In the case of Imperial Brazil, such data exist for virtually all types of institutions, save economic. Lists of corporation directors, particularly of railways, banks, and insurance companies, exist, as do qualitative descriptions of individuals as "wealthy coffee planter" or "large-scale merchant." However, no overall data on economic wealth appear to be available in a form that is suitable for systematic collection, such as tax returns or landowning records.

In light of these conceptual and evidential constraints, we decided to make the database for our project as wide as a coherent definition of the elite pool would permit. While literacy was never, as it has since become, a direct prerequisite for participation in politics, it was a practical necessity for anyone in any position of authority in Imperial Brazil. Literacy was simply the first step, the logical culmination of the educational process most often being the obtaining of an academic degree. The lists of the *bacharéis* constitutes one fairly precise and uniform criterion for inclusion in the elite pool. It is also a fairly universal criterion, our current estimate being that some fifteen thousand individuals living in Brazil from 1822 to 1889 held degrees. In addition, test results using a sample indicate a fairly strong correlation between the possession of a degree and political success.

Possession of a degree cannot, however, constitute the sole criterion for inclusion in the database. The late date of foundation and the uneven pattern of distribution of the Brazilian institutions of higher education introduce serious biases in geographical and time distribution. Preliminary research shows that a high percentage of the politically important in the early years of the Empire did not hold degrees. More significant, the behavior and knowledge requisite for success could be acquired without formal education. Qualitative evidence makes clear that association with college students, the frequenting of literary circles, residence in Europe, and even self-education quite often produced the same result. To compensate for these biases and also to permit the examination of career patterns, we have included not only some lower-level positions but social categories, the achievement of which did not so much aid in the procuring of

elite status as they were an acknowledgement of its already having been attained (although lower-level titles of nobility were often awarded to those of merely local significance).

Finally and most importantly, the database contains, as shown in the table, the holders of the top institutional positions not only in the area of politics but in administration, the judiciary, the military, religion, and, to a limited extent, in economic life. The particular criteria used in each area were developed independently primarily on the basis of qualitative research into the significance of that institution in the Empire.

Criteria for Inclusion of Information within Each Item⁷

The two basic criteria for the inclusion of each field, or piece of information, within an item are whether that field is a prime piece of biographical information, or contributes, either directly or in combination with other fields, to an analysis of the power structure of the Empire of Brazil. In order to facilitate the testing of a wide range of hypotheses, the second criterion has been interpreted loosely. Some fields contain information that is partly descriptive and subjective but are included to provide some substitute for hard data otherwise unavailable. In prosopographical terms, the range of our fields of information is neither unique nor unusual. Nor do the number of fields prove unwieldly, since our analytical programs are designed to select out only the relevant fields and items before starting on manipulation and analysis.

The use of the computer imposes certain requirements (but less than was formerly the case) on the process of regularization. If effective manipulation and analysis of the database is to occur, the fields of information in each item must be separately retrievable by the computer and, to achieve this end, each field must be uniquely identifiable within the item. In our database, each field of information is identified by its location in respect to the other fields in the same line. Each line must in turn be identified. The first three lines of each item, which are "nonrepeating" and so appear once in every item, are identified by their order of appearance in the item, with the first of the three lines beginning with a special character used to introduce a new item. The remaining lines of the item, made up of eight kinds of "repeating" lines with each kind appearing an optional number of times within the item, are identified by special introductory characters identifying the kind of repeating line.

The information in the three nonrepeating lines consists of basic vital statistics—birth and death date and place, parents' names and occupation, and nature of parentage. Nature of parentage refers to the legality of the birth and primarily reflects information found in university matriculation records. The repeating lines contain information about events that occur a differing number of times in the lives of individuals. It is the use of repeating lines that allows individuals to be treated evenly in the database in terms of their significance. These lines also provide depth and time perspective to the biographies. Each kind of repeating line can be repeated in any item as many times as necessary (within an allowable range of 0 to 50) to record all relevant information.

The first three kinds of repeating lines are used to record relationships and so make possible a broad analysis of the role and importance of familial ties in the creation and maintenance of the Brazilian elite.⁸ The "marriage" repeating line contains basic information about a marriage—ordinance, date, spouse's relationship, and spouse's vital statistics. Each marriage is recorded in a separate line. Separate instances of the "offspring" line are used to record the names of children, while instances of the "relative" line record other individuals with whom it is useful to note a relationship. The remaining five kinds of repeating lines are used for higher education, occupation, political affiliation, titles or honors, and key positions. The "key position" repeating line constitutes the heart of the database and records precise details about the holding of each position included in table 1, save that education categories are included in the "education" repeating line and social categories in the line on titles and honors.

