– Gedcom –

Service Programs

User Guide

Part 4

- Gedcom Start Center – in Part 1

- Gedcom Conversion – in Part 2

- Gedcom Reduction – in Part 3

- Gedcom Renumbering – in Part 1

- Gedcom Sorting – in Part 1

- Gedcom Structure Analysis – in Part 1

- **Gedcom Validate**

- Gedcom Duplicate / Merge / Sync – in Part 5

- Gedcom Listings – in Part 6

Diedrich Hesmer

Haldenstr 24

71083 Herrenberg

ofb.hesmer@gmx.de

All changes to the previous version are colored

Table of Contents

1   Introduction

This document is the part 4 of the User-Guide of the „Gedcom Service Programs". Described is the program

● Gedcom Validate
  is used to validate the data of Gedcom files. Verification can be done for

  ○ Logical errors for the date values. Available are fix comparisons (e.g. birth after death resp. marriage), where the date values will be compared 1:1. A second group of date comparisons using a flexible selection of ranges in days, months or years (e.g. birth of child after x years of birth of mother) are available. If witnesses, godfathers, etc. in the ged file are linked to the event, a check can be done, if the event took place between the majority and death of the witness, … In addition date ranges and missing dates can be checked.

  ○ Potential problems within the ged file. Groups of error checks warning checks (e.g. for the gender an erroneous swap of the sex from male/female and Mother/father can be checked) and checks for multiple tags within each record are available.

The details for

● Introduction

● Copyrights and License Conditions

● Support, Download & Installation

● Start of Program

● Gedcom Definitions

● Files used

  ○ Gedcom file (ged-file)

  ○ Language files

  ○ Protocol- / Log file

● Folders for storing files

are described in part 1 of the document.

When processing **gdz files**, the contained ged file is extracted and processed accordingly. See also chapter 2.4.4 Tab "TODO handling" for the optional modified ged file.

General information about gdz files and their loading can be found in the User Guide Part 1 in Chapter 5 "Files".

1.1   Reading the Data

After starting the validation following sequence will be followed

● Reading the data

● Checking of the data read. During this the character set used will be checked and, if necessary, criticized. For all date values, beside the check for Gedcom conformity, the year values will be checked if they are greater than the running year. Following changes to the raw data will be made without notice to the log file:

  ○ For given names surnames and place names multiple spaces will be replaced by just one.

  ○ Missing birth data resp. death data will be replaced always by christening data resp. burial data, if available.

  ○ If an attribute of a PLAC tag contains multiple commas or "comma space" sequences, those will be eliminated if th option is selected.

  ○ The actions during reading the ged-file defined by the base options.

● In case of abnormal data (e.g. empty line, missing tag, missing number + tag) an error message will pop up and the analysis may be canceled for correction of the data.

1.2   Checking of DATE values

These are described in chapter 2.2.2 Tab "Date Check options" on page 8. See also GSP User Guide 1 Chapter "Checking of DATE values" and at appendix "Date Formats".

1.3   Files

Beside the Gedcom file (ged-file) following additional files are used.

1.3.1   Validate Definition file (gvsd-file)

This file stores the options and definitions specified at the main screen. They will be created by this program and should be stored using an expressive name. It is recommended to store all gvsd-files in a common folder (e.g.. "C:\Data\Gedcom\files\").

1.3.2   Protocol- / Log file

The log file will be stored in the folder where the ged-file is stored. The name of the log file will be composed from the name of the ged-file with an attached "_val.log", e.g. "my.ged" results in "my_val.log". This file also contains all validation results. For details see the User Guide part 1 at section "Introduction".

1.3.3   Date List file

The date list file is a side by side list – see Fig 13: Tab "Ged date list" on page 22 – showing the ged file details of the 2 persons resp. person and family found as faultily. The file will be in the sub folder "\valid" of the ged file (see above). The name of the duplicate file will be composed from the name of the ged-file with an attached "_dat.html", e.g. "my.ged" results in "my_dat.html". The file will be created by pushing the appropriate buttons of the tab and may be viewed by a browser.

1.3.4   "No Date Error" File

This file is used to store date values of persons, which have been classified after manual check of the "Date List" as correct. The file name is freely definable, the file type is ".noda". The file can be expanded as often as required.
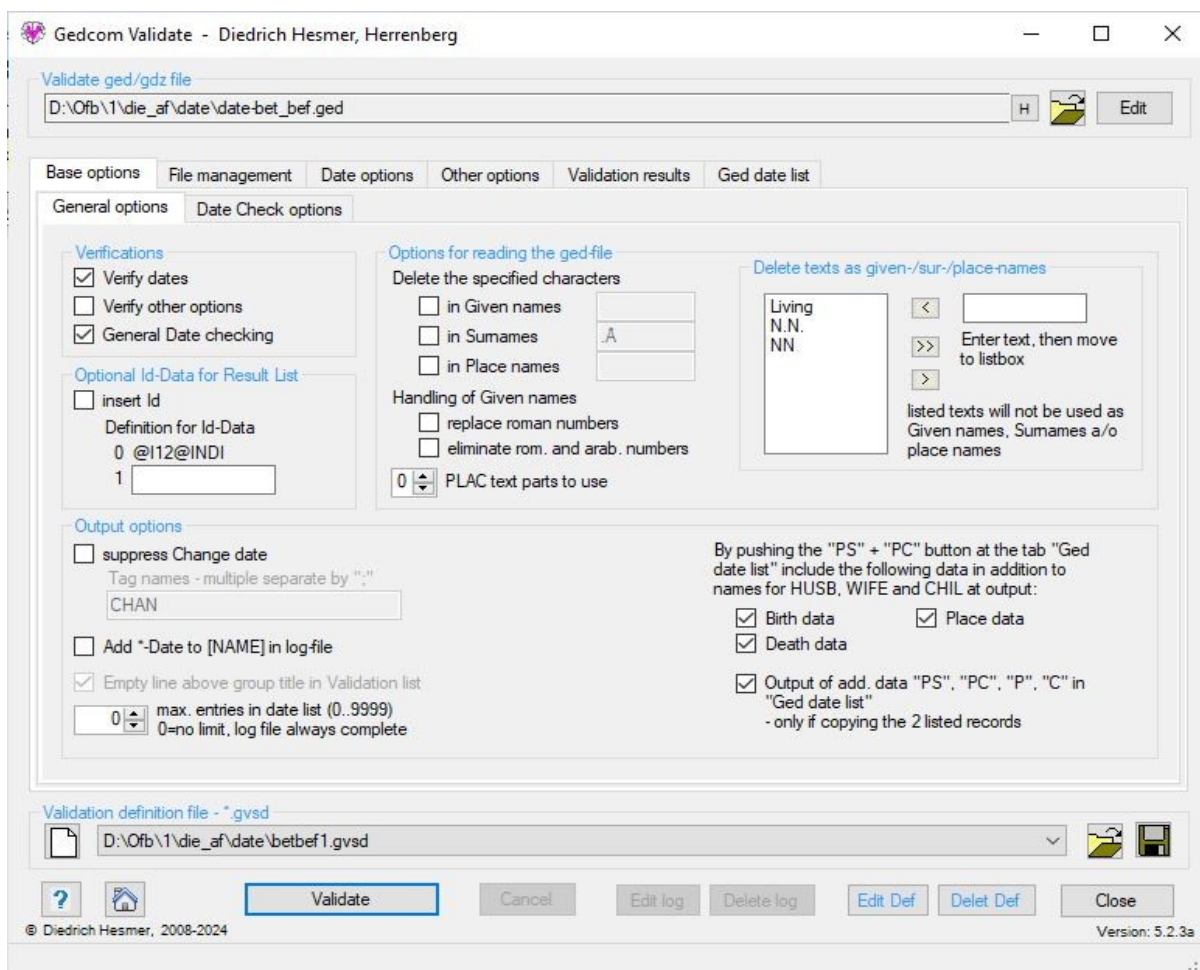
2 Gedcom Validation

