Database Developments
HISTORICAL METHODS, July–September 2010, Volume 43, Number 3

Routledge
Taylor & Francis Group

Constructing Relational Databases to Study Life Histories on Your PC or Mac

JOHANN PETER MURMANN
Australian School of Business
University of New South Wales

Abstract. In this article, the author presents a strategy for designing relational databases with the program FileMaker Pro (FileMaker) to study the histories of individuals and organizations. The approach facilitates efficiency in inputting data and flexibility for constructing statistical analyses from the raw data. The key feature of the strategy is to define the basic unit of observation in the database in terms of an agent, an event, and a date. Given that programs such as FileMaker can easily sort data by agent and date, once one structures the data correctly, he or she can construct well-ordered event histories for agents, even if the researcher enters the data in an unordered fashion. By using events that happened to an agent at a particular time as the basic unit of observation, one maintains maximum flexibility to do statistical analysis that aggregates basic data in different ways. This article illustrates the power of the approach by outlining ways to analyze changes in geographic distances between two events marking the life histories of chemists.

Keywords: aids for historical research, combining qualitative and quantitative data, event history analyses, relational databases, software tools

In a previous article in this journal, George C. Alter and Myron P. Gutmann (1999) provided a very useful overview of how to do formal event history analyses. But reflecting on the more basic steps of collecting, storing, and retrieving information in computer databases, they noted, "All of the projects with which we are familiar have required the services of skilled programmers to accomplish this task, and some researchers have done much of this work by hand. A need definitely exists for the development of advanced database management techniques in this area" (ibid., 172).

My goal in developing this technical note was to begin to fill this gap and share the strategies I have developed for collecting and storing together quantitative and qualitative information on social change. I will explain these database management strategies using the example of a new research project on the career histories of chemists. But I will also more generally extrapolate how these strategies are useful for all scholars who want to collect data on any entities that can be construed to have histories, regardless of whether the entities are people, groups, firms, organizations, nation states, and the like. As will become more evident later, it is *not* necessary to want to do formal event-history analyses to benefit from these strategies. The goal may simply be to present descriptive analyses of historical changes.

Trained as a social scientist who strives for representative samples and statistical power, I wanted to make my data readily analyzable with statistical software packages such as SAS, SPSS, and STATA.[1] Because it is so easy to import data from a spreadsheet program such as Microsoft Excel into these statistical software packages, it may seem natural for researchers to collect data using a spreadsheet. However, entering large amounts of data directly into a spreadsheet program is often an unwieldy process. Software solutions of professional statisticians such as SAS can be programmed to provide data entry interfaces, bypassing the need to use a spreadsheet program. I have found it much easier to work with FileMaker Pro (version 8.5; FileMaker), a database program that unlike Microsoft Access is user-friendly and available for both PCs and Mac.

Since my doctoral-student days, I have been designing electronic databases that track the development of firms and plants using both quantitative and qualitative data. For a new research project on the career histories of all German chemists from 1850 to 1914, I explored ways to make database construction more efficient while preserving maximum flexibility for subsequent work with the data. On completion of the data-entry process, I wanted to be able to aggregate and slice the stored data in different ways, answering analytic questions that I or other people accessing the database may not have foreseen at the outset of the project.

I was interested in constructing the career history of every German chemist before World War I to test statistically whether an argument I formulated with anecdotal evidence in my book on the synthetic-dye industry was correct. Based on nontrivial but nonetheless anecdotal evidence, I argued in *Knowledge and Competitive Advantage* (Murmann 2003) that synthetic-dye firms connected to leading chemists possessed important competitive advantages over their

Address correspondence to Johann Peter Murmann, Australian School of Business, University of New South Wales, Sydney NSW 2052, Australia. E-mail: hm@professor-murmann.net

	A	B
1	**Name**	**Career Description**
2	Baeyer, Adolf	Born: 31-Oct-1835 Birthplace: Berlin, Germany Died: 20-Aug-1917 Location of death: Starnberg, Germany University: PhD, University of Berlin (1858) Teacher: Privatdozent, University of Berlin (1860-66) Professor: University of Berlin (1866-) Professor: Gewerbe-Akademie, Berlin Professor: University of Strassburg (1871-75) Professor: University of Munich (1875-) Nobel Prize for Chemistry 1905 Davy Medal 1881
3	Hofmann, August W.	BIRTH: 1818 Giessen, Hesse-Darmstadt, Germany DEATH: 1892 Berlin, Germany Education and Qualifications 1843 PhD Awarding institution: Giessen University Institution of study: Career 1845 - 1865 Professor, Royal College of Chemistry 1845 - 1845 Associate Professor, Bonn University 1865 - 1892 Professor, Berlin University Memberships and Roles Institution: Position: Year(s): Chemical Society of London Fellow 1845 - 0 Chemical Society of London President 1861 - 1863 Chemical Society of London Past-President
4		

FIGURE 1. One record per chemist.

competitors lacking such ties. A few chemists, such as August Wilhelm Hofmann and Adolf Baeyer, seemed to occupy central positions in what I dubbed the *organic chemistry knowledge network*, which connected chemists at universities, governmental laboratories, and industrial firms.

Sociologists have developed sophisticated analytical tools to statistically examine the structure of social networks and their impact on the performance of individuals, firms, or even larger social aggregates. For examples, see the publications by Stanley Wasserman and Katherine Faust (1994) and John Scott (2000). One of my goals in constructing the chemists' database (Murmann 2010a, 2010b)[2] was to use the program UCINET (Analytic Technologies) to examine whether it was true that firms connected to more central chemists did better than firms that only employed chemists peripheral to the knowledge network. Hence, the electronic database I wanted to construct needed to be readily able to generate the quantitative network data required for UCINET. At the same time, I wanted to store more descriptive information in the database so that the database could become a valuable resource for all

	A	B	C	D	E	F
1	Person Name	Year	Description	University Name	Town Name	Country Name
2	Adolph Baeyer	1860	Works at University	University Berlin	Berlin	Germany
3	Adolph Baeyer	1861	Works at University	University Berlin	Berlin	Germany
4	Adolph Baeyer	1862	Works at University	University Berlin	Berlin	Germany
5	Adolph Baeyer	1863	Works at University	University Berlin	Berlin	Germany
6	Adolph Baeyer	1864	Works at University	University Berlin	Berlin	Germany
7	Adolph Baeyer	1865	Works at University	University Berlin	Berlin	Germany
8	Adolph Baeyer	1866	Works at University	University Berlin	Berlin	Germany
9	Adolph Baeyer	1867	Works at University	University Berlin	Berlin	Germany
10	Adolph Baeyer	1868	Works at University	University Berlin	Berlin	Germany
11	Adolph Baeyer	1869	Works at University	University Berlin	Berlin	Germany
12	Adolph Baeyer	1870	Works at University	University Berlin	Berlin	Germany
13	Adolph Baeyer	1871	Works at University	University Berlin	Berlin	Germany
14	Adolph Baeyer	1872	Works at University	University Berlin	Berlin	Germany
15	Adolph Baeyer	1873	Works at University	University Strassburg	Strassburg	Germany
16	Adolph Baeyer	1874	Works at University	University Strassburg	Strassburg	Germany
17	Adolph Baeyer	1875	Works at University	University Strassburg	Strassburg	Germany
18	Adolph Baeyer	1876	Works at University	University Strassburg	Strassburg	Germany
19	Adolph Baeyer	1877	Works at University	University Munich	Munich	Germany
20	Adolph Baeyer	1878	Works at University	University Munich	Munich	Germany
21	Adolph Baeyer	1879	Works at University	University Munich	Munich	Germany
22	Adolph Baeyer	1880	Works at University	University Munich	Munich	Germany
23	Adolph Baeyer	1881	Works at University	University Munich	Munich	Germany
24	Adolph Baeyer	1882	Works at University	University Munich	Munich	Germany
25	Adolph Baeyer	1883	Works at University	University Munich	Munich	Germany
26	Adolph Baeyer	1884	Works at University	University Munich	Munich	Germany
27	Adolph Baeyer	1885	Works at University	University Munich	Munich	Germany
28	Adolph Baeyer	1886	Works at University	University Munich	Munich	Germany
29	Adolph Baeyer	1887	Works at University	University Munich	Munich	Germany
30	Adolph Baeyer	1888	Works at University	University Munich	Munich	Germany
31	Adolph Baeyer	1889	Works at University	University Munich	Munich	Germany
32	Adolph Baeyer	1890	Works at University	University Munich	Munich	Germany
33	Adolph Baeyer	1891	Works at University	University Munich	Munich	Germany
34	Adolph Baeyer	1892	Works at University	University Munich	Munich	Germany
35	Adolph Baeyer	1893	Works at University	University Munich	Munich	Germany
36	Adolph Baeyer	1894	Works at University	University Munich	Munich	Germany
37	Adolph Baeyer	1895	Works at University	University Munich	Munich	Germany
38	Adolph Baeyer	1896	Works at University	University Munich	Munich	Germany
39	Adolph Baeyer	1897	Works at University	University Munich	Munich	Germany
40	Adolph Baeyer	1898	Works at University	University Munich	Munich	Germany
41	Adolph Baeyer	1899	Works at University	University Munich	Munich	Germany
42	Adolph Baeyer	1900	Works at University	University Munich	Munich	Germany
43	Adolph Baeyer	1901	Works at University	University Munich	Munich	Germany
44	Adolph Baeyer	1902	Works at University	University Munich	Munich	Germany
45	Adolph Baeyer	1903	Works at University	University Munich	Munich	Germany
46	Adolph Baeyer	1904	Works at University	University Munich	Munich	Germany
47	Adolph Baeyer	1905	Works at University	University Munich	Munich	Germany
48	Adolph Baeyer	1906	Works at University	University Munich	Munich	Germany
49	Adolph Baeyer	1907	Works at University	University Munich	Munich	Germany
50	Adolph Baeyer	1908	Works at University	University Munich	Munich	Germany
51	Adolph Baeyer	1909	Works at University	University Munich	Munich	Germany
52	Adolph Baeyer	1910	Works at University	University Munich	Munich	Germany
53	Adolph Baeyer	1911	Works at University	University Munich	Munich	Germany
54	Adolph Baeyer	1912	Works at University	University Munich	Munich	Germany
55	Adolph Baeyer	1913	Works at University	University Munich	Munich	Germany
56	Adolph Baeyer	1914	Works at University	University Munich	Munich	Germany
57	Adolph Baeyer	1915	Works at University	University Munich	Munich	Germany