The "occupation" repeating line differs from the other career lines in being not so much objective as suggestive. This line records contemporary descriptions of the principal means of economic support or livelihood during adult life. The descriptions can be recorded with considerable precision as to nature of occupation, degree of success, location, and duration. While such descriptions clearly vary in quality and are not necessarily consistent between sources of information, they do nonetheless partially compensate, if used with discretion, for the general scarcity of hard economic data.

Whether they appear in a repeating or in a nonrepeating line, fields of information are either of a "fixed" or "variable" length. A field appearing in an item must appear in every item. Fixed-length fields retain the same size in all items and contain blanks when the information is unavailable. Variable-length fields can be of any length with their termination always being indicated by a designated symbol at the end of the field. Where information for the field is lacking, only the symbol appears. For the retrieval of information, fixed-length fields are preferable, and it is for this reason that the only variable-length fields included in the database are those containing personal names and titles of nobility.

The need for regularity and uniformity in quantification combines with the computer's demand for conciseness and precision to make the use of codes virtually inevitable. The great advantage of using codes is that it does enforce a precise regularization of data. If, for instance, the raw information on Brazilian geographical locations, in particular on birthplaces, were used directly, the variety in type and format would be so great as to prevent any analysis. The adoption of a two-part code resolves the problem. The first part, based on the postal abbreviations, indicates the province. This information is available for the vast majority of birthplaces. The second part refers to standard administrative districts within the provinces, which we have taken to be the approximately five hundred judicial districts, or *comarcas*, as they existed at the end of the Empire. These *comarcas* are sufficiently small to permit precision but not too numerous to be unmanageable. The *comarca* codes, since they are made up basically of the first four consonants in the name of each *comarca*, can generally be both coded and identified directly, without reference to a separate codebook. The principal

work in coding is to determine from available sources (especially items 10, 46, 47, 49, 50, and 51 in the section on sources) in what comarca of a province a particular geographical place lay. A permanent codebook of all geographic locations is kept. The codes also make provision for unidentifiable place names and for the absence of data. The result in the database is fully comparable and compatible data on place names.⁹

Although the geographic codes are for practical reasons alphabetic, numeric codes are generally more advantageous and they are used in all other instances. They are not only more concise, usually needing no more than two digits, but they permit a degree of manipulation that neither the original format nor alphabetic coding can match. The utility of "collapsible codes," in which each character of the code has a particular level of significance, also militates against the general adoption of codes based on abbreviations. In addition, numeric codes are much easier to use than might first be feared. Each coding sequence rapidly loses its anonymity, and the separate codes within it become directly and instantaneously identifiable.

A good example of the advantages of numeric coding is contained in the repeating line for key positions. After the introductory symbol, each line starts with a two-digit code that allows the recording of up to 99 different positions of significance that Brazilians could occupy during their careers. Not only does each position possess its own code, but the positions of similar type all have the same first digit, an application of collapsible coding. Judicial positions all start with a 1 (10–19) and national political posts with a 3 (30–39). Following the position code comes a six-letter code, explained above, indicating the geographic location of the position, a shorter code (followed by a suitable number of blanks) indicating a nongeographic location, or six blanks if there is no location to be specified. The final sixteen digits contain the beginning and ending dates for the occupation of the position. Of the eight digits in each date, the first seven are the date itself and the eighth a numerical code indicating the validity or nature of the date. This refers to a range of possibilities from actual date of appointment or dismissal to simply the earliest or latest date at which the position is known to have been occupied by the individual. Thus, each instance of the key position repeating line contains only twenty-five coded characters but basically retains all the specificity of the original data.

UTILITY OF THE DATABASE

The prime and fundamental utility of the database is simply that, for the first time, a large number of disparate biographical sources relating to the Empire of Brazil have been brought together and collated into a standard format. Since the principal printed works of collective biography tend either to concentrate on the celebrated (for example, items 60, 84, and 97) or be limited in scope,¹⁰ the database is particularly useful as a source of information on people of second and third rank in the Empire.

The format of the database, designed primarily for manipulation and analysis by computer, does not prevent direct manual access to its copious

biographical information. A printed copy of the database can be used to obtain information on a particular individual, or it can be employed prosopographically. The background of individuals conjoined in some significant respect—the signers of political manifestos, voters in important parliamentary rollcalls, members of local *câmaras municipais*, officers of the National Guard, and even slave holders—can be investigated with relative ease. Manual consultation of the database lies at one end of a spectrum of utility that ranges through manipulation to statistical analysis. Several programs, some of them especially written, are employed for these purposes. The various methods of manipulation and analysis, which are here discussed separately, are usually employed in conjunction with each other, the output from one operation forming the basis for the next.

