

Strategies for the Integration of Genealogical Datasets

Kamila Klauzinska

(modified paper, first presented at the
29th IAJGS Conference of Jewish Genealogy
(Philadelphia, August 2-7, 2009)

Birth of the Project

Since the early 2001's, Prof. H. Daniel Wagner of the Weizmann Institute and I have been engaged in a project to document the graves in the Jewish cemetery of Zdunska Wola, in central Poland.¹

In September 2004, with the help of some high-school students, we were documenting section „A”, one of the 11 sections of the cemetery [slide 2].² As we came upon the headstone (*matzeva*) of Bajla Mindla, daughter of Henich GOLDHAMER, I sensed that I had already seen this tombstone somewhere. As I told Professor Wagner, it seemed to be the *matzeva* in an old photograph that he had placed on the website of the society for former residents of Zdunska Wola. I could not be sure, however, because the plaque of the *matzeva* we had just found bore no inscription. All that survived were the symbol at its head, the name of the deceased and the family name of her husband.

Later that day, we compared the new photograph with the old one. There was no doubt that the stone newly documented by us, *matzeva* A320, belonged to the KOBER family, whose history Prof. Wagner had placed on the Zdunska Wola website, together with an old photo of the *matzeva* from before the Second World War [slide 3].

We decided then and there to encourage the residents of Zdunska Wola to send us any old photographs of headstones from the cemetery in their possession. Our purpose was to compare them with the other tombstones without inscriptions still available at the cemetery, and indeed we were eventually successful in establishing family data for a number of tombstones.

¹ Klauzinska K., “Colored Tombstones in the Jewish Cemetery in Zdunska Wola”, *Ars Judaica*, vol. 5, 2009, pp. 121-128.

² Zdunska Wola is an industrial town located some fifty kilometers to the southwest of Lodz. It was built with great contributions of the Polish, Jewish and German communities. At the beginning of the 20th century, the Jewish community constituted nearly 40% of all residents, and in 1917 as much as 47.8%. The first mention of Jews living on the territory of the present Zdunska Wola dates back to 1788. The Township Rights Act, legislated by Tzar Alexander I in 1825, devotes its paragraph 5 to the Jewish community living in towns like Zdunska Wola. Jews were the owners of many homes in a town. Numerous Jewish craftsmen, Jewish societies and factories functioned there. They played an important role in the social and economic life of Zdunska Wola. More information about Jews of Zdunska Wola is available on the website of the Museum of The History of Polish Jews – Virtual Shtetl Project: Klauzinska K., „History of Jews of Zdunska Wola”, http://www.sztetl.org.pl/?a=showCity&action=view&cat_id=5&city_id=325&lang=en_GB (accessed 26 Oct. 2009)

Next, we decided to create an index for all the tombstones we had documented. All told, there were 3,505 *matzevot*, not all of which bore complete data. We often found that we had to compare the *matzeva* data with information contained in the town's birth and death records (so called „metrical records“). Given that this process was very time consuming, we were inspired to create a computer program that would search, compare and integrate data of this kind. Thus our project was born.

Background

The catastrophe of the Holocaust irreversibly changed the ethnic fabric of Poland. In their time, Jews had constituted the largest minority in the country, with a illustrious history going back over one thousand years. Over three million of them were murdered at the hands of the Nazis, and in 1968 many of those remaining departed, adding to the immeasurable and irrevocable damage done to the community by the Shoah. Today, virtually all that remains of former Jewish population of the country are material artifacts and the many documents stored in the national and municipal archives, registries, museums and libraries of Poland.

Since the end of the Second World War, there have been ongoing attempts to rescue and catalog these documents, in part to help families trace their relatives. The tragedy of the Holocaust made it impossible for most families to retain their memorabilia in the form of photographs of their parents and grandparents or family documents. The main purpose of contemporary documenting is to rescue memories of the residents of former *shtetls* (towns where Jews lived in significant numbers), largely through work now being done by the Museum of the History of Polish Jews in Warsaw and displayed on their Internet portal „Virtual Shtetl“.