2.1 Main screen

Fig 1: Main screen + Tab "Base Options"

After start of the program the following screen shows up. The name of the last loaded gvsd- file, if found in an ini-file, will be reloaded.

This screen has 6 tabs and allows the viewing and modification of the options and to watch the results of the validation.

Note: During verification of dates for missing birth data the christening data, if available, will be used and for all missing death data the burial data will be used. For all date checks simplified 30 days/months and 365 days/year will be calculated.

Above the tabs the group field for selecting the **ged file** is located:

● By 📂 the required Gedcom file will be selected

● Pushing *"Edit"* by the left mouse button opens the ged-file by the text editor for viewing, defined by the "Base-Settings" of the OFB resp. Gedcom Start Center. Alternatively the standard text editor will be used. Pushing the right mouse button starts "*GedShow*" with the ged-file.

● With the optional "*H*" button to the right of the loaded ged file the HEAD record of the ged file is displayed in the browser.

Below the tabs the management of the definion file and some general buttons are located:

● Group field for initialization, loading and saving the gvsd file:

   ○ By 🗋 all fields will be set to their initial value.

- By 📂 a previously stored validation definition file (gvsd file) will be loaded.
- By 💾 a folder (e.g. "C:\Data\Gedcom\files\") and name for the validation definition file will be defined and the file stored as "gvsd-file". This file contains all settings specified by all the tabs for reuse.

  A click on the "DropDown" Element for the gvsd-selection file opens a list with the recent used gvsd files, up to the number defined at "Settings" in the Start-Center. Selecting one of these files will load this file.

- Pushing *?* opens the user guide.
- Pushing the *house* opens the GSP homepage by the browser.
  To the right of "Help" an *Info* button (not shown here), shows up as long as no license for this program is found. After pushing the button a window pops up giving information what to do by a separate license for the program.
- Pushing *Validate* starts the reading and validation of the ged file. The results will be shown in tab "Validation Results" and will be written to a log file.
- Pushing *Cancel* will abort the started validation. It might use some seconds until the validation stops.
- Pushing *Edit log* opens the log file by the text editor, defined by the "base settings" of the OFB resp. Gedcom Start Center. Alternatively the standard text editor will be used.
- Pushing *Delete log* deletes the log file from disk.
- By *Edit Def* a file with all active definitions will be created and loaded by the text-editor defined at the OFB resp. Gedcom Start Center at "Base settings". Alternatively the standard editor will be used.
- By *Del Def* the file with the active definitions will be deleted.
- Pushing *Close* exits the program and creates an ini-file containing the filenames of the last used gvsd files for later re-use. If data have been changed a confirmation dialog for data saving will pop up.
- Messages will be written to the status bar at the bottom of the screen.

2.2 Tab "Base Options"

This tab consists of 2 tabs, described below.

2.2.1 Tab "General options"

4 group field to specify the base options are available:

- At *Verifications* the selection of the verification groups allows:
  - *Verify dates* activates the group field for date definitions
  - *Verify other options* activates the group field for other verifications
  - *General Date checking* activates the option field for the general date check in accordance to the GEDCOM specification.
- At *optional Id-Data for Validation Result List* an Id-Data for the persons affected may be added from the ged-file to the validation result list. This Id will be inserted between the individual record number (INDI number) and the name as "Id xxxx".
  - Selecting the option *insert Id* inserts the Id number.
  - At *Definition for Id-Data* the name of the tag from the ged-file has to be entered. Only level-1 tags are allowed.
- As *Options for reading the ged-file* are available:
  These options will be used for reading of the data for validation only. For a merging of persons and writing a new modified ged-file the data will be read again without applying these options.
  - *Delete the specified characters ...* – for *given names*, *surnames* and *place names* each special characters can be defined to be deleted during reading of the ged-file. The required characters have to be entered in a row. Spaces are not allowed.
    Special characters found will be deleted first from the names text.
  - *Handling of Given names* allows a modification of numbers included. In any case a closing dot "." always will be deleted:
    - *replace roman numbers* modifies those to Arabian numbers.