FIGURE 2. One record per chemist per year. Shaded entries do not add new information.

historians and social scientists interested in the career histories of German chemists.

The remainder of this article will unfold as follows: First, in *The Fundamental Issues in Constructing a Database*, I argue that the first task for any database management effort involves deciding what the fundamental unit of analysis will be. I advocate that using an agent, an event and a date as this unit creates not only efficiencies but also preserves highly desirable flexibilities in the data collection effort. Second, in *Relational Database to Reduce Labor Time in Entering Data*, I describe how to set up in FileMaker a relational database design that can save substantial labor in the data-entry process. Third, in *Key Challenges in Setting Up Relational Database*, I provide a tutorial on how to solve the only somewhat challenging task in FileMaker of setting up a many-variables-to-one-variable mapping between two distinct database tables. This problem is encountered, for example, when you want to add multiple reference sources into an event record by pulling all the reference information from a related list. Fourth, in *Application*, I outline an application of the relational database design method for the problem of analyzing how geographic distances between a chemist's location of study and his or her first job changed over time. Fifth, in *Conclusion*, I end this article by pointing out that the method of making an agent, event, and date the fundamental unit of observation also can be applied to objects not commonly construed as agents, such as newspaper articles, elections, and the like.

The Fundamental Issues in Constructing a Database

In designing an electronic database, a researcher has to decide what will be the most fundamental unit of observation. When, with Ernst Homburg, I constructed the database to record the development of synthetic-dye firms, the basic unit of observation was a particular firm in a particular year. As I was setting out to design the database for the chemists' career histories, it was not obvious what the fundamental unit of observation should be. Having one record per chemist containing all the information on him or her would make it very difficult to do statistical analysis with the database, without doing a lot more time-consuming work in reorganizing the data into a new format. (See figure 1.) Creating one record for every year in which each chemist lived, even though no new information might be available for a chemist in a particular year, would be a very inefficient process. (See figure 2.) Furthermore, there might be several events in a particular chemist's life in a particular year. Finding a way to store multiple events in each particular year in one record

	A	B	C	D	E	F
1	Person Name	Event Date	Event Type Descripton	University Name	Town Name	Country Name
2	Adolph Baeyer	18351031	Birth		Berlin	Germany
3	Adolph Baeyer	18600701	Starts employment at university	University Berlin	Berlin	Germany
4	Adolph Baeyer	18700101	Appears in directory		Berlin	Germany
5	Adolph Baeyer	18720630	Ends employment at university	University Berlin	Berlin	Germany
6	Adolph Baeyer	18720701	Starts employment at university	University Strassburg	Strassburg	Germany
7	Adolph Baeyer	18750630	Ends employment at university	University Strassburg	Strassburg	Germany
8	Adolph Baeyer	18750701	Starts employment at university	University Munich	Munich	Germany
9	Adolph Baeyer	18770101	Appears in directory		Munich	Germany
10	Adolph Baeyer	18860101	Appears in directory		Munich	Germany
11	Adolph Baeyer	18920101	Appears in directory		Munich	Germany
12	Adolph Baeyer	19010101	Appears in directory		Munich	Germany
13	Adolph Baeyer	19060101	Appears in directory		Munich	Germany
14	Adolph Baeyer	19140101	Appears in directory		Munich	Germany
15	Adolph Baeyer	19150630	Ends employment at university	University Munich	Munich	Germany
16	Adolph Baeyer	19170800	Death		Munich	Germany
17	August Wilhelm von Hofmann	18180408	Birth		Giessen	Germany
18	August Wilhelm von Hofmann	18360701	Starts study at university	University Giessen	Giessen	Germany
19	August Wilhelm von Hofmann	18410630	Ends study at university	University Giessen	Giessen	Germany
20	August Wilhelm von Hofmann	18410701	Receives (university) degree	University Giessen	Giessen	Germany
21	August Wilhelm von Hofmann	18430701	Starts employment at institute			Germany
22	August Wilhelm von Hofmann	18450428	Receives (university) degree		Bonn	Germany
23	August Wilhelm von Hofmann	18450630	Starts employment at university	University Bonn	Bonn	Germany
24	August Wilhelm von Hofmann	18450701	Starts employment at university	Royal College of Chemistry	London	Great Britain
25	August Wilhelm von Hofmann	18640201	Starts employment at university	University Berlin	Berlin	Germany
26	August Wilhelm von Hofmann	18700101	Appears in directory		Berlin	Germany
27	August Wilhelm von Hofmann	18770101	Appears in directory		Berlin	Germany
28	August Wilhelm von Hofmann	18860101	Appears in directory		Berlin	Germany
29	August Wilhelm von Hofmann	18920101	Appears in directory		Berlin	Germany
30	August Wilhelm von Hofmann	18920502	Death		Berlin	Germany

FIGURE 3. One record per distinct event in the life of a chemist.

of the database and still make it easy to analyze the data statistically seemed difficult. For example, within a short period, a chemist might graduate from university and then take employment in a particular firm. I needed to choose the fundamental unit of analysis so that my research team would have to enter as little data as possible by hand while still maintaining the greatest flexibility for slicing and aggregating data in different ways later. It is always possible to aggregate data (e.g., to calculate the average age of a group of chemists), but it is impossible to move from aggregated data (the average age of a group of chemists) to the basic units of information that went into an aggregated statement (the ages of the chemists making up the group).

The fundamental unit of information in a database can have multiple pieces of information. You can add categories of information to a fundamental unit later in a project, but at the outset, you have to decide what the fundamental unit of observation will be. In the context of a spreadsheet file, the fundamental unit of observation would be a row in the spreadsheet, and the different pieces of information connected to this fundamental observation would be the columns in the spreadsheet.

After considerable reflection, I realized that the most efficient and flexible database design would be to define the fundamental unit of observation as a specific event in the career of a chemist (see figure 3). The minimum information associated with every record would be a specific event (e.g., birth, graduation from university, onset of first job, death) in the life of a specific chemist (e.g., August Wilhelm Hofmann) on a specific date (e.g., January 1, 1870).[3] As I will subsequently discuss, other information can be added to each basic unit of observation.

There are several significant advantages in using each event in the career of a chemist as the basic unit of observation and analysis.