Since the 1980s, Jewish genealogy has developed rapidly. Whereas in Poland, the study has been treated marginally as a domain supporting history, anthropology and sociology, in the world at large there has been a genealogical boom, as carried out both by family historians and social scientists.³ Computer technology and the Internet have enabled programs to create genealogical trees and specialized databases. The most developed and popular site among researchers of Jewish genealogy is „JewishGen“.⁴

Genealogical studies have become more complex and multi-faceted.⁵ The more databases have been indexed and the more Jewish cemeteries have been cataloged, the more the center of gravity of Jewish genealogy has moved towards a systematic merging of these databases, a process that, among other things, helps the recreation not only of individual family trees but also kinship groups and indeed communities. The dataset merging project using the Jewish cemetery of Zdunska Wola as a pilot falls into the latter category.⁶ The project is being conducted with the sponsorship and support of the International Institute for Jewish Genealogy in Jerusalem, Israel.

³ Klauzinska K., „A Modern Approach to the Genealogy of Polish Jews: Zdunska Wola as a Test Case“, *Scripta Judaica Cracoviensia*, vol. 5, 2007, p. 39-51

⁴ <http://www.jewishgen.org/>

⁵ Wagner D. H., „Genealogy as an Academic Discipline“, *Avotaynu*, vol. 22, No 1, Spring 2006, pp. 3-11; Lamdan, N. „Jewish Genealogy: Moving Towards Recognition as a Sub-branch of Jewish Studies“, *Avotaynu*, vol. 25, No 2, Summer 2009, pp. 3-8

⁶ Zdunska Wola's Jewish cemetery dates back to 1826. The oldest tombstone found there dates back to 1829, and the most recent ones come from 1941. At some date, probably just after the World War II, somebody put up an obelisk memorializing relatives who perished in 1943

Analytical Concept ⁷

While the implementation of our merging project is technically complex, the concept behind it is relatively simple. Let us say we seek to merge two datasets, “A” and “B”, where “A” are birth certificates, and “B” death certificates [slide 4]. If, for example, we have the birth certificate of Yisrael MOSZKOWICZ, we will try to find and match the death certificate of Yisrael MOSZKOWICZ. In a particular town, however, there may be two or three death certificates for individuals named Yisrael MOSZKOWICZ.⁸ To be sure of identifying the right one, the actual document has to be located and examined.

Various data may be at our disposal – so, as above:

A. The birth record contains the following information: father’s name, occupation, mother’s name [sometimes maiden name], date and place of birth.

B. The death record contains similar information: father’s name [sometimes his occupation], mother’s name [sometimes her maiden name], age of the deceased [sometimes the exact date and place of death]; and surviving relatives of the deceased are occasionally listed - wife, husband, children.

The more information available about a given person in the discrete datasets, the more confident we can be of proper data merging [slide 5]. The actual merging of datasets „grows“ the amount of information known about a particular person.

Let’s demonstrate what happens when several datasets are merged, using for our pilot study of Zdunska Wola, where several significant datasets have survived [slide 6], including:

- 35,000 metrical records from the years 1808-1942
- 28 Books of Permanent Residents (Ksiegi Ludności Stałej), containing over 3,500 records of Jewish families [now extracted and held in the author’s private archives]
- Applications for identification papers for the years 1930-1934, among which are nearly 600 Jewish applicants, with their photos [private archives of the author]
- 3,505 collections of data on deceased buried at the Jewish cemetery in Zdunska Wola, including photos of the *matzevot*
- A list of 91 Jews, who returned to Zdunska Wola in 1946 [private archives of the author]
- A list of 445 individuals who paid the tithe tax in 1907 [private archives of the author]
- 1,083 entries in the Polish Business Directory from 1929
- A list of 2,300 necrology records in the „Yizkor Book“ (Memorial Book) of the Jewish community of Zdunska Wola
- Over 300 records of Jewish families in the Residents Register [Rejestr Mieszkańców] [private archives of the author]
- Local newspapers – for instance *Kaliszanin* and *Gazeta Kaliska*, that contain pertinent information concerning the town’s history and its residents, including the Jewish community [private archives of the author].