Because the database is so large, some sixty thousand lines by early 1977, it is not efficient or economical to use in its entirety. The first step in virtually any manipulation or analysis is to create a working dataset that contains only the relevant items. The ability to extract from the database all items sharing a common characteristic or condition is a fundamental kind of manipulation. The condition for extraction can be relatively simple (the presence or absence of a specific code or codes in a designated field or fields) or complex. It is possible, for instance, to postulate a condition involving holding a particular kind of position either before or after or at a designated point in time, between two designated dates, or for a period longer or shorter than a designated time period. If desired, all items that do not contain at least one instance or contain more than a designated number of instances of a particular repeating line can be extracted. The condition for extraction can be a separate list of names independent of the database.

The extracted items can be outputted either in their existing format or in a new one. The new format need contain only selected fields of information from the original item that can be located on the output in any order or position. The ability to reformat the item is extremely important. Not only can printed output be given maximum legibility, but data from an item can be organized so as to be compatible with the requirements of a standard analytical program such as SPSS.

The manipulation programs also include the facility, subject to certain technical limitations, to reorder a list of items upon fields other than the personal name field on which they are ordered in the database. If the field on which the items are being reordered is part of a repeating line, such as a date of appointment as minister, each repeating line to be reordered is treated as though it were part of a separate item.

The practical utility of the manipulation programs is exemplified in our use of their facilities to check and regularize the data we had gathered on judicial appointments from the original decree books and annual ministerial reports. The computer first extracted all those individuals who ever held a judicial appointment and then ordered them by type of position held, by province, by judicial district, and by appointment date. The output was arranged on the page to facilitate the discovery of gaps between date of dismissal and that of the next appointment to the same district. Errors and omissions in the original data were

thus easy to identify, and mistakes made in the two transcriptions of the material (from the originals to research notes and then onto computer cards) were easily caught. Such a check could not have been carried out manually.

A facility of the programs that combines manipulation and analysis involves the production of "computed fields" from the performance of a mathematical operation on two or more conceptually numeric fields within a single item. At present virtually all such fields involve dates (of birth, graduation, appointment, and so on), and the mathematical operations can be performed on the entire date or on the year alone. In the first case, the result can be expressed either as a number of days or as a number of years with the remainder given in days or as a percentage of the year. Thus, the age of an individual when appointed to a designated key position can be calculated by subtracting date of birth from that of appointment and the duration of the appointment by subtracting appointment date from that of dismissal. Computed fields can be calculated from those already produced, all in the course of a single manipulation. So, not only can the duration of an appointment be calculated but, for individuals holding an office more than once, the average time period during which the office was held can be determined.

The uses of the database so far discussed have basically involved manipulation or analysis of information within single items rather than analysis between items. Such analysis is normally performed upon data that have already undergone extraction, reordering, formatting, and possibly computation. The simplest, and in some cases the most efficient, means of analysis is not through the computer but by use of a calculator to produce statistics from data in a printed output. The quickest method, for example, of discovering the average age of ministers in particular cabinets is, once the computer has been employed to calculate the age of each minister and to reorder the relevant items chronologically by date of appointment, to use a calculator to find the average age of each cabinet. However, most statistical analysis involves enough data to make manual calculation tedious. Limited statistics, including frequency tables, can be obtained directly from the data. This would be the most efficient way to determine, for instance, the range of ages and the number of each age for students matriculating annually at a designated university or the number of appointments as *juiz de direito* made annually to each province.

For more complex statistical procedures, we are using the standard statistical packages of computer programs designed for the social sciences, in particular SPSS and MIDAS. The limitations in these programs, which make them impractical and overly simplistic as a basis for organizing the original database, do not seriously affect their suitability for statistically examining selected portions of that database. The speed, reliability, and efficiency of these well tested programs and the sophistication of their output facilities far outweigh their inability to handle data only in limited formats. Thus, rather than designing a new analytical program which, although directly compatible with our database, would be cruder and less trustworthy than SPSS or MIDAS, we have included the necessary reformatting capacity in our general manipulation program. This is one reason, for instance, why computed date fields can be outputted in three

different formats. Full descriptions of the statistical capabilities of SPSS, MIDAS, and OSIRIS are generally available, and it would be redundant to discuss them here.¹¹

No statistical procedures can, however, fully compensate for limitations and deficiencies in the original materials. In our case, as in so many projects using historical data, the items vary in completeness and the fields in reliability. Even for such a central group as cabinet ministers, several birthdates are missing or uncertain. Not only must special care be taken so that no statistical procedure is more powerful than warranted by the data, but deficiencies in particular fields, including missing data, must be taken into account when using the results of the procedure.

FUTURE DEVELOPMENT OF THE DATABASE

As we anticipate its development, our project involves the establishment of a permanent set of historical data that can be utilized by many historians in a wide range of projects. The categories of individuals in the left column of table 1 have been incorporated into the database. During a sabbatical in Brazil and Europe, 1977–78, we intend to search for data on the categories in the right column and to seek further information on some categories in the left column.