1. It is easy to analyze statistically particular event types. For example, you can determine the total entry of chemists into the labor force in a given year or decade by calculating the entries of all individual chemists.
2. The amount of data that needs to be entered by hand into a database is limited.
   a. Using a relational database design, you can enter a numerical code for an event type (e.g., 1 = person is born), and the program will automatically connect

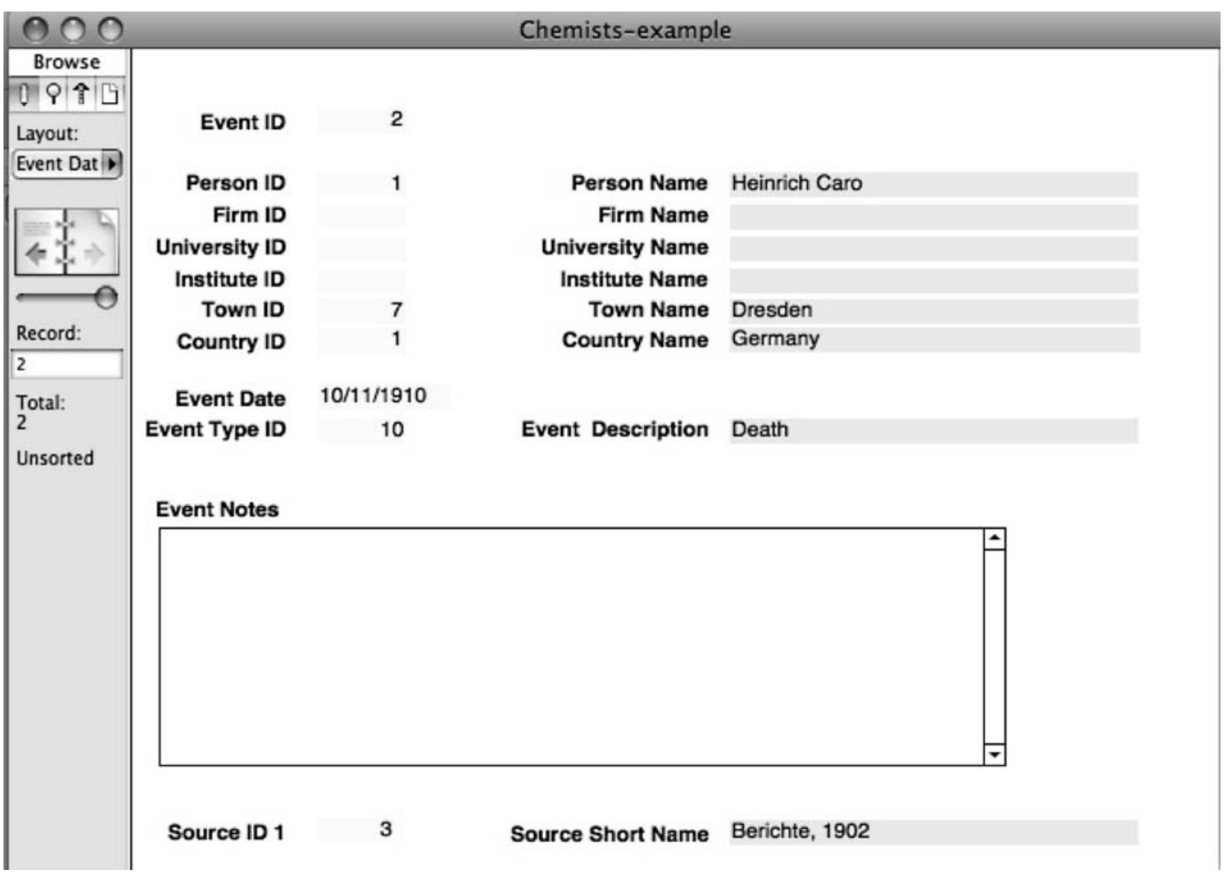


FIGURE 4. Main interface window of the chemists' event history database.

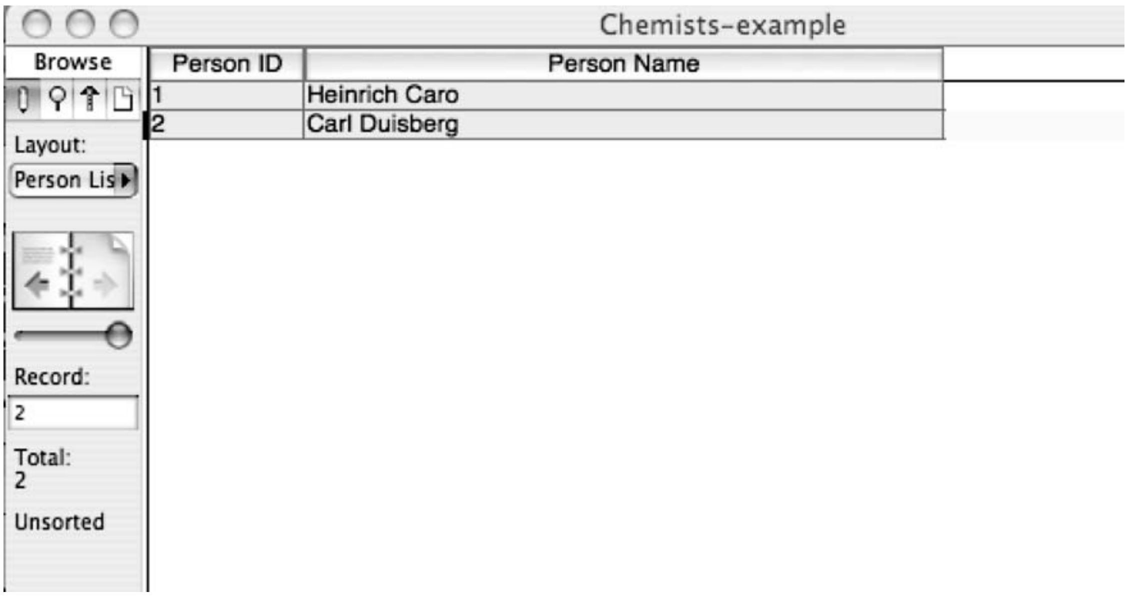


FIGURE 5. Related person list database.

it to the verbal description of the event. No text will have to be entered in a new record.

b. A computer program such as SAS can automatically fill in records for intervening years when this is required for particular analytical purposes. For example, if the database contains a record that a chemist worked at the Bayer Company in 1870 and a record that he or she worked at the firm in 1875, you can let a computer program fill in records for 1871, 1872, 1873, and 1874.

3. You do not have to enter data for a particular chemist chronologically. Besides directories of chemists, information on events in a career of a chemist comes from many diverse sources. You can first add data from one source and then add data from other sources that may refer to earlier events. Just as a detective often receives clues in a disorderly fashion, a scholar drawing on many sources encounters information on a particular agent in a disorderly fashion. Even when data on a chemist's career events are not entered chronically, it is very

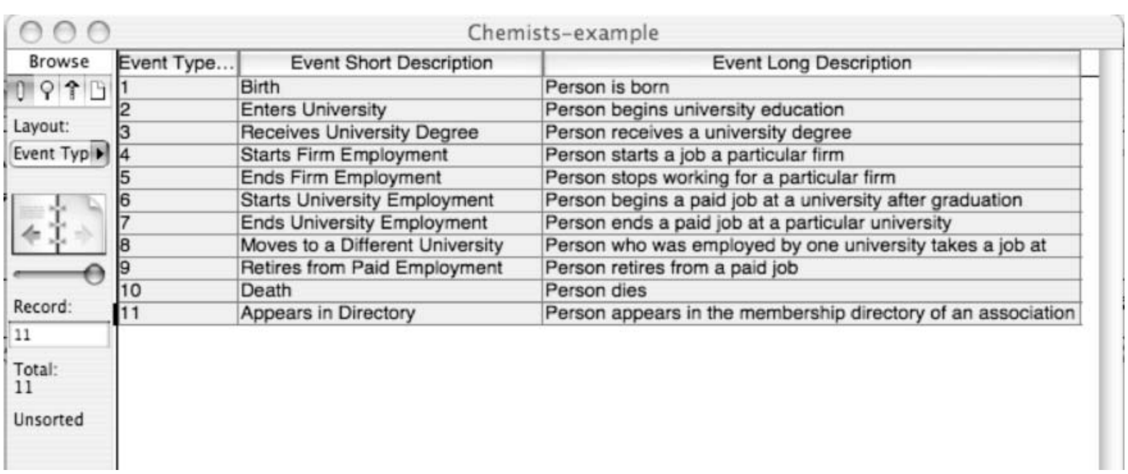


FIGURE 6. Related event–type database.

