

Tombstone Identification through Database Merging

by H. Daniel Wagner

This article originally appeared in the Fall/Winter issue of Roots-Key, the journal of the Jewish Genealogical Society of Los Angeles. It is reprinted with permission—Ed.

The development of sophisticated tools designed to integrate—or merge—family data from different sources and databases from different repositories is currently viewed as a major objective in the virtual reconstruction of past Jewish families and communities. The rewards and complexity involved in database integration are examined here by means of a specific example viewed as a pilot project: the merging of the recently created burial list for the Jewish cemetery of Zdunska Wola (the cemetery was in use between 1828 and World War II) with the Zdunska Wola metrical death listings found in the Lodz archives (for the years 1808–1906) and in the local town hall (for 1907–42).

Introduction

Most Jewish communities of Europe were devastated during World War II. Towns in Poland and elsewhere were emptied of their Jewish residents, and entire families disappeared, often leaving no descendants to attest to their very existence. After the war, exhaustive efforts were invested to collect the names of victims and survivors. Attempts of a new kind are currently underway to memorialize the victims by virtually reconstructing the family trees of individuals on a community-by-community basis and to recreate the Jewish populations that existed just before the Holocaust. This increasingly sophisticated approach to Jewish genealogy—and to genealogy as a whole—is evolving into a field of far-reaching information and knowledge: in the last 15–20 years, the character of genealogical research has increasingly become an intricate, multifaceted academic pursuit, as detailed elsewhere.* A large number of local and regional databases have survived in Europe. Such databases, if suitably combined, inevitably lead to a much improved biographical depiction of individuals, to a more correct (and probably much larger) account of the number of victims, but also to the multi-generational reconstruction of many lost branches of the Eastern-European Jewry. The issue of database integration forms the core of the present article.

Concept of Database Merging

The simplest case of database merging may be illustrated as follows: two databases, A and B, containing genealogical material that includes both overlapping (usually names of individuals) and non-overlapping information are compared and then combined. The objective is to attempt to gather as much information as possible about a given individual (for example, one's grandparent). An example is the merging of a town's metrical birth and death records. In addition to details regarding the newborn, birth records usually include

names, ages and occupations of the parents; death records, on the other hand, usually include age at death and often identify surviving family members. The merging of these two databases obviously increases the amount of genealogical information for a given individual and his/her family. The main problem, of course, is to ensure that the merging procedure links identical individuals, namely that the birth and death documents of Jacob Goldberg, for example, are those of the same Jacob Goldberg. In other words, merging criteria of "identicalness" must be defined as accurately as possible. (Figure 1)



Figure 1—Integration of information contained in databases A and B. The merged database includes overlapping and non-overlapping information, thus increasing the useful amount of data for an individual.

Obviously, the complexity (as well as usefulness) of the integration process grows when more than two databases are available for merging, for example metrical birth, death and cemetery data. As a concrete example, we will consider the Polish town of Zdunska Wola, for which various types of databases exist for its pre-World War II Jewish population, such as:

- 35,000 metrical data (from three original databases: births, marriages, deaths) for the years 1808–1942
- 3,500 burial records recently extracted from tombstone transcriptions in the local Jewish cemetery
- 2,300 necrology records from the Jewish Memory Book (yizkor book)
- 1,100 surnames from the 1929 Polish Business Directory data
- Records from additional Zdunska Wola databases, such as the house-by-house (KLS) registrations, the 1933 applications for identity cards (which include sometimes unique pictures), names of Auschwitz survivors, surnames on the Zdunska Wola memorial monument erected in Tel Aviv by survivors, etc.

Here the focus is on the merging of the metrical death data and the cemetery listings, which can be viewed as a pilot project. Why these two specific sets?

- A stricter interpretation of Jewish tradition does not allow including surnames on tombstones, as evidenced by

about two-thirds of all tombstones in the Jewish cemetery of Zdunska Wola. The integration of the cemetery listing with the death metrical dataset makes it possible to assign a surname to most of those tombstones, allowing their exact identification, which is of utmost importance to family members and descendants.