Information from collective biographies (listed in part C of Sources) has already been added to particular categories of individuals in the database about whom analysis is currently taking place. Such information is now being added in a more systematic fashion. Reference to the biobibliographies (items 31 and 52) greatly facilitates use of the major collective biographies. We have also developed a supplement to the collective works in the form of a printed sheet identical in format to the fields of a complete item. This form is used to record relevant information about potentially eligible individuals that we encounter in the course of our general research. The information is added to the database when suitable.

The addition of information from an increasing number of sources brings to the fore a limitation of quantification in general and use of the computer in particular. It is highly impractical, virtually prohibitive, to record the provenance of the separate pieces of information in each item.¹² The provenance of certain fields, such as those on education or key positions where the original data is found in one of the categories for inclusion, is easily determined. Where information in a field, such as that of birthdate, could have come from several sources and a conflict arises, two partial remedies exist. Manual lists of conflicting evidence are maintained, to be used when the next opportunity arises to obtain data from an original source, such as the birth certificates usually included in university matriculation records. Second, a one-digit code is used in the database to indicate validity. The misrecording of digits in dates is a major weakness in Brazilian collective biography, and this coding facility points out a problem within a date.

The utility of the database as a research tool has naturally increased as it has grown towards completion. We are, for instance, using the database to

supply the quantitative materials for a major study of education, career selection, and political recruitment in the Brazilian Empire, looking at such questions as the role of geography in education and elite recruitment, the formation of generations and the significance of generational conflict, the selection of occupations among post-secondary graduates, the relationship of politics to both the several branches of national administration and to private professions, the socio-economic basis of political affiliation, and the significance of familial relationships in the process of elite recruitment. The database is also being used to study separate topics, including marriage patterns, the transmission of foreign technology through education, and even the cultural influences on changing Christian name usage.

As the database develops and our own research proves its utility and validity, we would hope for increasing cooperation with other researchers. At present we are particularly interested in the exchange of information concerning relevant categories of individuals to be included and would very much like to establish contact with any historians who have collected such data themselves or have access or knowledge of it. Knowledge of economic data would be particularly useful, and we would prefer to set up an exchange by which access to the database would be traded for permission to incorporate relevant data into it. We would also like to hear from any researchers with a reaction to the project. Criticism, advice, ideas for future development, and suggestions for cooperative research are especially welcome.

NOTES

1. See, in particular, Richard Graham, "Landowners and the Overthrow of the Empire," *Luso-Brazilian Review* 7(1970):44-56, an article whose approach is extended to the whole of Latin America in his "Political Power and Landownership in Nineteenth-Century Latin America," in R. Graham and Peter H. Smith, ed., *New Approaches to Latin American History* (Austin: University of Texas Press, 1974), pp. 112-36; Eul-Soo Pang and Ron Seckinger, "The Mandarins of Imperial Brazil," *Comparative Studies in Society and History* 14(1972):215-44; Eugene Ridings, Jr., "The Merchant Elite and the Development of Brazil: The Case of Bahia during the Empire" *Journal of Inter-American Studies and World Affairs* 16(1973):335-53, and "Elite Conflict and Cooperation in the Brazilian Empire: The Case of Bahia's Businessmen and Planters," *Luso-Brazilian Review* 12(1975):80-99; William Dudley, "Institutional Sources of Officer Discontent in the Brazilian Army," *Hispanic American Historical Review* 55(1975):44-65; and Thomas Flory, "Judicial Politics in Nineteenth-Century Brazil," *Hispanic American Historical Review* 55(1975):664-92. Although on the late colonial period, John Kennedy, "Bahian Elites, 1750-1822," *Hispanic American Historical Review* 53(1973):415-39, adopts the same approach. The unpublished doctoral dissertation of José Murilo de Carvalho, "Elite and State-Building in Imperial Brazil," (Stanford University, 1974) was not available for consultation at the time of writing.
2. See "The Role of the Law Graduate in the Political Elite of Imperial Brazil," *Journal of Inter-American Studies and World Affairs* 18, no. 4 (November 1976):423-50, revised from a paper presented to the American Historical Association, December 1974, and "A formação dos grupos dirigentes políticos do Segundo Reinado: a aplicação da propografia e dos métodos quantitativos à história do Brasil Imperial," delivered in November 1975 to the Congresso de História do Segundo Reinado, Rio de Janeiro, and to be published in the *Anais* of the Congresso.
3. We would particularly like to thank the Canada Council for their generous assistance. The Associate Director of the UBC Computing Centre, Mr. Al Fowler, and the Gen-

eral Programmer, Mr. John Coulthard, both gave us excellent advice and guidance in the early stages of the project. To his intelligent cooperation, marked ability, and willingness to work long hours, we owe more than we can properly acknowledge to our programmer, William Kershaw.