- ■ *eliminate rom. and arab. numbers* eliminates these numbers.
- ○ *PLAC text parts to use* defines the numbers of text parts from the PLAC tags, separated by comma. 0 = No checking, the complete text will be used.
- ○ At *Delete texts as given-/sur-/place-names* texts can be defined, which should not be used as given names, surnames or place names during reading of the ged file. These texts will be deleted and handled as missing text. Deleted will be those names which match 1:1 with the texts. These texts may contain spaces.
  - ■ Enter the text to the right text field.
    Note: Potential special characters have already been deleted from the text. This has to be considered at entering the texts.
  - ■ By `‹` the text will be moved to the list. The list is sorted.
  - ■ By `››` all texts in the list box will be deleted.
  - ■ By `›` only selected texts of the list box will be deleted.
- ● Group field *Output options*
  - ○ By *Suppress Change date* the output of change date and time can be suppressed for the screen and the new ged-file. The corresponding level-1 tags must be entered into the text box, multiple separated by a ";". Standard is CHAN, some programs are using in additional _CREAT for creation date.
  - ○ *Add *-Date to [NAME] in log-file* adds in the log-file the birth date to the name of the person.
  - ○ Empty line above group title in Validation list inserts a line for better visibility.
  - ○ By defining *max. entries in date list* the output of date errors at the tab of "Validation results" may be limited. Allowed are the values 0.. 9999. Entering "0", all date errors will be listed in the tab. Independently the log file will include all results.
  - ○ Selection of data fields to be added to main data of the concerned person at the "Ged duplicate list". The insertion will be done by the toggle *"F"*, *"I"* and *"C"* at the tab "Ged duplicate list" and only at the reference Id's of HUSB, WIFE and CHIL of the family data. The name of the persons always will be inserted.
    - ■ For Birth data + Death data the dates will be added to the names.
    - ■ For Place data the places will be added to the dates.
  - ○ By *Output of add. data ...* the behavior of button 🖹 at figure 13 on page 22 will be set to list the data of the 2 side-by-side listboxes 1:1, which includes the additional data by "PS", "PC", "P", "C" in the duplicate list.

2.2.2 Tab "Date Check options"

Fig 2: Tab "General options"

The general options for the two types of date checks are defined here.

● Group field ***Options for "Verify dates"***
These options only apply to the checks selected in the "Date options" tab.
Accordingly, all date fields starting with "2 DATE" or "3 DATE" are checked for validity only for INDI and FAM data records, regardless of the options selected, in accordance with the restrictions listed below. See also GSP User Guide 1 Chapter "Checking of DATE values" and at appendix "Date Formats". Restrictions are:

From date values with 2 dates, e.g. "FROM xxx TO xxx" only the 1st value will be used for comparison. This is also valid for text like "xxx/xxx", "/xxx" and "xxx/", where also the "-" can be used as separator. Existing prefix text, e.g. ABT, FROM, … will not be used for the comparison.

In case of multiple DATE lines for an event, only the first DATE line will be tested.

DATE values ending with "B.C." resp. BCE (before Christi) or containing none-Gregorian text, will NOT be used for validation and internally stored as "0000.00.00".

○ Group field ***List Date Text with …***

   ■ ***Free form text lines*** those date lines will be listed, who's text is completely bracketed and therefor so called "free form text" – Example:: "2 DATE (Eastern 1920)". The list will be copied to the log-file and the validation results tab.

   ■ ***BCE/B.C. - before Christi*** those date lines ending with "B.C." will be listed. They are NOT used for verification. The list will be copied to the log-file and the validation results tab.

   ■ ***None-Gregorian DATE*** those date lines, which contain none-Gregorian date values, will be listed. They are NOT used for verification. The list will be copied to the log-file and the validation results tab.

   ■ ***Suppress DATE errors***, these are not listed in the log file, but the sum of the findings is listed. This is recommended for a high number of such errors so that the other errors are recognized.

● Group field ***Options for "General Date checking"***
These options only apply to the "General date check".
All DATE lines of all data records are checked according to the GEDCOM specification. Excluded from this check are free text lines (bracketed texts), BCE and B.C. texts (before Christ) and non-Gregorian values, which are assumed to be correct without a check.

o Group field ***During Date checking ...***
- ■ ***Accept German text phrases*** rates following texts as correct:
  - ● Months shortcuts: "MÄR", "MRZ", "MAI", "OKT", "DEZ".
  - ● Date prefixes: "AB", "BIS", "CA", "CA.", "ERR", "ERR.", "GESCHÄTZT", "NACH", "SEIT", "UM", "VON", "VON xxx BIS", "VOR", "ZW xxx UND", "ZW. xxx UND", `"ZWISCHEN xxx UND"`.
- ■ ***Accept format DD.MM.YYYY*** rates this German notation as correct. As separator only the "." (dot) will be accepted.
- ■ ***Accept DATE xxxx/xxxx*** accepts date values in the form "year/year", 4 digits each.
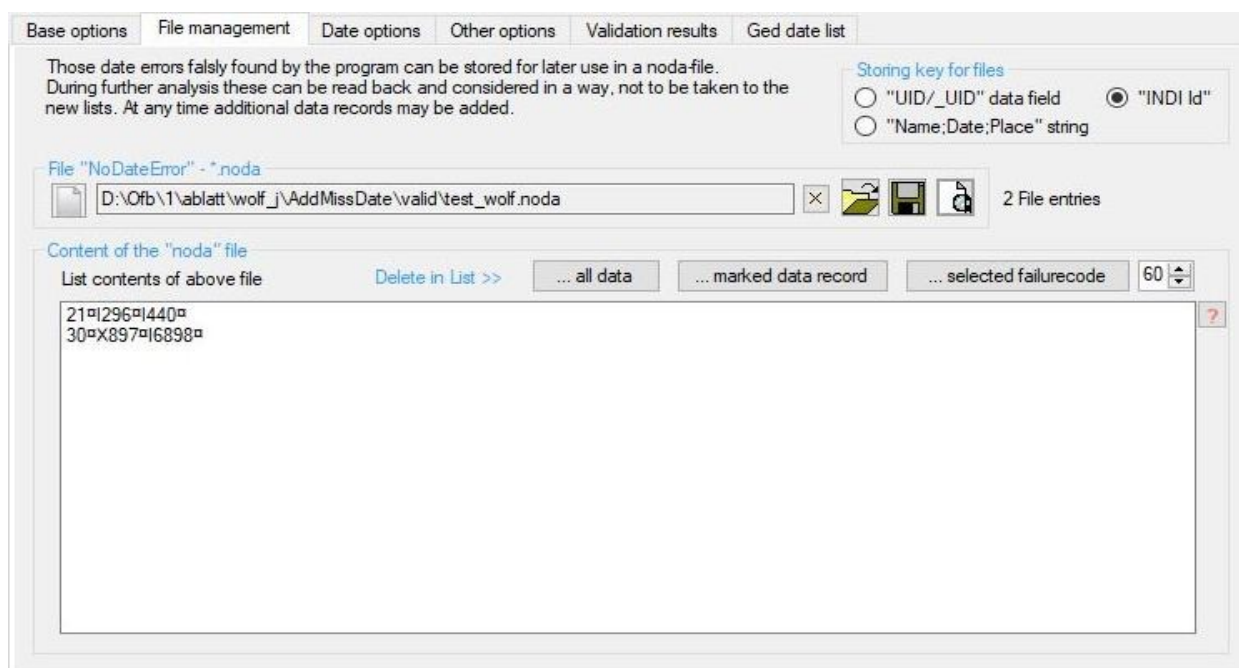
2.3 Tab "File management"

