Final Report

STRATEGIES FOR THE INTEGRATION OF GENEALOGICAL DATASETS

_____

H. D. Wagner (Israel) – daniel.wagner@weizmann.ac.il
K. Klauzinska (Poland) – kamisiakk@poczta.onet.pl
J. Zajdel (Poland) - jakubzz@wp.pl

Submitted to

THE INTERNATIONAL INSTITUTE OF JEWISH GENEALOGY
December 2008

Contents of this report

1. Background and Original Objectives

A large number of local and regional Jewish genealogical databases have survived in Europe. Such databases, if suitably integrated –or merged-, can lead to an improved biographical depiction of individuals, to a more correct (and probably much larger) account of the number of victims of the Holocaust, but also to the multi-generational reconstruction of many lost branches of the communal Jewish family tree. This is the heart of the present research proposal.

The <u>general objective</u> of the present research project is to integrate details about individuals listed in different databases relevant to a town or region, thereby increasing our knowledge about these individuals to make possible the future reconstruction of their family trees.

The <u>specific objective</u> is to develop algorithms/software for full data extraction from genealogical/Jewish-oriented databases and for processing the fetched data so as to merge separate databases and contribute to the reconstruction of family trees. The approach adopted was to focus on a well-defined issue. The merging was limited to the *metrical death data* and *cemetery listings* for a specific ancestral town. This venture, restricted to two databases only, was intended to serve as a pilot project to compare results from a small-scale manually merged database with those from the algorithms developed in the project.

The <u>long-term objectives and significance</u> of this kind of research consist in developing so-far non-existent integration tools for dealing with the merging of very large Jewish genealogical databases of different types.

2. Achievements

The research work was divided in two distinct phases:

Phase 1 - The preparation by systematic manual extraction of a "well-behaved set" of metrical death data of a given town (Zdunska Wola in this case), over a number of years. This will be detailed below.

Phase 2 - The development of an automated merging procedure. The idea was to create a computer program that would merge two databases, namely the metrical database created in step (1) above, with a database of the cemetery records previously created by us. The latter

records have been produced in the recent past through a 5-year effort by a group of dedicated volunteers in the Jewish cemetery of Zdunska Wola, under the leadership of Kamila Klauzinska and Daniel Wagner. Details of Phase 2 are also detailed below.

Phase 1 – This necessary step took many months of time-consuming manual extraction of data at the local USC (Town Hall). The following Table provides a summary of the cumulated metrical data collection performed (deaths only, except for 1909), with a total of 2137 records:

YEAR	NUMBER OF RECORDS
D/1907	131
D/1908	117
B/1909	54
D/1909	114
D/1910	93
D/1911	124
D/1912	80
D/1913	71
D/1914	77
D/1923	77
D/1924	84
D/1930	78
D/1931	77
D/1932	89
D/1933	80
D/1934	86
D/1935	67
D/1936	69
D/1937	82
D/1938	78
D/1940	71
D/1941	74
D/1942	264
TOTAL	2137

Each individual metrical record included the following items (when available):

(i)   Surname;
(ii)   Given name;
(iii)   Record type (birth, marriage, death);
(iv)   Civil year of event;
(v)   Akt (record) number;
(vi)   Exact civil date of event (Gregorian or Julian, depending on period);
(vii)   Age at death;
(viii) (Approximate) year of birth;
(ix)   Father name
(x)   Mother name
(xi)   Spouse name

This set of data was to be compared and merged with the cemetery listing. The latter comprised 3505 burials/graves from 11 sections, over the 1828-1943 time frame. The data for each burial included the following, when available:

(i)   Surname;
(ii)   Given name;
(iii)   Grave location (section and position in section);
(iv)   Hebrew death date
(v)   Gregorian death date (manually translated from the Hebrew death date)
(vi)   Father name
(vii)   Possible corresponding metrical data, obtained by manual merging;
(viii) Spouse name
(ix)   Comment

As seen, the data from both data sets include some overlapping information (sometimes with spelling differences in names, and accuracy in dates and so on), as well as non-overlapping information. For example, and most importantly, the surname of the deceased often does not figure on a grave. Similarly, the mother's name almost never figures on a gravestone. These however always figure in the metrical data. On the other hand, the Hebrew date figures on the grave but not in the metrical data; other more personal information also sometimes figures on graves but not in the metrical data, such as revealing symbols, nicknames, circumstances of death (in a fire for example), and so on. This increase of information about an individual is the fundamental motivation for the merging procedure developed here. In particular, the most important task for descendants is often to identify the correct grave of an ancestor, which is usually difficult without a surname on the tombstone: details that appear both on the grave and in the metrical data (the first name, date of death, and the father name) usually lead to the correct assignment of a surname on a grave and thus to the sought after identification of the grave.