⁷ Wagner D. H., „Tombstone identification through database merging”, *Avotaynu*, vol. XXIV, Spring 2008, p. 8-10

⁸ The frequency of information relating to individuals with the same name and last name increases with the size of the town and correspondingly its metrical birth and death records.

For our pilot project, we chose to utilize two datasets: metrical death records and the data collected from the Jewish cemetery.

Why These Two Datasets?

We chose these two sets for the following reasons:

1. The need to expand the data on the *matzevot*: In accordance with Jewish tradition, there are no family names (surnames) on the *matzevot*, a practice which was widely observed in Zdunska Wola. Only 629 (18%) of the tombstones bear family names, while 2,876 have no family names (82% of all tombstones). Moreover, not all tombstones have survived in their original state. Many of them are broken and no identification of them was possible. Merging the two datasets allows families to locate and identify more tombstones which belong to their family members. In fact, by manually merging the data, we identified another 1,541 family names, or an additional 54%. Adding these to the original 629, we obtained 2,170 tombstones with identified family names, 62% of the total (in comparison with 18% at the outset).
2. Manual merging of the two datasets proved the process valuable, though it consumed a good deal of time. The process had the added advantage in that allowed us to improve and optimize the performance of the program.

How the Program Was Created

To prepare the software program we divided our work into two separate phases **[slide 7]**:

1. Forming the death records database: The database was created by manually extracting and indexing all data from the death records of the years selected to test our project, and inserting them into an Excel file. The table includes 2,137 metrical records arranged specifically for the program in Polish and English (many of the original documents were written in Russian).
2. Building the computer program to execute automatic data merging.⁹ In this phase, we integrated the data from the death records and the data recorded from the *matzevot*.

The collection and documentation of the data from the cemetery has taken six years to date, done with the invaluable assistance of a group of local volunteers **[slide 8]**.¹⁰

⁹ Here I describe the design and use of the program, rather than the information technology upon which it was built. Technical questions may be directed to Jakub Zajdel, IT specialist and member of the team that created this program: jakubzz@wp.pl

¹⁰ Klauzinska K., „Colored Tombstones in the Jewish Cemetery in Zdunska Wola”, *Ars Judaica*, vol. 5, 2009, pp. 121-128. It is important to mention that the extensive research conducted at this Jewish cemetery is one of a few large projects carried out with the participation of volunteers in Poland. It produced the first modern documentation ever created for the Jewish cemetery in Zdunska Wola. Burial registries have not survived.

As mentioned above, each metric death record contains the following data: surname, given name(s), record number, year of record, type of record [in this case D for death], date of death, age of the deceased person at death, father's name, mother's name, name and family name of spouse (if alive), and sometimes civil status (widower, widow) **[slide 9]**.

We then compare this set of data with the data set from the *matzevot* from the cemetery, which may contain some or all of the following information: given name(s), surname, location of grave [section, number], Hebrew date of death, father's name, name and family name of husband, plus an epitaph, sometimes with further personal information.

To carry out this operation, the program „merge.exe“ was created, the prototype of which is shown in the accompanying slide **[slide 10]**.

1. At the upper left, we have the area of search. The user can use the Soundex option, if desired.
2. At the bottom left, we have the area where results from metrical birth and metrical records are displayed.
3. At the upper right, we have the photos of *matzevot* and additional documents [the photo can be enlarged by clicking on the small image].
4. At the bottom right, we find the search results from the database of the cemetery. By clicking on „Search“, the program shows all results for a particular search. By clicking on „Export Selected Records“, the user transfers the data regarding a particular person to a printer friendly file **[slide 11]**.

Let us look at three examples:

Example I [slide 12]:

When we look at this fragment of a tombstone found in the cemetery, we can read the name: „(...) shon [?]“, son of Moshe, died „2“ or „22“ Tamuz 5665 (according to the Hebrew calendar). We may now enquire as to his full name and family name.

First, by comparing this data with the death records which have the family names of the deceased, we can establish the identity of the deceased. The first step is to convert the Hebrew date into standard dating, according to which a son of Moshe died on July 25, 1905, leading us to look for the death record for Gershon in 1905. Luckily, in 1905 we find only one death record for the name Gershon **[slide 13]**.