Fig 3: Tab "File management"

This allows the management of the file "*No Date Errors*" – noda-file. These file allows the storage of found date errors during validation and avoid a re-listimg of these data during a follow-on analysis.

● The selection of **Storing key for files** defines, how the pairs of persons resp. date errors will be stored, to recognize them in a later validation, even after potential modification of the data.

  o **"_UID" Data field** is most likely the best method, in case this field is provided by the genealogy program, e.g. by Ages, Legacy, PAF and Reunion. "_UID" is an unique marking.

  o **"Name;Date;Place" String** is an alternative, if the genealogy program does not provide "_UID". For each person to be stored a long string will be created by "Surname;Given Name;*Date;*Place;+Date; +Place". In case the birth data or death data are missing, the christening resp. burial data will be used. As a consequence, changing those data may result in a none recognition of a "no duplicate" marked person during a later validation.

  o **INDI Id** can only be used for ged files exported from programs with constant INDI record Id.
    Note: There are very few programs that use constant INDI Ids.

  Depending on the selected date options, additional data may be added like marriage or event data. If a NoDu line was added by the Sync function, a "¤SYNC" will be appended to the line.

● The group field **File "No Date error" - *.noda** allows the recording of dates, found by the program as potential error, but which are by a manually decision correct. The group field provides buttons to initialize, load, store and show a noda-file.

  o By ☐ a new, empty file will be created. It is filled using the **NoDat: xx** button in the Tab "Ged date list" on page 22.

  o By ☒ the text of the noda-file and their data will be deleted.

  o *By* ☞ an existing noda-file may be loaded.

  o By ☐ a folder (e.g. "C:\Data\Gedcom\files\") and name for the "No Date error"-file will be defined and the file stored as "noda-file". This file will store all pairs of persons, defined as "No Date errors".

  o By ☐ the content of the file will be shown at the list box below line by line.

  o Right beside the field the number of date errors stored in the file will be shown.

● The group field **Content of the noda file** shows, depending of the action, the content of the noda-file at the list box. In addition the text for the failure code for the marked line will be listed above the list box. 3 buttons above the list box allow a selective deletion of data records:

- By **... all data** the text of the file displayed in the list box will be deleted and the number of data set to "0".

- By **... marked data record** a marked line will be deleted from file.

- By **... selected Failure code** all records of the list box will be deleted, which comply to the code defined by the field right to the button. The first 2 chars of every line are the failure code.

- The **"?"** at right of the list box opens the Screen "Explanation of Error codes" on page 16 for displaying the error codes. By a left mouse click the failure codes of the shown codes in the list box, by a right mouse click all possible codes with their explanations are shown.

Note: Don't forget to save the file if you would like to permanently delete the content of the file.

2.4 Tab "Date options"

This tab consists of 2 tabs, described below.

2.4.1 Tab "Standard Events"

Fig 4: Tab "Date options – Standard events"

At **Date definitions** the date definitions to be verified may be defined:

● 9 options with years differences are available:
Birth child after birth mother (2x min and max), birth difference between wife and partner, birth difference between siblings (2 consecutive births by the same mother), birth difference between eldest and youngest siblings (births by the same mother), birth child before marriage of parents, birth child after birth father, marriage and death after birth.

  ○ By **only if wife older than partner** the findings for "birth between wife and partner" are only listed if the woman is older than the partner.

● 2 options with months differences are available:
Birth child after death father and after/before birth siblings.
For the latter differences up to 2 days will be count as twins and not reported.
Note: These are only checked if the year and month are contained in DATE, the day may be missing.

● 2 option with days differences are available:
Christening after birth and burial after death.
Note: These are only checked if the year, month and day are contained in DATE.

● 8 options without any time value are available:
Birth of child after death of mother, birth after death, christening after death, christening before birth, burial before death, marriage before birth and after death, divorce before marriage.

● 2 options with year values are available:
Year value > actual year and year value less than a defined value – 100 to 1000 are available.

● By the group field **No checking of ...** various checking can be skipped to improve the accuracy of the comparisons, depending on the DATE content.

  ○ **"DATE BEF YYYY"** will skip the checking if the DATE value is only BEF, followed by a years value. This will be used for all date options.

o **DATE with AFT BEF BET FROM TO …** will skip the checking if the DATE value starts with AFT, BEF, BET, FROM, TO or the corresponding German codes NACH, VOR, ZW, ZW., ZWISCHEN, AB, VON, SEIT, BIS.

o **DATE with ABT CAL EST …** will skip the checking if the DATE value starts with ABT, CAL, EST or the corresponding German codes BER, CA, CA., ERR, ERR., GESCHÄTZT, UM, UNG.

Note: DATE values containing a bracket "(" are excluded from all checks.

● By the group field **Replace missing date of …** following data, if applicable and available, will be adjusted.

o By **BIRT by CHR** missing date of birth is replaced by the date of christening. The PLAC data is replaced simultaneously.

o By **DEAT by BURI** missing date of death is replaced by the date of burial. The PLAC data is replaced simultaneously.

o By **CHR by BAPM** missing date of CHR is replaced by the date of BAPM The PLAC data is replaced simultaneously.

The specification of years, months and days are selectable in a wide ranges.