• Manual merging, although very time consuming, is relatively easy to perform and will allow comparison with future automated merging, leading to software refinement and optimization. A few examples of manual merging follow, with incremental levels of difficulty.

Illustrative Examples of (Manual) Database Merging

A. Consider the headstone (Figure 2) of a young girl named Sara Rywka, daughter of Zeew Wolf, who died on the *rosh chodesh* (first day) Cheshvan 5672. What was her surname? By merging this information set with the death metrical database—which always includes surnames—it usually is possible to assign a surname. This is accomplished by first translating the Hebrew death date into a civil date (in this case, October 22, 1911) and then by searching for a Sara Rywka towards the end of the 1911 list of deaths. In this case, we are fortunate, because she is the only Sara Rywka in the 1911 list; her surname is found to be Jakubowicz, (record #101 out of 124 records in 1911).

B. Next, consider the small fragment of tombstone shown in Figure 3. The first name (probably Sara) is in acrostic form, with a second name starting with ‘P’ or ‘F’. Again, there is no surname. The Hebrew date is 15 Kislev 5671 (16 December 1910). Searching towards the end of the 1910 metrical death listing, one finds Sura Perla Berkowicz (record #83 out of 93 records in 1910), a high-probability match. Another candidate record exists: Sura Pacanowicz (record #19), with a much lower probability, however, because (i) the month of Kislev is located towards the end—rather than the beginning—of the listing for 1910, and (ii) surnames are almost never part of an acrostic.

C. The next example is the headstone of Rywka Necha, daughter of Szymon (Figure 4).

Figure 2 – Headstone of a young girl named Sara Rywka, daughter of Zeew Wolf, who died on the 1st of Rosh Chodesh Cheshvan 5672. No surname is available.



The year of death is probably

Figure 3 – Fragment only, the date of death reads 15 Kislev 5671; the first name is an acrostic form of Sara with a second first name starting with “P” or “F”. No surname is available.



5693, thus 1932–1933, and there is no surname. Searching the approximately 35,000 metrical data (for births, marriages and deaths), one finds 15 Rywka/Ryfka Necha/Nacha. The closest death record is found to be Rywka Necha Halperin (record #23 in 1936), with, however, a late 1936 death registration date (which underlines the need to solve a difficult issue, late registrations, in any future merging software). Integration with additional databases in Zdunska Wola yields the following wealth of information:

- Marriage metrical list: Abram Sucher Halperin with Rywka Necha Szmulewicz, (record #40 in 1913)

- Identity card listing of 1933 for Rywka Necha Halperin, which includes in this specific case the only existing remaining photograph of Rywka Necha

- The Book of Permanent Residents (KLS) for the Halperin/Szmulewicz family: House #10, pp. 261–289, and



Figure 4 – Tombstone of Rywka Necha bat Szymon, deceased on (hard to read) 12 Iyar 5693 at age 46. No surname available.

more.

D. As a last example let us consider the broken stone shown in Figure 5.

Despite the paucity of information available from the headstone text, it is possible to discover, with a fair degree of probability, the identity of the deceased, with the condition, however, that the computerized metrical databases are improved. Indeed, by doing a search over all 35,000 metri-

cal data, one finds fifty-one Rasza/Raszka/Rasze (including nineteen double first names), with thirteen of these being death records. Elul is located approximately in July-August, thus in the mid-range of the civil year, which reduces the number of likely records down to six. Here, in fact, we are trapped: the only way to determine the correct surname is now to extract the father's name from the original metrical record (the father's name does not figure in the original index, only in the record itself). Thus, the computerized metrical index database must be augmented with (at least) the father's name by reading all original records. This is definitely hard work, but it is the only way to ensure a very high success rate for the merging procedure.