We then read the rest of the record. We see that Gershon, son of Moshe, died June 21, 1905, age 20. Obviously, the date does not match. We recall, however, that in 1905, Zdunska Wola was located in the Russian partition of Poland. Russian clerks wrote documents in the Cyrillic alphabet and used dates according to the Julian calendar, which in 1905 (i.e. the 20th century) differed by 13 days from the Gregorian calendar.

We still cannot be certain, however, that the *matzeva* found by us belongs to this particular Gershon **[slide 14]**. Other data from the death records are helpful in this regard **[slide 15]**. Looking at the death records preceding Gershon's, we notice that two other people died on the same day and their father's name was Moshe. In fact, in 1905 Zdunska Wola a total of eight people died that day and their deaths were recorded in eight consecutive records. Where could something more be found to determine this unique event?

From 1898 on, a daily newspaper *Gazeta Kaliska* was published in Zdunska Wola. In the volume for 1905, on July 7, a long article was published about a fire which took place in Zdunska Wola on that day [slide 16]. The article listed the names of all who died in it, with the result that the small fragment of the *matzeva* we found at the cemetery could be seen as belonging to Gershon WINTER, born in Widawa, son of Moshe.

Gershon was, it said, only 12 years old, though in the death certificate Gershon's age is 20. We cannot be sure which is mistaken, but we do know that Gershon lived at Laska Street, and at that address there was a post office, where a fire erupted. Together with Gershon, a total of 16 people died, eight of them Catholics and eight Jews. Amongst them were Gershon's siblings Ryfka Fajga and Itta, and the pregnant Matla RZESZOWSKA from Klodawa, and most probably her husband Lajzer Dawid RZESZOWSKI. Gershon's father, Moshe, was a widower. He was a carpenter.

Gershon's burial and that of the other victims took place, according to the Jewish tradition, within the next 24 hours, on July 6 at 2 p.m. (the fire broke out at 2 a.m. on the night of July 5-6, 1905). Jews and non-Jews took part in the burial. At 7 p.m. there was another burial, this time of the Catholics who died, with many Jews attending. After the funeral, the Jews, together with the Catholics, took up a collection on behalf of Tadeusz BIEGANSKI, whose wife and four children died in the fire.

This beautiful, if sad, lost moment in the history of Zdunska Wola was revealed and brought back to life, as it were, by that small fragment from the cemetery, and with the aid of merged datasets [slide 17].

Example II:

Another example is that of the tombstone of Rywka Nacha, daughter of Shimon, died 12 Iyar 5693 [slide 18].

By converting the date to the civil one, we can see that the deceased must have died in 1932 or 1933. Sifting through 7,000 death records for the name Rywka and Rywka Nacha, we find nearly 130 death records, among them only two death records for Rywka Nacha. The first one is from 1889 and the second from 1936. Obviously, 1936 is closer to the mark (Rywka Nacha HALPERIN, daughter of Shimon SZMULEWICZ and Ester LEWKOWICZ).

We must take into account the fact that death records were sometimes created after a delay, which may make a case more difficult to solve. In future, in such cases, we will need improved database merging. In this particular case, however, we had read the date on the tombstone incorrectly, due to the very poor condition of several letters. The last letter in the date we read - incorrectly - as a *gimel*, and not a *vov*, yielding mistakenly the year of 1933, instead of 1936.

However, by merging of the data available for Rywka Nacha with several other sources of information, we obtained surprisingly good results, thus proving that our merging of the first two databases is correct. We found a marriage record of Rywka Nacha SZMULEWICZ in a document for Aron Sucher HALPERIN (record # 40/1913). Furthermore, there is an application for an ID card: Rywka Nacha HALPERIN, daughter of Szymon, containing her only photograph (ID from 1933, p. 73). From the ID card, Rywka Nacha lived at Juliusza Street. Looking into the Business Directory from 1929, we found the entry of a grocery salesman, HALPERN

A., on Juliusza Street. Additional information is in the Books of Permanent Residents (vol. 10/261-289, p. 187, house 269) where there is a record of HARPEN–SZMULEWICZ.

Case conclusively proven!