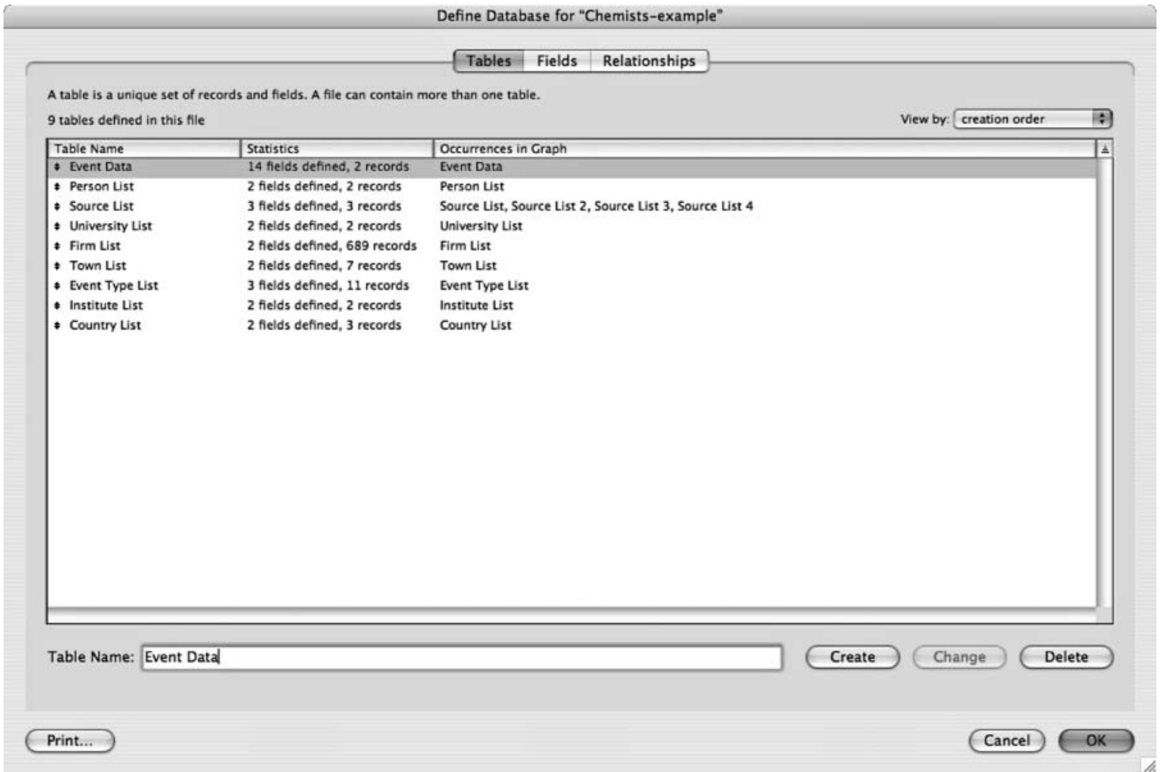
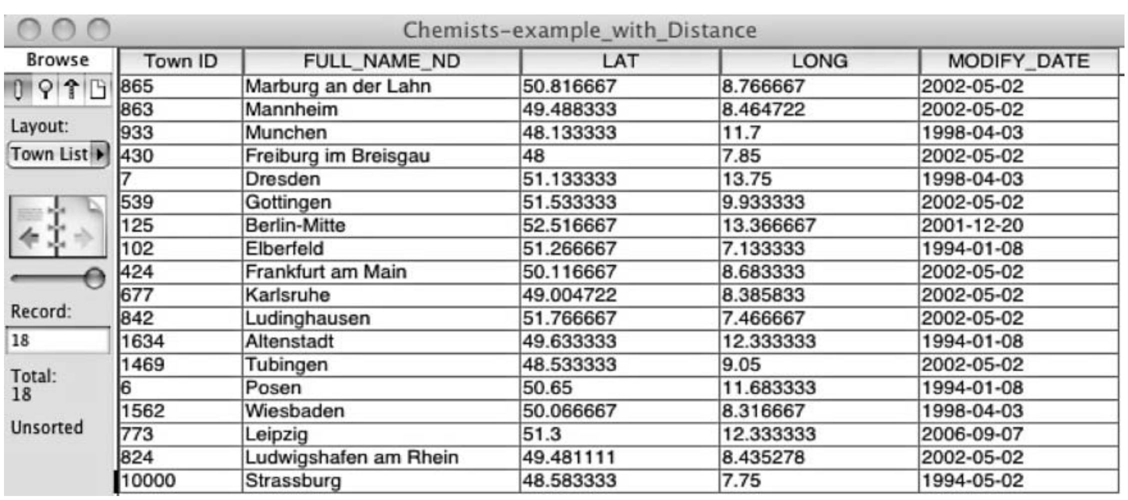


FIGURE 7. Inventory of tables related in database.

simple to create a chronology of a chemist's career by sorting the database based on date.

4. By using events in the life of an agent such as a chemist rather than the entire life or periods in the agent's life as the basic unit of observation, you maintain maximum flexibility to be able to aggregate the data for the different analytical questions you can address with such microdata.



Town ID	FULL_NAME_ND	LAT	LONG	MODIFY_DATE
865	Marburg an der Lahn	50.816667	8.766667	2002-05-02
863	Mannheim	49.488333	8.464722	2002-05-02
933	Munchen	48.133333	11.7	1998-04-03
430	Freiburg im Breisgau	48	7.85	2002-05-02
7	Dresden	51.133333	13.75	1998-04-03
539	Gottingen	51.533333	9.933333	2002-05-02
125	Berlin-Mitte	52.516667	13.366667	2001-12-20
102	Elberfeld	51.266667	7.133333	1994-01-08
424	Frankfurt am Main	50.116667	8.683333	2002-05-02
677	Karlsruhe	49.004722	8.385833	2002-05-02
842	Ludinghausen	51.766667	7.466667	2002-05-02
1634	Altenstadt	49.633333	12.333333	1994-01-08
1469	Tubingen	48.533333	9.05	2002-05-02
6	Posen	50.65	11.683333	1994-01-08
1562	Wiesbaden	50.066667	8.316667	1998-04-03
773	Leipzig	51.3	12.333333	2006-09-07
824	Ludwigshafen am Rhein	49.481111	8.435278	2002-05-02
10000	Strassburg	48.583333	7.75	1994-05-02

FIGURE 8. Town list.

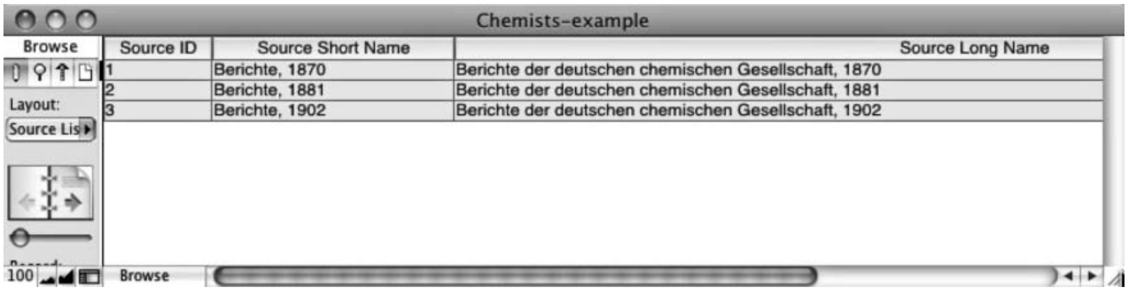


FIGURE 9. Source list table.

Please note that the four arguments certainly are not restricted to the case of chemists. They are equally applicable to any type of agent that can be conceptualized as having a history: a person, an organization, a state, and the like. To summarize, using as the fundamental unit of observation an agent, an event, and a date preserves a high amount of flexibility for one research project because you can later add additional event types or even entirely new variable types.

Relational Database to Reduce Labor Time in Entering Data

To reduce the work required in entering data, you can use a program such as FileMaker to create what is technically called a *relational database*. Conceptually, a relational database is the same as linked sheets in Excel. Again, I will use the example of the chemists' database to illustrate how such relational database designs reduce the work required in entering data. I am clearly not the first one to point out the