4. The standard article explaining record linkage is still Ian Winchester, "The Linkage of Historical Records by Man and Computer: Techniques and Problems," *Journal of Interdisciplinary History* 1(1971):107-24.
5. The input list to be collated does not have to contain all the fields of information already in the database, but it must contain the same fields for each individual in the list in the same order as they appear in the database.
6. The preferred form is, where possible, the name used by the Instituto Nacional do Livro (in item 52), otherwise the name which we judge to have been most commonly used in adult life. Name parts indicating a relationship, such as "Júnior," are abbreviated to their first letter and included after the last name.
7. A more precise description of the format of the item and of the construction and manipulation of the database is available in our article with William Kershaw, "Prosopography by Computer, the Development of a Database," *Historical Methods Newsletter* 10, no. 3 (June 1977).
8. To facilitate analysis of familial connections, any individual within a parentage or relationship line who has his or her own item is identified by a special character at the end of the name.
9. Foreign localities are treated in a parallel fashion through use of a two-digit code to indicate the separate countries, as of 1889, and a four-letter consonant code name of a city or town. The first thirty digits in the numerical code are reserved for Portugal, which is divided into its administrative districts (*concelhos*), based on item 48.
10. Scope is most often limited to provincial affiliation or a particular occupation. Coverage is extraordinarily uneven. The three volumes of Studart on Ceará (item 99) and the works of Guaraná and Bittencourt on Sergipe (items 73 and 59) provide complete information on anyone of importance born in or affiliated with these fairly minor provinces. But on Bahia, which produced so many people of importance during the Empire, the only collective biography is Souza (item 98), which is restricted to the celebrated. Similarly, while the upper ranks of the military are well covered (especially items 9 and 37), reliable works on provincial presidents are available for only five of the twenty provinces (items 65, 67, 71, 76, and 104).
11. The best descriptions of the programs are available in the respective instruction manuals; see Norman H. Nie et al., *Statistical Package for the Social Sciences*, 2nd ed. (New York: McGraw Hill Book Co., 1975), Daniel J. Fox and Kenneth E. Guire, *Documentation for MIDAS*, 3rd ed. (Ann Arbor, Mich.: Statistical Research Laboratory, University of Michigan, 1976), and *OSIRIS III*, vol. 1: *System and Program Description* (Ann Arbor, Mich.: Institute for Social Research, 1973).
12. A major attempt to create a machine-readable prosopographical database including provenance was BEAR (Berkeley Elites Automated Retrieval), described briefly in *Computers and the Historian* 4(1969):348. In a personal conversation in 1972, the head of the project, David Nasatir, stressed BEAR's almost prohibitively high cost due both to the inclusion of provenance and the storage of materials largely in a natural format.

SOURCES OF INFORMATION IN DATABASE

A. DATA ON CATEGORIES OF INDIVIDUALS INCLUDED IN DATABASE

Archival and Manuscript Sources

1. BAHIA. UNIVERSIDADE. FACULDADE DE MEDICINA. ARQUIVO
List of all graduates.
2. BRASÍLIA. SUPREMO TRIBUNAL FEDERAL. ARQUIVO
Manuscript books entitled "Martrícula dos magistrados."

3. RIO DE JANEIRO. ARQUIVO NACIONAL
Archive of the Ministry of Justice, IJ⁴, nos. 1–9, 16–18, 21–26, 28–30, 32, 34, 38, 45, 52, 333–334; and IJ¹ nos. 707–708, and 822.
4. RIO DE JANEIRO. ARQUIVO NACIONAL
Manuscript listing all members “do conselho.”
5. RIO DE JANEIRO. UNIVERSIDADE FEDERAL. ESCOLA DE ENGENHARIA. ARQUIVO
List of all students at Acadêmia Militar and Escola Central.
6. RIO DE JANEIRO. UNIVERSIDADE FEDERAL. ESCOLA DE MEDICINA. ARQUIVO
Graduation record books.
7. SÃO PAULO. UNIVERSIDADE. FACULDADE DE DIREITO. ARQUIVO.
Original student matriculation records.