Example III:

The third example is not very complex, but it is of interest [slide 19]. Let us return to something mentioned at the beginning of the paper: the old photographs we have from the cemetery.

In the „*Yizkor Book*“ for Zdunska Wola there is a photo, showing in the distance, on the left, a tall tombstone with a simple text Mirel BIRMA. Two questions arise:

1. Can this tombstone now be found somewhere in the cemetery?
2. Can we determine with precision relevant data: date of death, full family name, etc.?

First a manual comparison needed to be carried out, and then the computer had to be employed.

i. Manual Data Comparison:

Records from 1808 – 1942 show the name Mirel 25 times. The name Mirla occurs 201 times. However, in searching for BIRMA(N) or BYRMA(N), we find a singular death record of Mirla BYRMAN died in 1911 (death record # 84).

ii. Computerized Data Comparison:

Searching for Mirel in our program, we find seven records for the name Mirel/Mirla and 12 records from the cemetery (for the time being, the database of metrical records in the program encompasses only a limited number of years; were it to comprise 35,000 records, the number of occurrences of the name Mirel would certainly be higher). The comparison is fruitful. The *matzeva* of Mirel still exists in the cemetery, and the tombstone is marked A463. Merging the data from the cemetery and the Registry, we find that Mirel was a 82 year old widow after Moshek Gersh (Hersh/Hirsh) GOLDBART.

We returned to the cemetery to test the result of our Computerized Data Comparison. The *matzeva* identified by the computer was lying upside down. When we turned it over, we found the same inscription there that was visible in the old photograph: Mirel BIRMA(N) [slide 20]!

Pointers for Future Work on Developing Programs for Merging Datasets:

In developing further a computer program for the automated merging of data, many difficulties still have to be addressed. The essential problems to be solved will include:

- **The problem of reading and merging incomplete or incomparable data:**
 - As seen in the first example shown, a *matzeva* may provide only a fragment of a name – for instance: „(...) shon[?]”
 - Integrating names in specific datasets written in different languages: in Hebrew and Yiddish: Zvi in Hebrew (rendered „Cwi“ in Polish) = Hirsh in Yiddish; or Zippora/Cypora = Fajga in Yiddish, etc.
- **The problem of metrical records (especially birth records) prepared a few years after the event.**
- **Inconsistencies in the ways the date is written. Sometimes the date is written in Hebrew according to the Jewish religious calendar, as on *matzevot*. Metrical records have Julian or Gregorian dates, or Julian and Gregorian side by side.**
- **People of the same name (homonymous individuals), who may have died in the same year.**
- **The program should be able to rate and rank the results – indicating the person according to the highest probability (by reading and merging other data from different databases and, for instance, matching the father’s name, which is an important element in identification, appearing on the vast majority of *matzevot*).**
- **Key words – for example, by searching for “fire”, we should easily pull up the four *matzevot* at the cemetery which bear information on the fire of July 5-6, 1905.**

Clearly, manual checks will still be necessary, but our basic computer program can greatly narrow the range of probable data of interest. Thanks to the program, a typical family can now find its relatives’ tombstones faster and with greater probability than ever before. Moreover, it is possible to go several steps further and create the genealogical trees of families belonging to the same Jewish community in any given town, and eventually arrive at a „communal genealogical tree“ **[slide 21]**.

Kamila Klauzinska holds an M.A. degree in Ethnology from the Faculty of Philosophy and History, University of Lodz. Her thesis focused on the Jewish cemetery in Zdunska Wola. She is a PhD student in the Department of Jewish Studies at the Jagiellonian University in Krakow, Poland. Her dissertation is on Jewish genealogy.

http://www.judaistyka.uj.edu.pl/index.php?option=18&action=articles_show&art_id=621&lang=pl

As a researcher and consultant ***Kamila Klauzinska*** specializes in Jewish metrical data in the Books of Permanent Residents in the Polish Archives and Registry Office, mainly in Lodz, Sieradz, and Zdunska Wola. She was Co-leader of the Photographic and Topographic Census Project in the Jewish Cemetery of Zdunska Wola. She cooperates with the Museum of the History of Polish Jews in Warsaw – “Virtual Shtetl” project.