Guidelines for a Strategy in Future Automation of Database Merging

To optimize the future automated merging process of genealogical databases, a step-by-step procedure will have to be developed:

Step 1: Preparation of a "well-behaved set" of metrical death data: to maximize the probability of matching a tombstone having first name only and date with a specific metrical death record, the parents names of the deceased will have to be systematically extracted from the original death records and included in the data set. The identification of the father's name—always present on the tombstone—will ensure a valid match.

Step 2: Manual merging of pilot (calibration) sets of metrical death and cemetery data for future comparison with automated merging.

Step 3: Development of an automated merging procedure by creation of dedicated computer software.

Step 4: Application/calibration/refinement of the merging software by means of the pilot data sets.

Step 5: Implementation of software using the full scale shtetl database.

The development of a computer program to automate the merging of databases will necessarily have to deal with a number of difficult problems, including the development of a sophisticated searching soundex algorithm that will be needed to merge similar but incomplete or incoherent information: name spelling differences, partial names (double names vs. single names, for example), different names (names in Yiddish vs. names in Hebrew, such as Hersh for Tsvi), radical changes in surnames, different dates for a single event in different databases, and so on. Since the conventional Daitch-Mokotoff soundex algorithm may not be suitable (for name in Hebrew in one database vs. name in Yiddish in another, for example) or even irrelevant in some cases (obviously different names such as Szleifer and Zilber, nevertheless, possess the same Daitch-Mokotoff soundex code), the soundex will need to be reinforced by the creation of an extensive table of equivalence or dictionary, in which "equal" names (or surname change when known) will be listed.

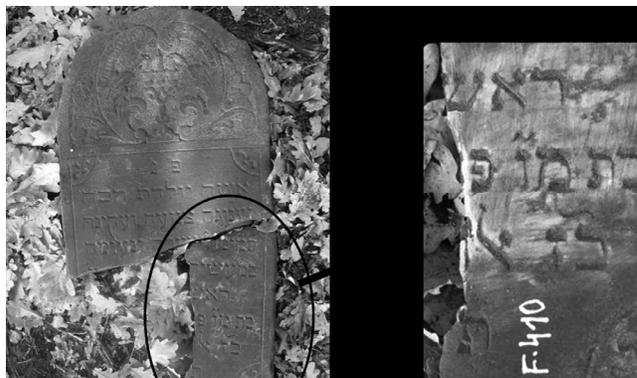


Figure 5 – The magnified insert reads *Rash[a]? Daughter of P... or F..., died 27 of Elul, without the year.*

Other difficulties arise: quite often, several individuals bearing the same first name may have died that year; the deceased may have been registered in a later year in the metrical books, and so on. But in the case of a positive identification in a merged database, the precious tombstone of the deceased family member becomes available to the descendants.

Note finally that the degree of success of the merged data set will have to be set by comparison with the manually merged listing. If a match of say 90% (or more) is found between both data sets, we will consider the integration to be successful even though some manual corrections will still be necessary.

In principle, merging could be applied to a large number of family trees existing in a given community (the "community forest") in an effort to create a single communal tree.

Concluding Comments

Genealogy continues to develop as an increasingly multidisciplinary activity. We Jews, in particular, have a specific reason to pursue the burgeoning field of genealogy: The tragedy of the Holocaust has created an enormous vacuum in practically all European families, a vacuum that begs to be filled. Our parents emerged from the Holocaust without most of their family, without documents and photo albums; without their past, which is also our past. Therefore, the development of sophisticated merging tools is becoming a priority in Jewish genealogy, where a major objective of a large number of researchers is the virtual reconstruction of one's ancestral town or shtetl.

Notes

*H. Daniel Wagner, "Genealogy As an Academic Discipline," *AVOTAYNU* Vol. 22, No. 1, (Spring 2006): 3–11.

H. Daniel Wagner is Professor of Materials Science at the Weizmann Institute of Science in Rehovot, Israel. He is the author of 150 scientific papers and 15 genealogical papers. Wagner is a member of the Israel Genealogical Society and was co-chair of the 24th International Jewish Genealogy Conference held in July 2004 in Jerusalem.