Merging with the metrical listing is therefore a necessity, even though it is often unavoidably ambiguous (dates on both data sets do not exactly fit --for various reasons such as simple mistakes, delayed registration, Gregorian date or Julian date choice, etc--, father name is absent because the grave is broken, etc). Manual merging is extremely tedious, however (it took us about 2 years to complete) and an automated procedure is necessary.

Phase 2 – The next step consisted in the elaboration of an automated merging procedure through the creation of dedicated computer software. A preliminary –tedious- task was needed as a means of asserting the correctness of the results by such a program: the manual merging of pilot sets of metrical death and cemetery data, for future comparison with automated merging. To specifically emphasize the importance of this point, in the cemetery of Zdunska Wola 3505 graves were found. Of these, only 629 graves had surnames indicated on the stones. This left 2876 graves with no surnames (82% of the total!). Manual merging, which consisted in comparing cemetery data with the metrical listing of deaths, led to the assignement of 1541 additional surnames (thus 54% of the graves with no surname were identified). The total number of positively identified graves in Zdunska Wola is thus 629 + 1541 = 2170, which represents 62% of the total (compared with the original 18% only of graves which had surnames).

Following the purchase of two dedicated software packages (Turbo Delphi and InterBase 2007), the applications were implemented:

i.   Tables were constructed (T_METRICAL and T_OTHER_SURNAME_MET), which store the data in a format more appropriate than the original EXCEL files.

ii.   A program was designed and implemented for automatic transfer of data from the metrical books (currently in EXCEL tables) to InterBase Table (source code available on demand).

iii.   A table T_CEMETERY was prepared.

iv.   The main program (source code, compilation, screenview) was constructed, using InterBase 2007.

v.   A small program was written to convert the age at death into a mathematical number.

vi.   A SOUNDEX program was created based on the Daitch-Mokotoff algorithm.

Running the merging program (termed **merge.exe**) is a simple matter and before showing some results (in the form of examples, see below) we present an overview description of the entry window that opens once the program is launched.