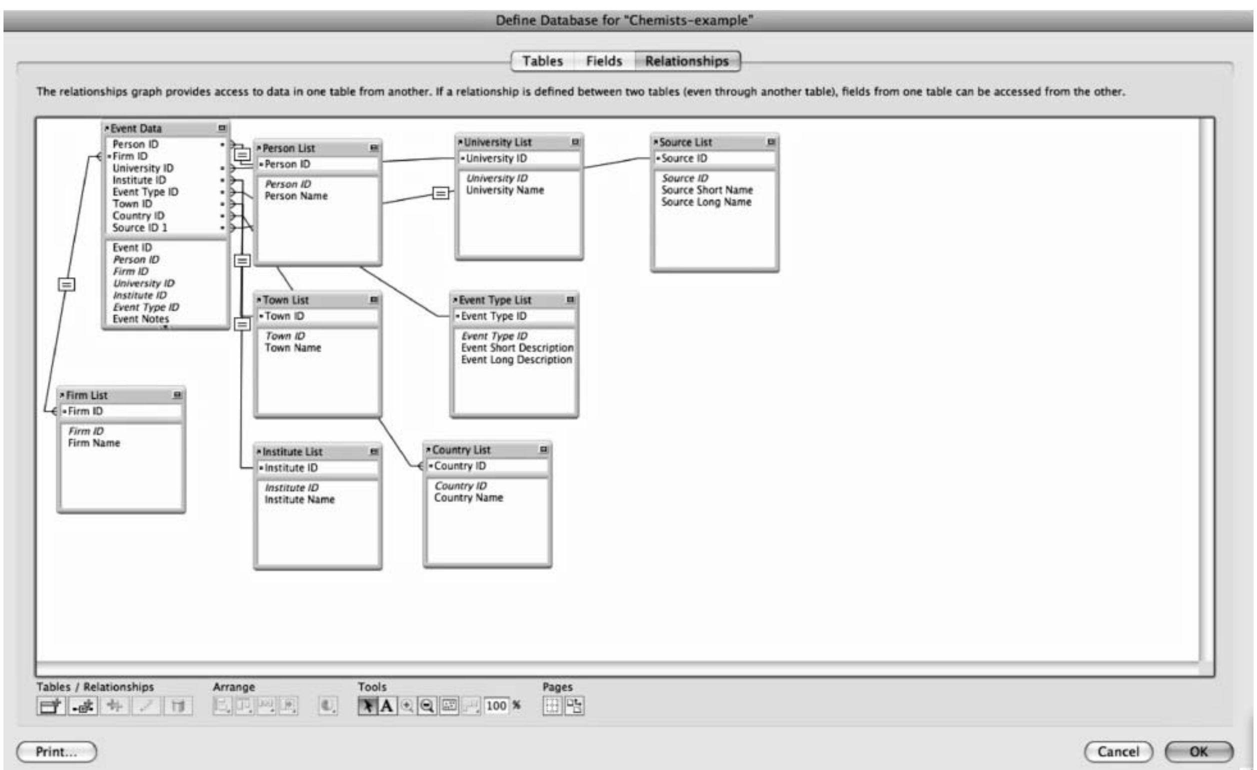


FIGURE 10. Relationships viewer of all tables in chemists database.

advantage of a relational database. My goal here is to describe how you can set up a relational database design in FileMaker and how you can potentially use already-collected information by importing that data as a table into FileMaker. As indicated earlier, the basic unit of observation in that database is a particular agent (e.g., the chemist Adolph Baeyer), a particular event (e.g., the birth of the person), and a particular date (e.g., February 13, 1834). This is the minimum information required for each record to be well formed.

Figure 4 shows the main data-entry interface for the chemists' event history database. This main database table has more fields than previously mentioned because I wanted to track additional information about individual chemists and their affiliation. The fields with a white background (left-side) indicate data that need to be entered by hand. The more darkly shaded fields (right side) are the corresponding pieces of information that are generated automatically from related databases (tables). (In FileMaker Pro 8.5, related databases are no longer stored in separate files but in one file, just as nowadays different spreadsheets can be stored as multiple sheets in the same file.)

To record the order in which data are entered in the database, you need to define the first entry field as an *event ID*, giving each record in the database a unique ID. Next comes the *person ID*. Every time data is entered for a particular person who does not already appear in the database, you need to enter in a related table, called "person list," a new person ID and the person name.

Once a person ID is created for a particular chemist and associated with a "person name," you can enter the person ID in the main event database table. An associated name will show up if you have linked the two tables in the database.

As previously mentioned, the two other required pieces of information for every record are the *event type* and the *event date*. Again, to create efficiency in the data-entry process, I set up a related table of event types that associates a particular numerical value with a particular event type. Entering the numerical value for the event in the main event history database pulls in the associated text description from the related event type database table.

All of the other ID fields (firm, town, source, etc.) work in the same way by pulling in information from a related table.

Once upon a time, before the PC was ubiquitous, scholars used index cards to record information. Aside from displaying information in the typical row-column format of a spreadsheet, as you see in figure 5 and figure 6, FileMaker has the ability to display information like an index card (see figure 4). A big note field to record any information useful to describe an event in the career of a chemist in more detail can be created. Large text fields make it possible to store a lot of qualitative information along with simple numerical information that characterizes most of the fields in the main table.

Figure 7 provides a summary of all related tables that make up the chemists' database.

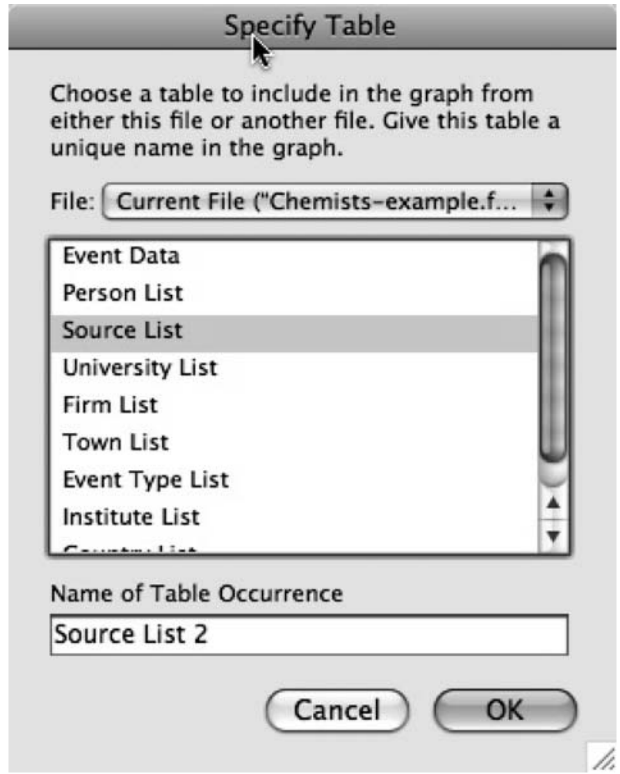


FIGURE 11. Creating a copy of a table window.

As one reviewer of this article pointed out, many sociologists, anthropologists, and historians are getting interested in computer mapping of individuals' travels, encounters, ships' journeys, and the like. The latitudinal and longitudinal information of the precise locations of towns that allows you to calculate the distance between two locations in FileMaker (see figure 8) or any other program is already available for download from Web sites (e.g., National Geospatial-Intelligence Agency).[4] Instead of entering this information by hand into the FileMaker database, you can simply set up a town table and import the location coordinates into the table from the publicly available data files.

For the purpose of the chemists' database, I imported this data into a table, called them *Town List*, added a variable for town ID, and then brought this information into the event table by making two relational links through the town ID variable. Before I explain how to analyze changes in geographic patterns in the *Application* section, I first discuss how to pull information connected to two different values of the same variable from one table into another table. This was the only challenge that I encountered in setting up a relational database in FileMaker.

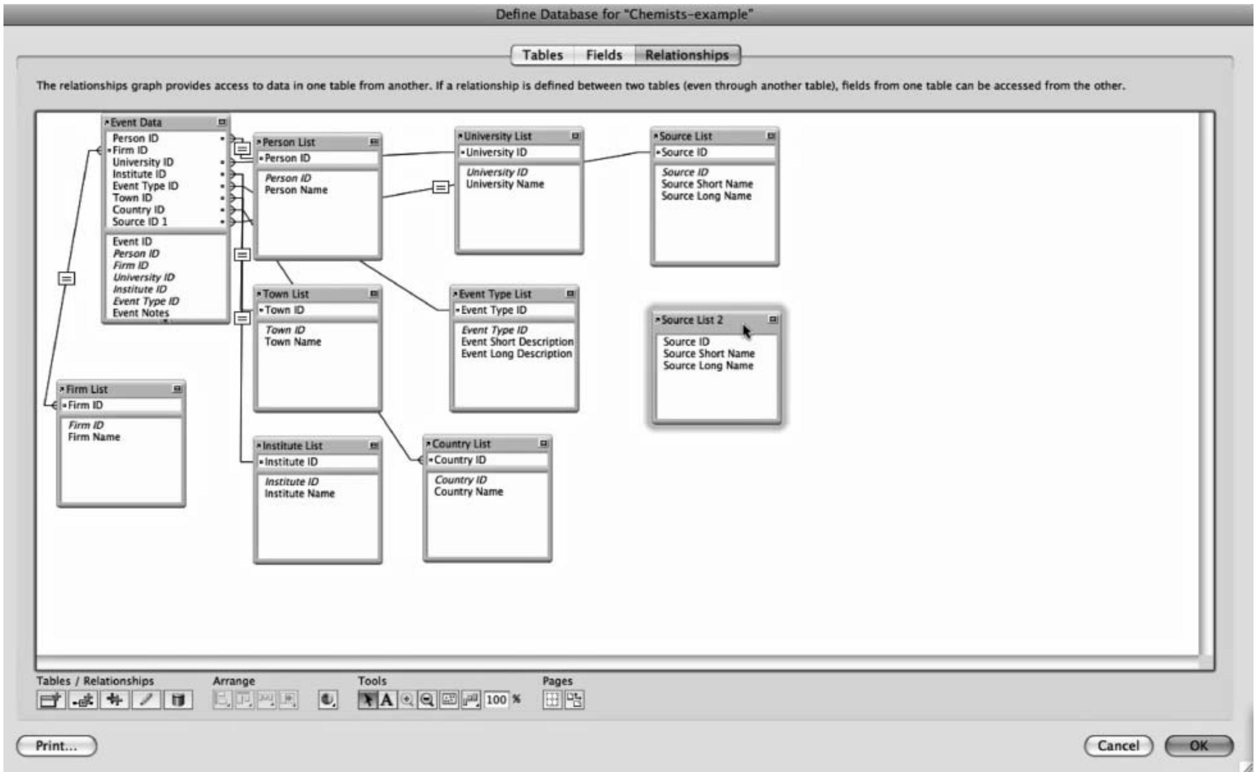


FIGURE 12. Activated table window.