By **on/off** all options may be switched on or off.

2.4.2 Tab "Special Events"

Fig 5: Tab "Date options – Special events"

At **Events in relation to death / birth** the following checks may be defined:

● Event/fact after death (the Gedcom tag CREM will not be checked).

● Event/fact before birth.

● If a witness/godfather/... has majority and is still alive at the time of the event. (Details see below). Selecting *"Years – "majority" up to death of witness at event"* will check if the person has at least the specified age and is still living at the day of the event. For this check the link within the ged-file must be realized by the ASSO.RELA tags. (Brothers Keeper user have to convert their _EVN tags by the conversion program before use.) In addition the following may be specified:

o Number of **Years life span if 1 date is missing**. In case the birth/christening date or the death/burial date is missing, this missing date will be calculated by this number. If this value is reduced to the minimum (here 50 Years), no calculation and no comparison will be done for the affected persons.

o By **if event date is missing, use birth/marriage date** for event without date, the date of birth/christening resp. marriage will be used.

- If at least one of the options is selected, 2 text boxes opens:
  - At *Tags without checking the DATE* level-1 and level-2 tags whose date should NOT be checked, can be entered. For entering multiple tags they have to be separated by ";" without space. By the system the level-1 tags CHAN (change date), CREM (cremation) and the _CREAT (creation of data record), used by some programs, will be skipped always.
  - At *RELA values for ASSO w/o checking the DATE* the RELA texts can be entered, where no date checking should be performed. For entering multiple texts they have to be separated by ";" without space.

At *Events in relation to marriage* only family events can be checked. This check allows a comparison if a date of an event/fact is took place falsely before resp. after marriage. For multiple marriage tags the 1st will be used.

For this option events can be checked, which are coded within the ged-file by a tag or by an EVEN or FACT with a following TYPE attribute (e.g. engagement). For input 4 text boxes for entering tag names and TYPE attributes will open. For entering multiple tags they have to be separated by ";" without space.

The *Init* button will set the text boxes to their system values. For the tags these are the values shown at the figure, the TYPE values will be blanked.

2.4.3 Tab "Time ranges / Missing dates"

Fig 6: Tab "Dat options – Time ranges / Missing dates"

The group *List persons in a selected time range for ...* allows the checking of date values, so that persons will be listed whose dates are within the defined limits.

- Selectable are the events *Birth*, *Christening*, *Death a*nd *Burial*.

- Following logical *Relations* for the comparison results of the events are available:
  - *And:* Each selected event must be within the defined limits.
  - *Or:* At least 1 of the selected event must be within the defined limits.

- As range the *from*-Date and *to*-Date has to be defined in the format „*YYYYMMDD*" without separation characters. Erroneous input will be marked in red color.

Note: A none-available date will be stored internally as „00000000". This allows to list persons with no date value at all by using „00000000" for the range of dates and „And" as relation.

The group *List persons with missing dates for ...* allows the checking of missing date values, so that persons will be listed whose dates for defined events are missing. Selectable are the events:

- **Birth and Christening** - the person will be listed, if both events are without date.
- **Birth or Christening** - the person will be listed, if at least 1 of the events are without date.
- **Death and Burial**   - the person will be listed, if both events are without date.
- **Death or Burial** - the person will be listed, if at least 1 of the events are without date.
  - By **… and also older than xx years** the group of persons found can be further reduced for the 2 Death/Burial options, so that only persons with missing death- and burial date, but who are older than the defined age will be listed. By this persons still alive with a known birth- or christening date can be eliminated.
  - **Also list, if tag is missing** lists such persons, if the appropriate tag does not exist and therefore also no date exists. If not selected, the person will only be listed if the tag BIRT CHR DEAT BURI exists but does not contain a date.
- **Birth and Christening and Death and Burial** - the person will be listed, if all 4 events are without date or events are missing. This option is alternative and excludes the above.
- **Marriage** - the person will be listed, if a MARR Tag exists without a date.
- **Divorce** - the person will be listed, if a DIV Tag exists without a date.
- By **Search only for defined Surnames** the group of persons found can be further reduced, so that only persons with a defined surname will be listed. For this input the desired surnames. Multiple names to be separated by Semicolon ";" without spaces. Wildcard characters "*" and "?" are allowed

2.4.4   Tab "TODO handling"

Fig 7: Tab "Date options – TODO handling"

This allows an insertion of a level-1 TODO structure with the text of the Validation results for the selected DATE findings.

At **Handling of ToDo entries into ged-file** can be defined, if and which results of the date checking should be added at the end of every concerned INDI resp. FAM record according the format defined below.

- For storing following findings are selectable
  - Options of the tab "**Standard events**" except "Christening after birth", "Burial after death", "Birth before marriage of parents", because very sophisticated data can exist.
  - The above excluded "Christening after birth", "Burial after death", "Birth before marriage of parents".
  - Options of the tab "**Special events**" … after death and before + after marriage.
- By **Delete all existing TODO …** all existing level-1 TODO entries according to the selected format will be

deleted during transfer from the original file to the new modified file. I.a. they will be replaced by the new analysis.

Group field ***The modified file will be saved as***

The new created ged-file will be saved under the name of the original file with an appended "_val" ("My-data.ged" will become "mmy-data_val.ged").

For gdz files, the name of the gdz file with appended "_val" is used for the modified ged file which is stored in the folder of the gdz file. The gdz file is NOT modified.

● ***TODO Storing*** defines the format for storing the findings. "xxx" stands for the text of the Validation results list. Available are:

  o ***No ToDo***: Nothing will be stored, even if the above options are set. This option may not be selected if the delete option is selected above.

  o ***_TODO.DESC xxx (Standard)***: This is the Standard setting. The data are stored in the format:
```
1 _TODO
2 DESC DATE: xxx
2 STAT auto
```

  o ***_TODO xxx***: This setting should be used only in special cases. The data are stored in the 1-line format:
```
1 _TODO DATE: xxx
```

  o ***EVEN.TYPE TODO+.NOTE xxx***: This setting should only be used if the standard setting is not understood by the receiving program of the ged-file. The data are stored in the format:
```
1 EVEN
2 TYPE TODO
2 NOTE DATE: xxx
```

  o ***EVEN.TYPE TODO+.CAUS xxx***: This setting should only be used if the standard setting is not understood by the receiving program of the ged-file. The data are stored in the format:
```
1 EVEN
2 TYPE TODO
2 NOTE DATE: xxx
```

  o By ***Save ged*** the modified file will be saved manually.

  o By ***Edit*** the file will be opened by the text-editor.

  o ***Automatic saving of the ged-file*** saves the file at the end of the validation automatically..