Printed Sources

8. BEVILAQUA, CLOVIS
1927 *Historia da Faculdade do Direito do Recife*. Rio de Janeiro.
9. BOITEUX, HENRIQUE
1915 *Os nossos almirantes*. 7 vol. Rio de Janeiro.
10. BRASIL. MINISTERIO DA JUSTICA
1847+ *Relatorio*. Rio de Janeiro.
11. BRASIL. MINISTERIO DA MARINHA
1855+ *Almanak*. Rio de Janeiro.
12. BRASIL. MINISTERIO DA MARINHA
1835+ *Relatorio*. Rio de Janeiro.
13. BRASIL. MINISTERIO DAS RELAÇÕES EXTERIORES
1939 *Ministros e altos funcionarios da antiga Repartição dos Negocios Estrangeiros . . . e membros do extinto Conselho de Estado (1808–1939)*. Rio de Janeiro.
14. BULCÃO SOBRINHO, ANTÔNIO DE ARAÚJO DE ARAGÃO
1961 *Chefes de polícia da Bahia no Império, 1842–1889*. *Revista do Instituto Histórico e Geográfico Brasileiro* (hereafter RIHGB). 253:12–33.
15. BULCÃO SOBRINHO, ANTÔNIO DE ARAÚJO DE ARAGÃO
1962 *A primeira e última legislatura da Câmara dos Deputados do Império*. RIHGB. 254:16–32.
16. CAMPOS, RAUL ADALBERTO DE
1913 *Relações diplomaticas do Brasil, contendo os nomes dos representantes diplomaticos do Brasil no estrangeiro . . . de 1808 a 1912*. Rio de Janeiro.
17. COMISSÃO BRASILEIRA DOS CENTENÁRIOS PORTUGUÊSES
1940 *Os portugueses na Marinha de Guerra do Brasil*. Rio de Janeiro.
18. FERREIRA, WALDEMAR
1928 *A congregação da Faculdade de Direito de S. Paulo na centuria de 1827 a 1927*. São Paulo.
19. FONSECA, LUÍS DA
1949 *Bachareis brasileiros—elementos biográficos (1635–1830)*. *Anais do IV Congresso de História Nacional*. 2:111–405.
20. GALVÃO, MIGUEL ARCHANJO
1894 *Relação dos cidadãos que tomarem parte no governo do Brazil no periodo de março de 1808 a 15 de novembro de 1889*. Rio de Janeiro.
21. INSTITUTO HISTÓRICO E GEOGRÁFICO BRASILEIRO
1839+ *Revista*. Rio de Janeiro.
22. JAVARI, JORGE JOÃO DODSWORTH
1889 *Organizações e programmas ministeriaes desde 1822–1889*. Rio de Janeiro. Also 2nd ed., 1962.
23. LAEMMERT, EDUARDO
1845+ *Almanak administrativo mercantil e industrial da corte e provincia do Rio de Janeiro*. Rio de Janeiro.

24. LAGO, LAURENIO
1936 Dados biographicos dos directores da Secretario de Estado de Guerra. Rio de Janeiro.
25. LAGO, LAURENIO
1944 Conselheiros de Guerra, Vogais e Ministros do Conselho Supremo Militar—Ministros do Supremo Tribunal Militar; dados biográficos, 1808–1943. Rio de Janeiro.
26. MAGALHÃES, FERNANDO DE
1932 O centenario da Faculdade de Medicina do Rio de Janeiro. Rio de Janeiro.
27. MAIA, JULIO JOAQUIM GONÇALVES
1900 Lista geral dos bachareis e doutores formados pela Faculdade de Direito de S. Paulo . . . ate 1900. Revista da Faculdade de Direito de São Paulo. 8:207–91.
28. MARTINS, HENRIQUE
1931 Lista geral dos bachareis e doutores que tem obtido o respectivo grau na Faculdade de Direito do Recife. 2nd ed. Recife.
29. MORAIS, FRANCISCO
1949 Estudantes na Universidade de Coimbra nascidos no Brasil. Coimbra.
30. NOBREGA, APOLONIO CARNEIRO DA CUNHA
1954 Dioceses e bispos do Brasil. RIHGB. 222:3–328.
31. NOGUEIRA, OCTACIANO and JOÃO SERENO FIRMO
1973 Parlamentares do Império. 2 vol. Brasília.
32. PEDROSA, MANOEL XAVIER DE VASCONCELLOS
1959 Estudantes brasileiros na Faculdade de Medicina de Montpellier no fim do século XVIII [1767–1793]. RIHGB. 243:35–71.
33. RHEINGANTZ, CARLOS G.
1960 Titulares do Império. Rio de Janeiro.
34. RIO DE JANEIRO. ASSOCIAÇÃO COMMERCIAL
1868+ Relatório. Rio de Janeiro.
35. RIO DE JANEIRO. BIBLIOTECA NACIONAL
1943 Estudantes brasileiros na Universidade de Coimbra (1772–1872). Rio de Janeiro.
36. RIO DE JANEIRO. ESCOLA POLYTECHNICA
1926 Jubileu da Escola Polytechnica do Rio de Janeiro. Rio de Janeiro.
37. SILVA, ALFREDO PRETEXTATO MACIEL DA
1940 Os generais do Exército brasileiro de 1822 a 1889 (traços biográficos). 3 vol. Rio de Janeiro. (Vol. 3 by Laurenio Lago)
38. SOUSA, ENNES
1909 Brasileiros que estudaram a engenharia de minas, na Academia Real de Minas de Freiburg, na Saxonia. Revista do Instituto Histórico e Geográfico de São Paulo. 14:245–50.
39. STOLS, EDDY
1974 Les étudiants brésiliens en Belgique (1817–1914). Revista de História. 100, pt. 2:653–92.
40. TAUNAY, AFFONSO ESCRAGNOLLE
1942 O Senado do Império. São Paulo.
41. TAUNAY, AFFONSO ESCRAGNOLLE
1950 A Câmara dos Deputados sob o Império. Anais do Museu Paulista. 14:1–252.
42. TORRES, OCTAVIO
1946 Esboço histórico dos acontecimentos mais importantes da vida da Faculdade de Medicina da Bahia (1808–1946). Salvador.
43. VAMPRE, SPENCER
1924 Memorias para a historia da Academia de São Paulo. 2 vol. São Paulo.
44. VASCONCELOS, ANTÔNIO DE
1938 Escritos varios relativos á universidade dionisiana. 3 vol. Coimbra.
45. ZUQUETTE, AFONSO EDUARDO MARTINS
1960 Nobreza de Portugal. 3 vol. Lisboa.