Upon running the merge.exe program, the opening windows looks like Fig. A (note: this picture was taken from the computer monitor, therefore the resolution is rather poor):
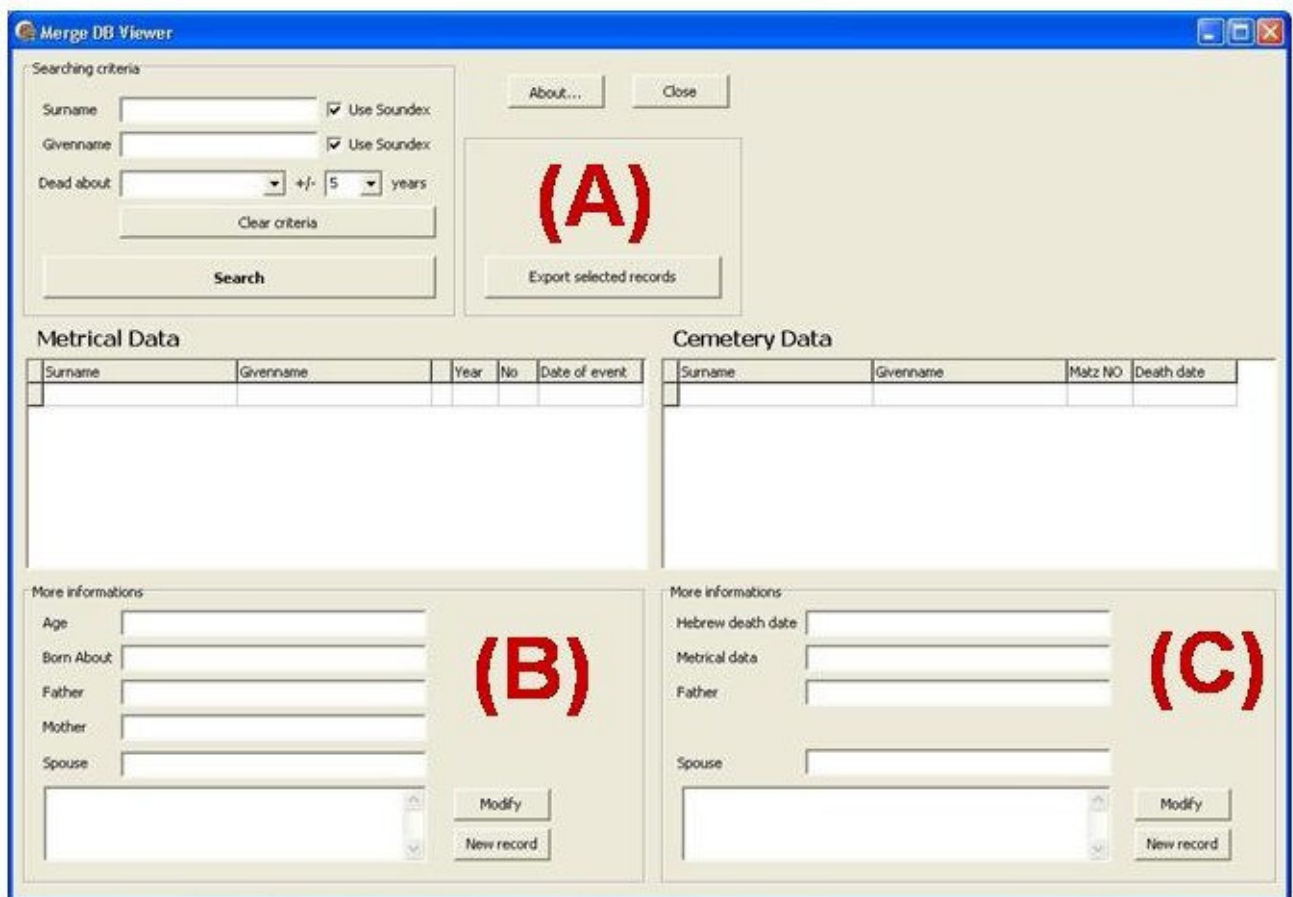

Figure 1 – Opening window of the merge.exe program.

There are three parts to this window, shown below in magnified versions:

        (i)   the upper-left part (A) is the search area;

        (ii)   the lower-left part (B) is the metrical data result area;

        (iii)   the lower-right part (C)is the cemetery data result area.
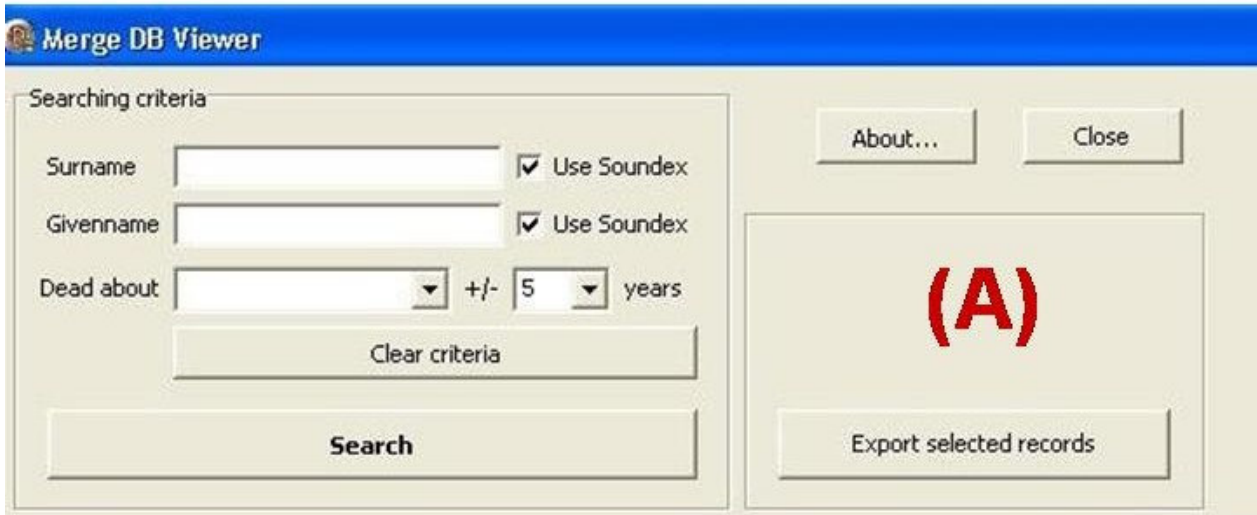

Figure 2 – Search area (or input) of the merge.exe program.