Key Challenges in Setting up the Relational Database

FileMaker, unlike Microsoft Access, is very easy to use, even for people who have no programming skills. I say this from experience because I initially started to create the chemists' database in Access. I figured that distributing the database would be easier with Access because most people have the program on their PC as part of Microsoft Office Suite. After I had to involve a professional programmer to set up a simple search function in Access that comes without any programming in FileMaker, I returned to using FileMaker, which I have used for more than 10 years and could set up myself. Based on this experience, I recommend that researchers not use Access—unless they are professional programmers or want to spend the money to hire one—and use FileMaker instead. Researchers will easily recoup the $180 purchase price for the academic version of the program based on the time saved in setting up the database.

Designing the chemist database with all the relationships between different tables of information was not difficult in FileMaker. The manuals and the tutorials available on the Web (http://www.Filemaker.com/products/fmp/video_tutorials.html) provided clear instructions on how to set up relational tables that minimize the need for entering a lot of data by hand.

The only complication that I encountered concerned the source ID fields (and later the two town location fields). Because I wanted to be able to cite multiple reference sources from the source table, using the one link between the main event table and the reference table would not work. Every source ID field and the associated pieces of information (in this case, the short name of the reference and the long name of the reference; see figure 9) needed to have their own unique relationship to the references list table.

To understand how you can create many-fields-to-one-list links, I will walk you through the process of setting up a second source ID field in the main-event data table and then the process of creating a working link to the existing reference database table. (To get a better understanding of the process, you can obtain a trial version of FileMaker and download the chemist sample database from http://professor-murmann.info/files/chemists-example.fp7 and try out my instructions.)

1. Add a new field to the Event History Database by going to File -> Define -> Database and clicking on the Tables tab field.
2. Select the Event Data table.
3. Click on the tab called Fields.
4. Put into the Field Name the text "Source ID 2" and select Number under the Type menu.

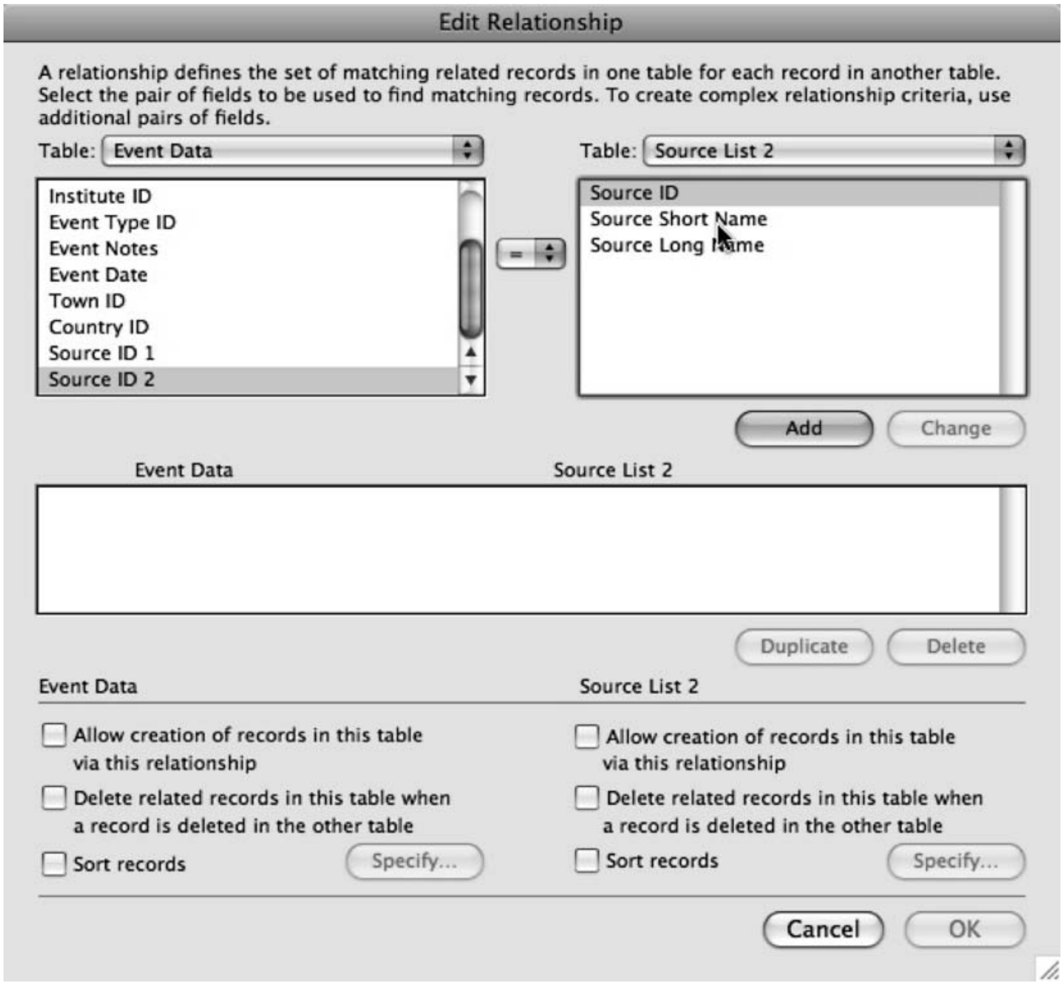


FIGURE 13. Defining relationship between variables in different tables.

5. Click on the blue Create button.

Now that you have created the source ID 2 field in the event database, you can link this to the source list database table.

6. Click on the Relationships tab.

You will see existing relationships among the tables of database (figure 10).

There is already one *Source List* window that represents the already-existing link from the event database to the source-list database. Now we will add a second one.

1. At the bottom left of this window, click on Tables.

You will now see the dialog in figure 11.
Click on the Source List item in the list. FileMaker automatically enters “Source List 2” in the *Name of Table Occurrence*. Click OK.

Now you will see a little window surrounded by a yellow shadow (see figure 12).

All you have left to do is to create a link between the event data window and the source list 2 window. You need to link the source ID 2 name at the bottom part of the event data window. It is the only name not italicized.

There are two ways to create the link. One way is simply to click on the source ID 2 field and then start dragging the mouse to the right. A line will appear. Drag it all the way over the source ID variable in the source list 2. If you have some trouble with this, there is another way that requires less eye-to-hand coordination. Click on the second icon to the right under the Tables/Relationships icon strip that contains a green plus (“+”) sign and symbol for relationship. This dialog in figure 13 will appear.