B. SOURCES OF INFORMATION FOR CODES

46. ALMEIDA, CANDIDO MENDES DE
1868 Atlas do Imperio do Brazil. Rio de Janeiro.
47. BARBOSA, WALDEMAR DE ALMEIDA
1971 Dicionário histórico-geográfico de Minas Gerais. Belo Horizonte.
48. BETTENCOURT, E. A. DE
1885 Dictionario chorographico de Portugal e ilhas adjacentes. Lisboa.
49. FORTES, AMYR BORGES et al
1966 Dicionário geográfico brasileiro. Porto Alegre.
50. INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATISTICA
1971 Índice dos topônimos da carta da Brasil ao milionésimo. Rio de Janeiro.
51. INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATISTICA
1972 Carta do Brasil ao milionésimo. Rio de Janeiro.
52. SOUSA, J. GALANTE DE
1963 Índice de biobibliografia brasileira. Rio de Janeiro.
53. WORLD HEALTH ORGANIZATION
1967 International Classification of Diseases. 8th ed. Vol. 1. Geneva.

C. MAJOR COLLECTIVE BIOGRAPHIES

(Also see items 9, 18, 24, 30, 37, and 45)

54. ABRANCHES, DUNSHEE DE
1918 Governos e congressos da Republica dos Estados Unidos do Brazil. 2 vol. São Paulo.
55. ALEGRE, ACHYLLES PORTO
1918 Homems illustres do Rio Grande do Sul. 2nd ed. Pôrto Alegre.
56. ALMEIDA, ANTÔNIO DA ROCHA
1961 Vultos da pátria. 4 vol. Pôrto Alegre.
57. BARBOSA, RAYMUNDO RODRIGUES
1952 Historia do Supremo Tribunal Militar. Rio de Janeiro.
58. BITTENCOURT, AGNELLO
1973 Dicionário amazonense de biografias. Rio de Janeiro.
59. BITTENCOURT, LIBERATO
1917 Homems do Brazil. Vol. 1 (Sergipe) and vol. 2 (Parahyba). Rev. ed. Rio de Janeiro.
60. BLAKE, AUGUSTO VICTORINO ALVES SACRAMENTO
1883 Dictionario bibliographico brasileiro. 7 vol. Rio de Janeiro.
61. BOITEUX, LUCAS ALEXANDRE
1933 Ministros da Marinha; notas biographicos. 3 vol. Rio de Janeiro.
62. BRASIL. MINISTERIO DA VIAÇÃO E OBRAS PÚBLICAS
1962 Dados biográficos dos ministros. Rio de Janeiro.
63. CALMON, PEDRO
1972 História do Ministério da Justiça. Vol. 1. Rio de Janeiro.
64. CARMO, JOSÉ A. PINTO DO
1944 Ministros da Fazenda. Rio de Janeiro.
65. CARNEIRO, DAVID
1960 História do periodo provincial do Paraná (Galeria de Presidentes da provincia). Curitiba.
66. CARVALHO, MARIA TEIXEIRA DE
1937 Nobiliario sulriograndense. Pôrto Alegre.
67. CASCUDO, LUIS DA CAMARA
1939 Governo do Rio Grand do Norte. Natal.
68. CORRÊA FILHO, VIRGILIO
1954 Grandes vultos da nossa engenharia ferroviária. In: 1° [Primeiro] centenário dos ferrovias brasileiros. Rio de Janeiro.