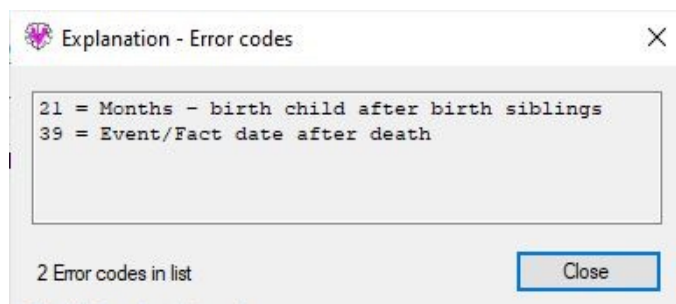## 2.4.5 Screen "Explanation of Error codes"

Fig 8: Screen "Explanation of Error codes"

This screen lists dynamically the error codes of the listed items. It is a permanent screen and can be moved to any place of the screen to be visible side by side to the output screen. It will be opened by the small button "?" on Tab "File management" on the right side.

The button ***Close*** closes the screen.

2.5 Tab "Other options"

Here 2 tabs are available:

2.5.1 Tab "Tag checking"

Fig 9: Tab "Other options - Tag checking"

● Group field **Checking - Multiple existing tags** checks the existence of multiple tags in every data record

  ○ **Names** checks NAME

  ○ **Birth, Christening, Death, Burial** checks BIRT, CHR and BAPM, DEAT, BURI.

  ○ **Child in family (FAMC)** checks the multiple existence of tag FAMC for a person. However, this is allowed for different child to family links. Only those FAMC with no sub-tags will be counted.

  ○ **Partnership, Divorce** checks MARR, DIV.

  ○ **Event, Fact** checks for the combinations of "EVEN.TYPE xxx" resp. "FACT.TYPE xxx".
    This selection activates a text box, where those TYPE texts xxx can be entered, which should not be tested for multiple occurrences, e.g. Residence.

  ○ **Other lvl-1 tags** checks all level-1 tags specified in the text field. The tag names already displayed in the group field cannot be checked here. This option and the following option are mutually exclusive.

  ○ **All others** checks all other tags, e.g. OCCU, RESI, … The tags FAMS, CHIL, CHAN will be skipped. This selection activates a text box, where those level-1 tags can be entered, which should not be tested for multiple occurrences, e.g. ASSO. This option and the option above are mutually exclusive.

● Group field **Checking - Not existing tags / texts** checks the absence of tags in every data record

  ○ **FAMS + FAMC (single persons without familiar relations)** checks the absence of the 2 stated tags in every INDI record, i.e. persons without any relation to a family of parents resp. partner will be listed.

  ○ **HUSB + WIFE + CHIL (family with no data about persons)** checks the absence of the 3 stated tags in every FAM record, i.e. families without any relation to persons will be listed.

  ○ **HUSB/WIFE/CHIL only 1 entry (no family)** checks if only 1 entry of the 3 stated tags are existing, i.e. families with only 1 person will be listed.

  ○ **HUSB + WIFE (family without partners, only children)** checks the absence of the 2 stated tags in every FAM record, i.e. families only with children, but no parents, will be listed.

- o **HUSB|WIFE + CHIL only 1 parent + child(ren)** checks if only 1 parent exists, i.e. families with only 1 parent will be listed.
- o **SEX tag not existing** checks the existence of the tag SEX in the INDI records.
- o **NAME tag not existing** checks the existence of the tag NAME in the INDI records.
- o **INDI/FAM Tags with missing text** checks, if for following level-1 and leve-2 tags the attribute text is missing.
  - GEDCOM 5.5.1
    - Level-1 INDI: CAST, DSCR, EDUC, FACT, IDNO, NAME, NATI, OCCU, PROP, REFN, RELI, RESN, SEX, SSN, TITL
    - Level-1 FAM: REFN, RESN
    - Level-2 INDI + FAM: ADDR, AGE, AGNC, CAUS, DATE, PEDI, PLAC, RELA, RESN, STAT, TYPE
  - GEDCOM 7.x
    - Level-1 INDI: CAST, DSCR, EDUC, FACT, IDNO, NAME, NATI, OCCU, PROP, REFN, RELI, RESN, SEX, SSN, TITL
    - Level-1 FAM: REFN, RESN
    - Level-2 INDI + FAM: ADDR, AGE, AGNC, CAUS, DATE, PEDI, PLAC, RELA, RESN, STAT, TYPE
- o **INDI/FAM Tags with invalid text** checks, if for following level-1 tags no attribute text or only a single "Y" exists.
  - GEDCOM 5.5.1
    - Level-1 INDI: BIRT, CHR, DEAT
    - Level-1 FAM: MARR
  - GEDCOM 7.x
    - Level-1 INDI: BAPM, BARM, BASM, BIRT, BLES, BURI, CENS, CHR, CHRA, CONF, CREM, DEAT, EMIG, FCOM, GRAD, IMMI, NATU, ORDN, PROB, RETI, WILL
    - Level-1 FAM: ANUL, CENS, DIV, DIVF, ENGA, MARB, MARC, MARL, MARR, MARS
- ● Group field **Checking – Special tags** checks the existence of the selected tags within every INDI + FAM record.
  - o **List all RESN tags** checks the existence of RESN tags (restricted use of the data). All by RESN marked records and events will be listed.
  - o **List all SEX unequal F or M** checks the existence of SEX, if the type is not F or M, and lists all those found.
- ● Group field **Error – Missing tags / records / references / pointer** checks for missing records, references and pointer within the whole ged-file.
  - o **Record missing** checks for every pointer (e.g. HUSB @I123@) of the ged-file, if the related data record exists. The pointer @VOID@ will NOT be criticized.
  - o **Record orphaned** checks for every record of the ged-file, if at least one related pointer (e.g. HUSB @I123@) exists.
  - o **Defective ref: FAMS↔HUSB/WIFE + FAMC↔CHIL** checks for every FAMS pointer if the appropriate HUSB/WIFE pointer and for every FAMC pointer if the appropriate CHIL pointer exists. And vice versa.
  - o **Multiple identical record-Id's** checks for every record of the ged-file, if the record number exists multiple.
  - o **Missing TYPE tag for EVEN/FACT/IDNO** checks the existence of "2 TYPE xxx" after "1 EVEN" resp.. "1 FACT" for individual (INDI) and family (FAM) records and "1 IDNO" for INDI records.
  - o **Missing sub-tag {1:n}** checks if mandatory sub-tags acc. Gedcom 5.5.1 are existing at least once {1:n}.