In the search area (A) the user can enter the information that is known to him/her, even as little as just a year, or a first name. The user has the option to use a soundex, and a year range if desired. Pressing the 'search' button will show all relevant results in both the metrical and cemetery data areas (the lower-left (B) and lower-right (C) parts of the window).
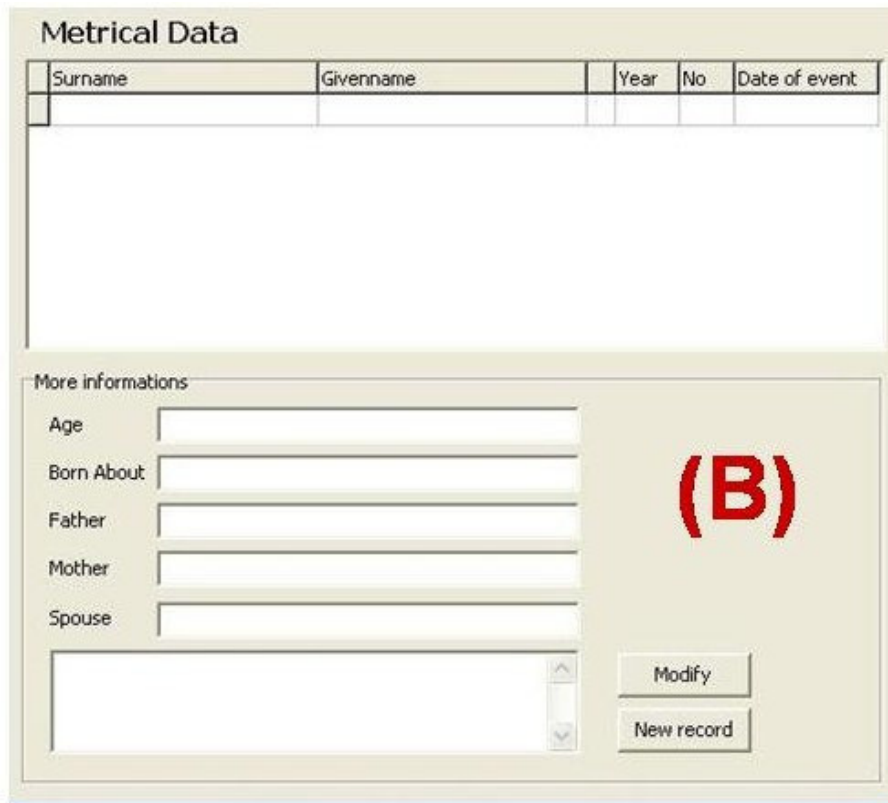

Figure 3 – Metrical data result (output) window of the merge.exe program.

Figure 4 – Cemetery data result (output) window of the merge.exe program.

All the user has to do then is to select the individual -in both the right and left windows- who best fits the request. The lower parts show additional relevant data ('more information'). Finally, the user can export the results into a merged file (by pressing the 'export selected records' button, in area A) for the searched individual.

The following examples demonstrate the use and power of the final program.

EXAMPLE 1

The first example is taken from an illustration of the merging concept described in a recent article published in Roots-Key [1] and reprinted in Avotaynu [2]. It is examined by manual merging, then by computer merging. Consider, below, the headstone of Rywka Necha, daughter of Szymon. Her date of death is 12 Iyar 5693 (?), thus 1932-1933. No surname appears.

Figure 5 - Tombstone of Rywka Necha bat Szymon, deceased on (hard to read) 12 Iyar 5693 at age 46. No surname available.

(i)   MANUAL MERGING: Searching the approx. 35,000 metrical data (for births, marriages and deaths), one finds 15 Rywka/Ryfka Necha/Nacha, with only 2 being death records. The closest 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 merging software).