69. COSTA, FRANCISCO AUGUSTO PEREIRA DA
1882 Dicionario biographico de pernambucanos celebres. Recife.
70. CUNHA, RAYMUNDO CYRIACO ALVES DA
1970 Paraenses illustres. 2nd ed. Belem.
71. EGAS, EUGENIO
1926 Galeria dos Presidentes de São Paulo. 3 vol. São Paulo.
72. GALVÃO, SEBASTIÃO DE VASCONCELLOS
1908 Dicionario chorographico, historico e estatistico de Pernambuco. 4 vol. Rio de Janeiro.
73. GUARANÁ, MANOEL ARMINDO CORDEIRO
1925 Dicionario bio-bibliographico sergipano. Rio de Janeiro.
74. GUIMARÃES, ARGEU
1938 Dicionario bio-bibliographico brasileiro de diplomacia, politica exterior e direito internacional. Rio de Janeiro.
75. JORNAL DO BRASIL.
1931 Galeria nacional. 10 vol. Rio de Janeiro.
76. LACOMBE, LOURENÇO LUIZ
1973 Os chefes do executivo fluminense. Petrópolis.
77. LAGO, LAURENIO
1940 Supremo Tribunal de Justiça e Supremo Tribunal Federal; dados biográficos, 1828–1939. Rio de Janeiro.
78. LAGO, LAURENIO
1949 Nobiliarquia brasileira: titulares do sexo feminino. Anais do IV Congresso de História Nacional. 2:47–66.
79. LAGO, LAURENIO
1954 Acréscimos e retificações ao Arquivo Nobiliárquico. Anuário do Museu Imperial 15:83–219.
80. LEAL, ANTONIO HENRIQUES
1873 Pantheon marenhense; ensaios biographicos. 4 vol. Lisboa.
81. LOPES, THEODORICO and GENTIL TORRES
1946 Ministros da Guerra do Brasil, 1808–1945. Rio de Janeiro.
82. LYRA, AUGUSTO TAVARES DE
1926 O Senado do Imperio. RIHGB. 153:229–70.
83. LYRA, AUGUSTO TAVARES DE
1946 Os Ministros de Estado da independência à republica. RIHGB. 193:3–104.
84. MACEDO, JOAQUIM MANUEL DE
1876 Anno biographico brasileiro. 4 vol. Rio de Janeiro.
85. MARQUES, MANUEL EUFRAZIO DE AZEVEDO
1879 Apontamentos historicos, geographicos, biographicos, estatisticos e noticiosos da provincia de S. Paulo. 2 vol. Rio de Janeiro.
86. MENDONÇA, RUBENS DE
1953 Dicionário biografico mato-grossense. São Paulo.
87. MORAIS, TANCREDO
1954 Resumo histórico antropogeográfico do estado de Alagoas. Rio de Janeiro.
88. MOYA, SALVADOR DE
1939+ Anuário genealógico brasileiro. 10 vol. São Paulo.
89. NOBREGA, APOLONIO CARNEIRO DA CUNHA
1964 Bacharéis paraibanos de Olinda e Recife de 1832 a 1960. RIHGB. 262:3–240.
90. NOGUEIRA, JOSÉ LUIZ DE ALMEIDA
1907 A Academia de São Paulo; tradições e reminiscencias. 9 vol. São Paulo.
91. NONATO, RAIMUNDO
1960 Bacharéis de Olinda e Recife; norte-riograndenses formados de 1832 a 1932. Rio de Janeiro.
92. PARANHOS, JOSÉ MARIA DA SILVA (BARÃO DO RIO BRANCO)
1946 Efemérides brasileiros. Rio de Janeiro.

93. RIBEIRO, JOSÉ JACINTO
1899 Chronologia paulista. 2 vol. São Paulo.
94. SANTOS, PRESALINDO DE LERY
1880 Pantheon fluminense. Rio de Janeiro.
95. SILVA, INOCENCIO FRANCISCO DA
1858 Dicionario bibliographico portuguez. 22 vol. Lisboa.
96. SILVA, MANUEL FRANCISCO DIAS DA
1871 Dicionario biographico de brasileiros celebres. . . . Rio de Janeiro.
97. SISSON, SEBASTIÃO AUGUSTO
1861 Galeria dos brasileiros illustres (os contemporaneos). 2 vol. Rio de Janeiro.
98. SOUZA, ANTÔNIO LOUREIRO DE
1949 Bahianos illustres, 1564–1935. Salvador.
99. STUDART, GUILHERME
1910 Dicionario bio-bibliographico cearense. 3 vol. Fortaleza.
100. VASCONCELLOS, RODOLPHO SMITH DE
1918 Archivo nobiliarchico brasileiro. Lausanne.
101. VEIGA, JOSÉ PEDRO XAVIER DA
1897 Ephemérides mineiros (1664–1897). 4 vol. Ouro Preto.
102. VELHO SOBRINHO, JOÃO FRANCISCO
1937 Dicionario bio-bibliographico brasileiro. 2 vol. Rio de Janeiro.
103. WERNECK, FRANCISCO KLORS
1947 História e genealogia fluminense. Rio de Janeiro.
104. WILDBERGER, ARNOLD
1949 Os Presidentes da província da Bahia. Salvador.