2.5.2 Tab "General checking"

Fig 10: Tab "Other options – General checking"

● Group field ***Checking – Names***
   Note: All NAME-related checks refer to the first NAME entry found.

   o ***Surname missing*** checks empty name fields.

   o ***Surname contains comma*** checks if this character exists.

   o ***Surname only 1 chars*** checks on surnames with only 1 character.

   o ***Given name missing*** checks empty name fields.

   o ***Given name*** only 1 chars checks on surnames with only 1 character.

   o ***only 1 "?" in given name or surname*** checks the question mark as the only character. For this the question mark may not be deleted during reading of the ged-file (see "Base options").

   o ***Surname in capital letters*** checks if surname is in capital letters.

   o ***same surname for husband + wife*** checks for same names.

   o ***no point after initial s in given name*** checks the existence of a point after a single character.

   o ***no space after point in given name*** checks for a space after each point.

   o ***Special chars in names field*** checks the existence of the specified special chars in given name and surname.

      ▪ The ***Special chars*** have to be entered without any space. Spaces are not allowed.

      ▪ For ***multiple*** a warning will be indicated, if several special chars are found or individual special chars several times.

      ▪ For ***any number*** a warning will be indicated already if one special chars will be found one times.

   o ***Text phrase in names field*** allows the check of any text fragments in given names and surnames. In case of checking multiple text fragments, these fragments have to be separated by a ";".

   o ***Check _RUFNAME*** checks for correct matching. "Rufname" is the underlined first name in German birth certificates for multiple first names, mandatory during certain periods. The system checks whether there is only one first name as a "Rufname" and whether this occurs in NAME.

● Group field ***Checking - Sex***

- o *Verify sex male/female* – persons, defined in the ged file as wife, must have the sex code "F" and husbands must have "M" . Errors will be listed in the log.
- o *List same sex partnerships* lists the FAM-Id and the stored Ids and names of HUSB and WIFE and their sex value.
- ● Group field *Checking – Other warnings*
  - o *Path in FILE for gdz file* checks at gdz files the path specifications of FILE at the contained gedcom.ged file. All locally stored media files are expected without path specification.
- ● Group field *Checking Records / Pointer*
  - o *Empty records* lists all records that consist of only 1 line with no additional text.
  - o *Multiple identical records* is checking for NOTE, SOUR, OBJE, _LOC, …, but not for INDI + FAM the existence of records with identical content and will output this in the following format:

```
*** Multiple record analysis ***
Base record + identical records
------------------------------------------------
0 @N24@ NOTE unehelich geboren
  Identic are Id: @N79@, @N80@, @N81@, @N82@, @N83@, @N85@, @N86@,
0 @N29@ NOTE Nottaufe
  Identic are Id: @N30@, @N31@, @N32@, @N33@, @N50@, @N51@, @N54@
0 @N45@ NOTE ledig geboren
  Identic are Id: @N46@, @N47@, @N48@, @N142@, @N143@
0 @N108@ NOTE ledig
  Identic are Id: @N133@, @N140@, @N141@, @N148@, @N149@, @N161@,
0 @N111@ NOTE geht nach Amerika
  Identic are Id: @N112@, @N498@, @N503@
0 @N115@ NOTE unehelich geboren, Vater unbekannt
  Identic are Id: @N126@, @N511@, @N512@
```

Fig 11: Output format "Multiple records"

If such records have been found, these can be adjusted by the Conversion program.

2.6 Tab "Validation results"


```
Base options | File management | Date options | Other options | Duplicate options | Validation results | Ged date list | Ged duplicate list

Created on 01.07.2009 21:09:05
Base: "C:\Daten\1\die_af\AF_dh_neuTest.GED" dated: 24.06.2009 20:38:26
Line-nbr: @Record-Id@ -> Comment or Gedcom Tag -> faulty  => converted
    58: @I2@ => 1 CHR => 2 DATE xxxxx 24 APR 1945 => 1945.04.24
 Lines read - 560
 Lines faulty - 1

*** found records (quantity type):
23 INDI, 10 FAM, 0 NOTE, 0 SOUR, 0 OBJE, 1 SUBM

*** Validation results ***
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++

*** Verification Other options ***
----------------------------------------------------------------
Verify sex male/female:
 => For INDI nbr "I7" wrong sex "M" => should be "F"
Surname missing:
 => For INDI nbr "I8"
 Faulty entries - 2

*** Verification Date (YYYY.MM.DD) ***
----------------------------------------------------------------
DATE: sequ nbr: xxx =>
                                    Results found faulty: File  1, Others 2, Date 2, Duplicates 2
```

Fig 12: Tab "Validation results"

The tab is one output area showing the results of reading the ged file and the validation. The same text will be written to the protocol/log file in the folder where the ged file resides.

All possible duplicates will be numbered sequentially. The same numbers will be used at the next tab for selecting the duplicate persons.

At >100 duplicates the screen output will be canceled due to performance reasons, but all results are available by the log file.