(ii)   COMPUTER MERGING: If the name "Rywka Necha" is entered in the search area of the opening window of the merge.exe program (without any other information), only one result appears, identical in both windows, and therefore the merging in this case is most likely correct. The data can be exported and either saved or printed. This is shown on the next page. As seen, the cumulated information resulting from the merging process is, in this case, quite significant.

(iii)   Note that integration with other databases in Zdunska Wola would yield even more information (for example, using the Zdunska Wola Marriage metrical list, one would discover the existence of a 1913 marriage record (#40) of Abram Sucher HALPERIN with Rywka Necha SZMULEWICZ; Proper data extraction from the Marriage listings would immediately provide Rywka Necha's parents names, ages, profession etc.)

```
_____ Genealogy Merge Data Base _____
                 2009-02-22


   Metrical Data
   ------------------
   Name........... HALPERIN Rywka Nacha
   Act............ D 1936 No. 23
   Date of event.. 1936-05-03
   Born about.....
   Father......... SMULWICZ Szymon
   Mother......... LEWKOWICZ Estera
   Spouse......... HALPERIN Abram Sucher
   Comment:


   Cemetery Data
   ------------------
   Name........... HALPERIN [*] Rywka Necha
   Death date......1933-05-08
   Heb. death date.12 Iyar 5693
   Tombstone No....K-23
   Father......... Szymon
   Spouse.........
   Metrical data.. Late registr.: 1936 / D23 ?
   Comment:
   Died at 47
```

EXAMPLE 2

The second example is also taken from the Roots-Key article [1]. Again, it is first examined by manual merging, then by computer merging. Consider the small fragment of tombstone shown in Figure 6.

Figure 6 – Fragment of headstone, the date of death reads 15 Kislev 5671, the first name is in acrostic form (Sara) with a second first name starting with 'P' or 'F'. No surname is available.

(i) <u>MANUAL MERGING</u>: 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 acrostich.

(ii) <u>COMPUTER MERGING</u>: If the partial name "Sara P" is entered in the search area of the opening window of the merge.exe program (without any other information), five different results appear in the metrical records window, and eight records appear in the cemetery window, including Sara P. All "Sura Bs" (Bajla etc), which have the same soundex as Sura P, can immediately be discarded. Only three metrical data records remain: Sura Pessa MENDELZON (died 1913), Sura Perla BERKOWICZ (died 1910), Sura Perla ROSTSZTAJN (died 1914). However, since we know that the date of death is 1910, only Sura Perla BERKOWICZ remains. Had we initially entered both "Sura P" and "1910 ± 0", the only metrical records we would have obtained would have been Sura Berkowna SZUSTAK (died 1910) and Sura Perla BERKOWICZ (died 1910). The data can again be exported and either saved or printed, see below.

```
_____ Genealogy Merge Data Base _____
                 2009-02-22

    Metrical Data
-------------------
Name........... BERKOWICZ Sura Perla
Act............ D 1910 No. 83
Date of event.. 1910-12-04
Born about.....
Father.........
Mother......... DESSO Brucha
Spouse.........
Comment:

    Cemetery Data
-------------------
Name........... BERKOWICZ [*] ?? Sara  F... or P...? [Sura Perla ?]
Death date......1910-12-16
Heb. death date.15 Kislev 5671
Tombstone No....A-9
Father.........
Spouse.........
Metrical data.. 1910 / D83 ?
Comment:
Broken stone/ Fragment of text
```

EXAMPLE 3: The third example is not particularly intricate but it is interesting because it relates to an old picture found in the Yizkor Book of Zdunska Wola (p. 583), shown below in Figure 7. The only readable stone is the high one in the background on the left, which contains the mere words: "*Mirel BIRMA… Fin Shirad…*".

Figure 7 – Old photograph, date unknown, taken in the Jewish cemetery of Zdunska Wola (source: Yizkor Book of Zdunska Wola).

Here two interesting questions arise, which might be potentially solved by the merging method: (i) Is the stone shown on the old photograph still to be found among those remaining in the cemetery? (Indeed, many stones were regularly stolen from Jewish cemeteries in Poland, to be used in walkways and buildings). (ii) Can we fully identify the deceased (its full name, date of death etc)?