Above the left window, choose Event Data from the table list. Above the right window select Source List 2 from the table list. Now all the variables associated with each table will appear in both windows. In the left window click on

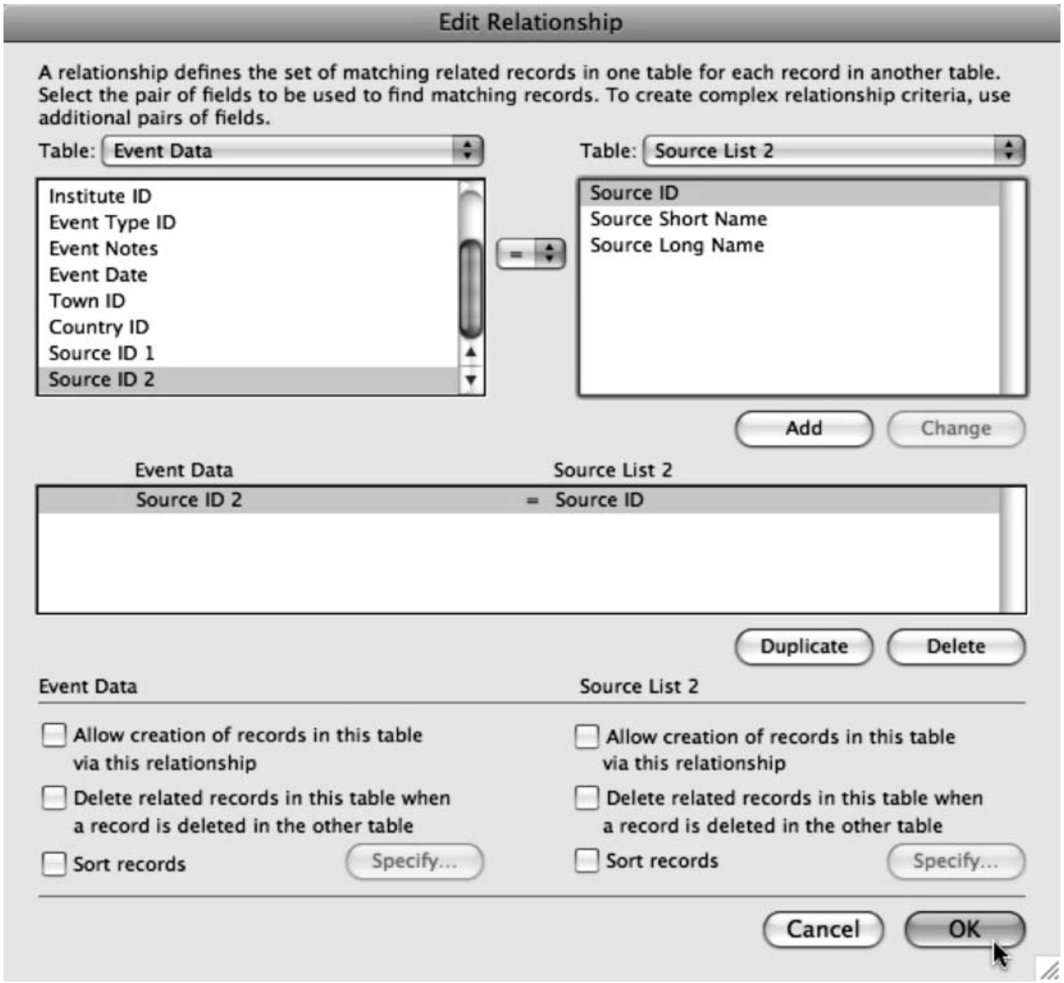


FIGURE 14. Creating relationship between variables in different tables.

Source ID 2. The background of the variable will turn blue. Now select the source ID variable in right left window. The background for the variable will also turn blue. Click on the Add button below the right window to create the link. The new link will be displayed as shown in figure 14.

Finally, click on the OK button at the bottom right corner. Now you have finished creating the relationship between the second source ID field and the reference list table. You will see it once you have clicked OK. Because you now have a second unique relationship between the event data table and the source list table (source list 2), you can pull in different information through the source ID link and the source ID 2 link. For example, to pull the short name of the source into the event data, you can add this field from the source list 2 just like you would for any relational data table field. You need only to make sure that you pick the variables from the source list 2 and not the first source list, otherwise you will find the name of the journal referred to by the number in the source ID 1 field in the short name field.

Please note that we could have used any other table in the FileMaker database to make such a many-to-one mapping of variables. For the distance analysis in the *Application* section, I needed to record in the event table both the town in which a chemist did his or her university studies and the town in which he or she took his or her first job. Just as in the case of making two references to the source list, in this case, we are making two links to the town list. We have to follow the exact same procedure as before when we made a second link to the source ID list.

Application

As previously mentioned, many sociologists, anthropologists, and historians are getting interested in computer mapping individuals' travels and encounters, ships' journeys, and the like. Among the many interesting research questions you can ask, one relates to the movement of chemists throughout their careers. In my book (Murmann 2003), I presented

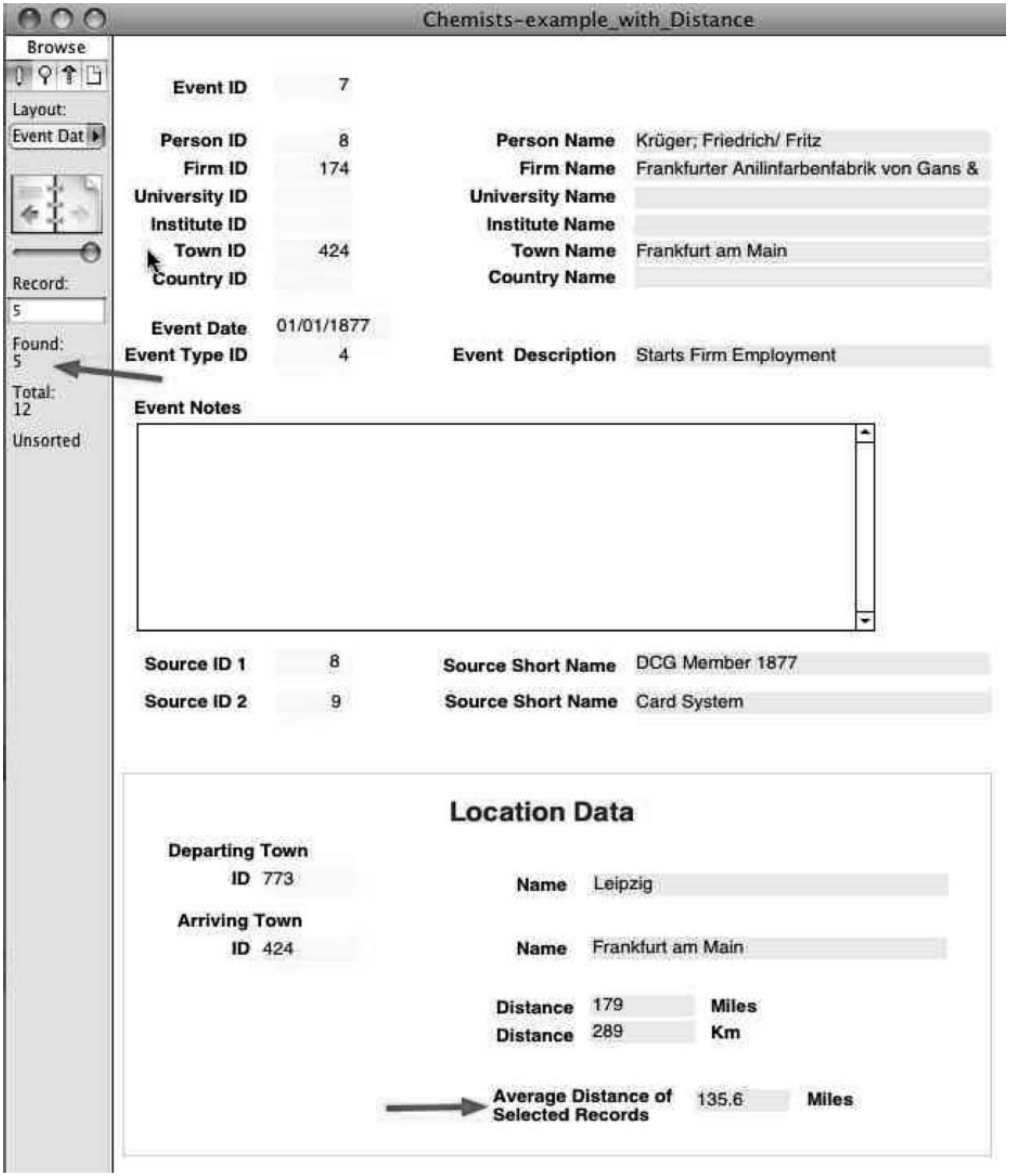


FIGURE 15. Distance of first job from university location before 1881.

anecdotal evidence that chemists throughout history moved from university labs to firms and back. With our event-based histories on all German chemists it is possible to analyze these migration patterns systematically. For example, you can ask if in the early period of the development of the industry (1857–80) chemists would travel further between their location of university study and their first job location than in a later period (1881–1914). Given my own programming skills, I could do this easily by importing the data from FileMaker into the SAS program.