On the right side below of the output area the number of findings will be shown.
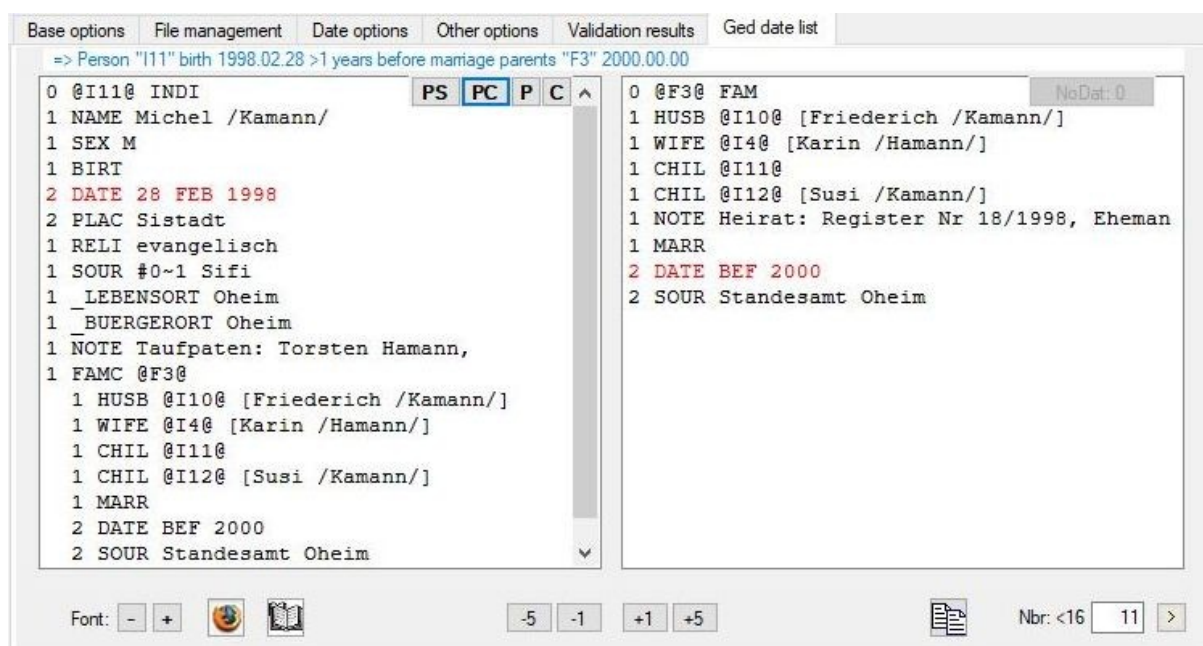
2.7 Tab "Ged date list"

Fig 13: Tab "Ged date list"

This tab has 2 output areas side by side showing the ged file details of the two persons resp. family found as faulty for the date values. The date values in question are marked in **Red** color, in some cases in **DarkSkyBlue**. Above the 2 areas the respective line as shown at the "List of Validation results" is listed.

This representation should simplify the manual review of the correctness and a potential correction of the original data. This correction has to be done by the genealogy program where the original data reside.

The buttons at the top border of the list boxes are available for:

● **"F"** toggles the output of family record defined at the FAMS tag. This is the family, where the person is a spouse and shows the data of his/her partnership and the children. The content is listed indented directly after the FAMS line.

● **"C"** toggles the output of family record defined at the FAMC tag. This is the family, where the person is a child and shows his/her parents and siblings. The content is listed indented directly after the FAMC line.

● **"P"** toggles, if "F" resp. "C" is active, the output of the persons data defined at the tab "Duplicate options" for those persons defined by the tags "HUSB" resp. "WIFE" (partner). The data will be, enclosed by […], attached to the concerning line.

● **"I"** same as "P", but for the childs defined by the tag "CHIL".

By pushing the button **NoDat: xx** the date error shown at the screen will be stored as "No Date error" at the noda-file. This date error will no longer be listed as potential date error during a later validation as long as those noda-file will be used as reference where this error has been stored. The "xx" is the number of actual stored date errors in the current noda-file.
Note: This button is only active if a noda-file has been defined by the tab "File management" and a program license has been found on the PC.

The buttons above the bottom border will allow to

● Change the **Font** size of the 2 output areas.

● By 🦊 the **browser** will start and opens the duplicate file, if created by one of the next 2 buttons.

● By 📖 all **date errors** will be listed side by side in a date list file (see page 4). The data will be appended to an already opened file, otherwise a new file will be created. After adding of all data the date list file will be closed and may be opened and viewed by a browser.

● By 📄 only the **actual shown pair date error** will be copied to the date list, optional the additional data by "PS", "PC", "P", "C". During the initial call after a validation a new file will be created. As many date errors as you like may be copied to this open file. The file remains open. She only will be closed by

pushing the buttons "Browser", "Validate" or "Close" or after pushing "all date errors" and all have been appended to the open file. Thereafter the file may be opened and viewed by a browser.

- The desired sequence **Nbr** can be entered at the input field on the right side and by pushing ⏵ the data of the 2 affected persons will be listed at the output areas.
- In the middle are various buttons, depending on the number of found duplicates, with different link targets. By activating you will jump back and forth according to the number on the button and the concerning number is shown at the right field. This last 2 items should help for a fast navigation.
- Clicking in the 2 list fields at a line with a record Id, this record (NOTE, SOUR, _LOC, ...) will be listed in a separate screen, described below. This is not established for INDI and FAM.

2.8   Data Record screen

Fig 14: "Data record screen"

In this screen the content of several data record may be displayed. This can be NOTE, SOUR, OBJE, REPO, _LOC or other records. INDI and FAM records can't be displayed, because they are already part of the validation screen. The screen will be held in the foreground to have the data visible always. Any number of screens can be opened in parallel and the access to the Validation screen is still possible.

- **Close** closes the screen.

This screen can be increased to any size.

2.9   Remarks for validation

For the validation by this program following parameter of the ged file are important:

- Given name and surname. Only the tag "NAME" will be used, no sub tags.
- Birth, christening, death, burial: date and place of the event. For christening the tags "CHR" or "BAPM" may be used. In case of multiple occurrences only the first will be used.
- Marriage/partnership: date of the event. In case of multiple occurrences only the first will be used.
- Events/Facts: date of event.
- Does the date contain 2 dates , e.g. from xxx to yyy resp. between xxx and yyy, only the first date xxx will be used. All date prefixes themselfs, e.g. "abt", "bef", "aft" …, will not be considered but their date values.
- Father/mother/child(ren) relations of the ged file within a family/partnership.
- Sex of a person.
- Record Id numbers and their related pointer.

All further data will not be used for validation.