(i)   MANUAL MERGING: Between 1808 and 1942 the first name Mirel appears 25 times in the metrical listings, while Mirla appears 201 times. However, searching the surname BIRMA(N) or BYRMA(N) yields Mirla BYRMAN died in 1911 (death record #84).

(ii)   COMPUTER MERGING: Eleven records for Mirel/Mirla appear in the cemetery window, and seven records appear in the metrical listing window (remember that the metrical data used in the computer program include only partial (but fully extracted) XXth century data as listed in the Table above). But the results are indeed gratifying: the grave of Mirel BYRMAN can still be found in the cemetery (grave 463 in Section A, see Fig. 8 below); The metrical record number is 84 in 1911; Mirla's father was Moszek Gersz (Hirsz) GOLDBART; And most significantly, she was an old widow from Sieradz, which confirms the indication ('Fin Shirad…') on the old photograph of Fig. 7. The exported data is shown below.

```
_____ Genealogy Merge Data Base _____
                2009-02-23


    Metrical Data
-------------------
Name........... BYRMAN Mirla
Act............ D 1911 No. 84
Date of event.. 1911-08-24
Born about.....
Father......... GOLDBART Moszek Gersz
Mother.........
Spouse.........
Comment:
from Sieradz widow


    Cemetery Data
-------------------
Name........... BYRMAN [*] ?? Mirel
Death date......1911-09-06
Heb. death date.13 Elul 5671
Tombstone No....A-463
Father......... Mosze Hirsz
Spouse.........
Metrical data.. 1911 / D84 (or 1911 / D122 [JELENOWICZ] ?)
Comment:
Old woman
```


Figure 8 – Grave A-463 of Mirel BYRMAN in the Jewish cemetery of Zdunska Wola

3. Suggestions for improvements and directions for future research

The following steps should be considered in future research:

(i)   All metrical data of Zdunska Wola starting in 1808 (including births and marriages) should be fully extracted and included in the database of the program that was created in the project.

(ii)   Depending on the time frame, the language used in Polish metrical data was either Polish or Russian, and the civilian calendar used in metrical data was either Julian or Gregorian. However, the date on Jewish gravestone was in Hebrew, and translation into the Civil calendar should be performed systematically into the Gregorian calendar. In the future, a note should therefore be added to indicate this, and to indicate what was the original language used in the metrical data for a specific record. The merged result in Example 2 above clearly shows the discrepancy between both civilian calendars.

(iii)   The merge.exe software currently accommodates two input data sets, and provides data merging based on these two data sets (merging is effected by comparing data shown in areas B and C). Future extensions of the program should include more input data sets, leading to more output windows similar to B and C. For example, photographs of individuals taken from various sources (Yizkor book, applications for identity cards or passports, membership cards, and so on) could be included in a single database and merging could then provide an additional, very concrete aspect to the personality of the relevant individual. Other databases (Kahal listings, business directory listings and so on) are additional examples.

(iv)   A strategy should be developed to generate 'restricted' family trees (thus, for each surname).

(v)   A strategy should be developed to subsequently integrate all restricted family trees into 'connected trees'.

4. References:

1. H.D. Wagner, "Tombstone identification through database merging - A tool for the virtual reconstitution of vanished Jewish Communities", Roots-Key **27** (3/4) (Fall/Winter 2007), 32-34

2. H.D. Wagner, "Tombstone identification through database merging - A tool for the virtual reconstitution of vanished Jewish Communities", Avotaynu XXIV (1) (Spring 2008),8-10 (reprinted with permission from Roots-Key)

5. Financial balance ($ US)

Items	Cost	Paid?
First term: 1. Manpower for database creation/update (Klauzinska) and software programming (Zajdel): (<u>$12/hour each</u>), including travel expenses for Klauzinska & Zajdel (within Poland):	3,500	Yes
2. Software purchase for Zajdel - Interbase & Delphi:	554	Yes
Second term: 1. Manpower for database creation/update (Klauzinska) and software programming (Zajdel): (<u>$12/hour each</u>), including travel expenses for Klauzinska & Zajdel (within Poland):	5,784	No
2. Software purchase for Wagner - Interbase	60	Yes
3. Travel/accommodation expenses for Wagner (May 22-27, 2008, Berlin-Warsaw-Zdunska Wola and return to Warsaw only).	102	No