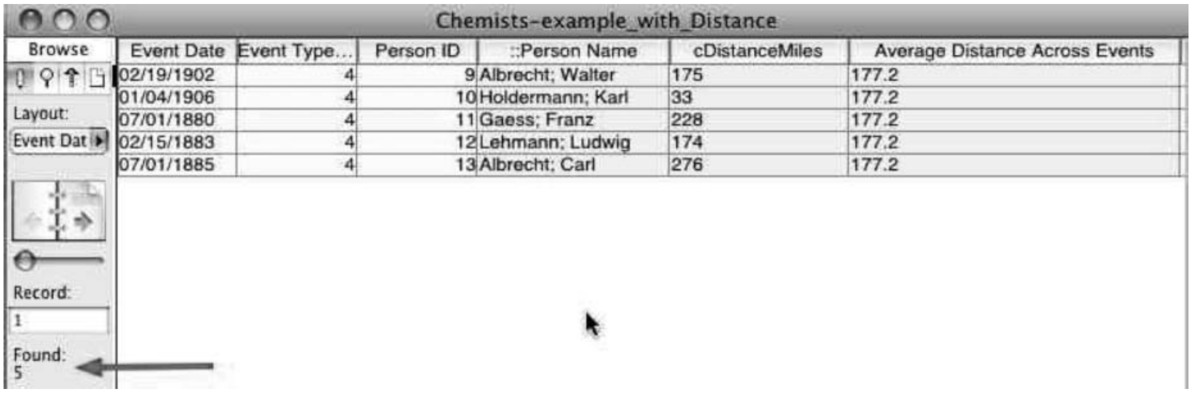


FIGURE 16. Distance of first job from university location after 1880.

But I want to illustrate here how you can set up basic distance analyses in FileMaker. For that purpose I added to the event table two geographic locations fields. The arriving town ID is simply a copy of the town ID field that already existed in the database (in FileMaker, when you enter a number in one field it will automatically display the data in the copy of that field). However, I added a departing town ID to the event data table so that I could track the distance that chemists traveled as part of the event. Both fields are linked to the town list table that stores the precise geographic location data and from which you can calculate distance by creating a new variable in the event table called *distance*. Instead of troubling myself with programming this formula in FileMaker from scratch (this would have required several hours of programming and also forced me to buy FileMaker Pro Advanced, which for academics is $80 more expensive than the regular version), I contracted with Hi-Voltage's FileMaker database developers to add this function to the chemists' database for AU$250.[5]

I added to the example database 10 event records of chemists who started their first job. Each event also tracks the city in which the chemists resided before taking their first job. In each record, the distance between the two locations is calculated. I also added a variable to the event database table called *average distance of selected records* (see figure 15). FileMaker can calculate averages across selected records very easily by creating a new variable that has the field type *summary* and then selecting *average* as the desired option for the summary variable. Below you see the entries that were found in the event table of the second chemist example database (a link to it is provided in the reference list) when I searched for all events that marked the start of the first employment earlier than January 1, 1881. Five out of the twelve records fall into the period (see figure 15). The average distance of the five events is 135.5 miles.

I then performed a second search for all entries of the event-type *starts employment in firm* after December 31, 1880. This yielded five records (see figure 16). You can see all records on one screen if you go to the view menu and selects View as Table option.

The average distance for these five records is 177.2 miles. This means the five chemists who took up their first employment in the period (1881–1914) traveled 42.3 miles farther from their point of university study than did the five chemists in the period from 1857 to 1880. (The sample in this example is too small to make any reliable inferences about whether this pattern holds for the entire population of chemists.)

Once our research team has collected data for all German chemists from 1857 to 1914, you can see how easy it is to get very quickly a definite answer to this and many other questions with the FileMaker search function. You can find out the distance patterns for any subsample that can be defined by selecting values for particular variables and searching for these values. If you search for a specific person, for example, you could obtain the average distance this person traveled across all movements in the career of this person. If you want to know whether firms differed in terms of how far away they recruited chemists, you can search by firm and compare firm patterns. What is more, you can export the FileMaker data as a spreadsheet table and import it into statistical software such as SAS or STATA to do more sophisticated analysis.

Conclusion

This technical note used the example of the chemists' career history database to illustrate how to design efficient and flexible databases for research projects. The key idea is to make an event in the life of an agent the basic unit of observation. The approach can be applied to any conceivable kind of agent (people, groups, firms, organizations, nation-states, etc.). It can even be applied to objects not commonly construed as agents. Examples of such agents include academic articles, books, television shows, newspaper articles, strikes, elections, demonstrations, and buy-or-sell offers.

Quantitative social historians have used relational database designs before (see, e.g., Tilly 1995, appendix I), but these previous efforts have required significant technical abilities. With FileMaker, every researcher can readily create such databases for himself or herself.

NOTES

1. SAS can be accessed at http://www.sas.com/software/sas91/. SPSS can be accessed at http:///www.spss.com/software/statistics/statistics-base/. STATA can be accessed at http://www.stata.com/stata11/.

2. I have posted the chemists database files (Murmann 2010a, 2010b) on the Internet as an example so that you can see how such a relational database looks and experiment with it before setting up your own database. There are two versions of the database. The first version is the version before adding the additional variables. This file is fully unlocked, allowing you to make any additions you want, including those outlined in this article. A second version has all the additions you would make if you followed the tutorial in this article. It also has the functionality to calculate distance between two geographic locations. Because I had the programmers of Hi-Voltage add this functionality to the chemists' database for AU$250, the file is semilocked, making it impossible to add new variables but allowing you to add and delete data.

3. In the technical literature of event history analysis, such as Alter and Gutmann (1999), a distinction is made between a strict *event* and an attribute of an individual at a specific point in time, which is called a *declaration*. This distinction is very important for formal event history analyses but is not important for the design of a database. In our chemists' database, the event list table contains both strict events (e.g., birth) and declarations (e.g., appearance in directory).

4. Data files for geographic coordinates of towns (GNS) are available from the National Geospatial-Intelligence Agency Web site, http://earth-info.nga.mil/gns/html/entry_files.html (accessed June 30, 2010). You can download the data for individual countries or multiple countries. In the case of Germany, whose data I downloaded and imported into FileMaker, the database had 166,929 locations. We are not done with entering all biographical data, but so far the Geonames database covered all the German towns that appeared in the biographies of chemists. This, of course, is not surprising, given the German version has almost 170,000 locations in it.

5. The alternative would have been to buy FileMaker Pro Advanced (academic price was $250) and buy the custom function LatitudeZERO from Hi-Voltage for AU$175 to calculate distance between geographic location based on longitude and latitude coordinates or zip codes. For someone who does not already own the regular version of FileMaker, this would amount to the same price as my having the functionality added by Hi-Voltage to my chemists' database. However, I saved myself time by learning how to work with custom functions in FileMaker Pro Advanced and installing the function myself. I know enough programming to do this, but I wanted to present readers of this article with a solution to getting distance calculation as simply as possible. Hi-Voltage will add a field and the required formula to calculate distance for AU$250.

REFERENCES

Alter, G., and M. P. Gutmann. 1999. Casting spells: Database concepts for event history analysis. *Historical Methods* 32:165–76.

Analytic Technologies. *UCINET*. Computer software. Analytic Technologies. http://www.analytictech.com/ucinet/.

FileMaker, Inc. *FileMaker Pro.* Computer software. FileMaker, Inc. http://www.filemaker.com/.

FileMaker, Inc. *FileMaker Trial.* Trial version of computer software. FileMaker, Inc. http://www.filemakertrial.com/.

FileMaker, Inc. *FileMaker Demos.* Online tutorials for computer software. FileMaker, Inc. http://www.filemaker.com/products/fmp/video_tutorials.html.

Hi-Voltage. *LatitudeZERO*. Computer software. Hi-Voltage. http://www.hi-voltage.com.au/latitudezero.php.

Murmann, J. P. 2003. *Knowledge and competitive advantage: The coevolution of firms, technology, and national institutions*. New York: Cambridge University Press.

——— 2010a. *Chemists' example with distance*. Personal Web site. http://professor-murmann.info/files/chemists-example-with-distance.fp7.

——— 2010b. *Chemists' example without distance*. Personal Web site. http://professor-murmann.info/files/chemists-example.fp7

National Geospatial-Intelligence Agency. *Geonames.* Computer software. National Geospatial-Intelligence Agency. http://earth-info.nga.mil/gns/html/entry_files.html (accessed June 30, 2010).

SAS. *SAS.* Computer software. SAS. http://www.sas.com/.

Scott, J. 2000. *Social network analysis: A handbook*. Thousands Oaks, CA: Sage.

SPSS. *SPSS*. Computer software. SPSS. http://www.spss.com/.

STATA. *STATA*. Computer software. STATA. http://www.stata.com/.

Tilly, C. 1995. *Popular contention in Great Britain, 1758–1834*. Cambridge, MA: Harvard University Press.

Wasserman, S., and K. Faust. 1994. *Social network analysis: Methods and applications*. New York: Cambridge University Press.