



Prüfung und Umwandlung von PLAC Daten

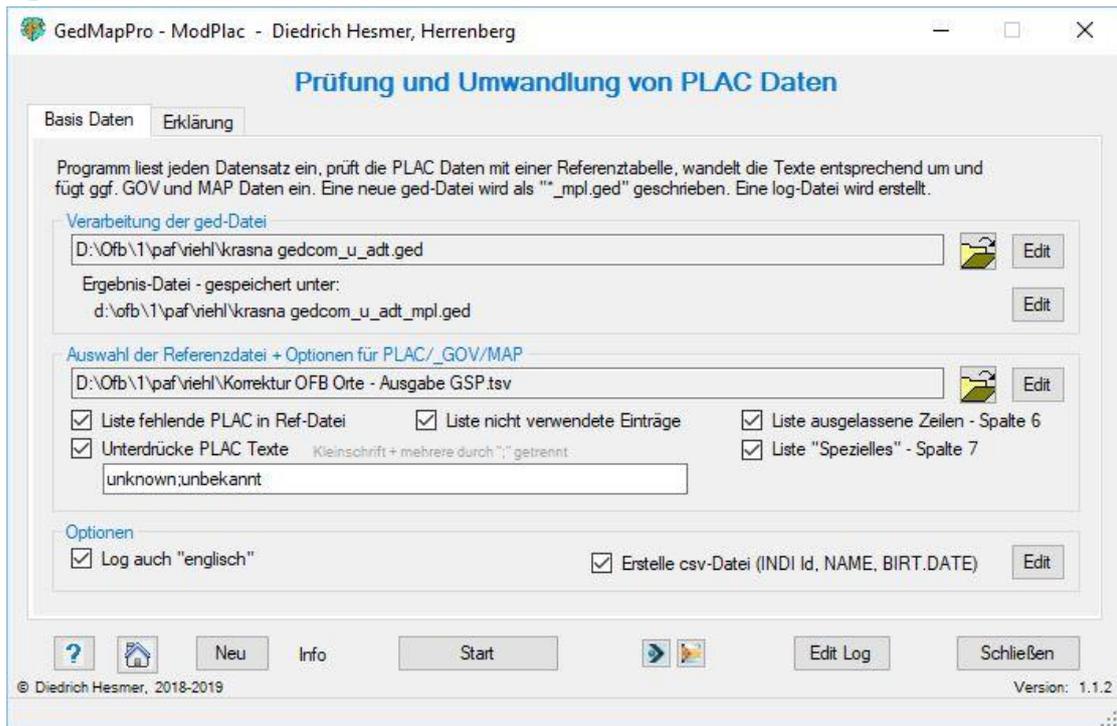
Inhaltsverzeichnis

1	Hauptfenster.....	2
1.1	Funktionalität	2
1.2	Beschreibung	2
1.3	Verarbeitung.....	2
1.3.1	Vorbereitung der ged-Datei.....	2
1.3.2	Einlesen der Daten.....	2
1.3.3	Ausgabedateien	3
1.3.4	PLAC Referenzdatei	3
1.3.5	Optionen.....	3
1.3.6	Spezielle Eigenschaften	4
1.3.7	Schaltknöpfe	4
1.4	Fix kodierte Umwandlungen	4
1.4.1	Allgemein	4
1.4.2	INDI + FAM.....	4
1.5	Erklärung des Ablaufs	4
1.6	Logdatei	6
1.6.1	Beispiel.....	6
2	Anhang.....	7
2.1	Urheberrechte, Lizenzvereinbarungen, Programmfehler und anderes	7
2.2	UTF-8 Kodierung und Prüfung	7
2.3	PLAC Referenzdatei – Beispiel	7
2.4	Konvertier Steuerdatei: "date.gcsd"	8

Änderungen zur letzten Ausgabe sind gelb hinterlegt.



1 Hauptfenster



1.1 Funktionalität

Das Programm ergänzt ged-Dateien, die diverse Daten nicht enthalten, mit diesen Daten. Ausgangspunkt waren ged-Dateien aus PAF 4.0. Dazu modifiziert es PLAC Daten der INDI und FAM Datensätze entsprechend einer Referenzdatei.

Das Programm liest jeden Datensatz ein, prüft die PLAC Daten mit der PLAC Referenztable, wandelt die Texte entsprechend um und fügt ggf. GOV¹ und MAP Daten ein und schreibt eine neue ged-Datei. Eine log-Datei mit den Ergebnissen wird erstellt.

1.2 Beschreibung

Nach Programmstart erscheint obiges Fenster. Nach Wahl der ged-Datei, der Referenzdatei und der Optionen erfolgt die Verarbeitung durch "Start". Die Textfelder der Dateien sind pink, sofern die Datei nicht existiert.

Die beiden farbigen Schaltknöpfe zwischen "Start" und "Edit Log" erscheinen, sobald Programm DiffMerge u/o WinMerge – siehe Map Homepage > Hilfsprogramme – auf dem PC installiert ist. Mit diesen können die beiden ged-Dateien miteinander verglichen werden. Bei kleinen Dateien ist der linke, bei großen Dateien der rechte Schaltknopf empfohlen. Bei sehr großen Dateien ist das aus Speicher- und Zeitgründen jedoch fast unmöglich.

1.3 Verarbeitung

1.3.1 Vorbereitung der ged-Datei

Zur Absicherung eines einwandfreien Laufs des Programms ist die Original ged-Datei bei Bedarf mit dem GSP-Konverter vorher zu korrigieren. Hierbei werden die Kodierung auf **utf-8** geändert und die fehlerhaften DATE Werte entsprechend angepasst – siehe "Erklärung des Ablaufs" auf Seite 4 und im "Anhang" ab Seite 7.

1.3.2 Einlesen der Daten

Die PLAC Referenzdatei (*.tsv) wird eingelesen und wird intern für die Vergleiche und Übertragung vorbereitet. **Achtung:** Es wird davon ausgegangen, dass die Datei eine **UTF-8** kodierte Textdatei ist. Siehe hierzu auch Kapitel 2.2 UTF-8 Kodierung und Prüfung im Anhang.

¹ GOV: Geschichtliche Orts-Verzeichnis, ein Projekt der "Verein für Computergenealogie, e.V." (CompGen)



Gedcom Map Programm – "ModPlac"

Die ged-Datei wird datensatzweise eingelesen. INDI und FAM Datensätze werden entsprechend den Daten der Referenzdatei und den gewählten Optionen umgewandelt. Daher setzt dies ebenfalls eine **utf-8** kodierte ged-Datei voraus. Alle anderen Datensätze bleiben unverändert. Die einzelnen Datensätze werden als neue ged-Datei ausgegeben. Damit ist die Reihenfolge der Datensätze beider Dateien identisch.

1.3.3 Ausgabedateien

Die Originaldatei bleibt unverändert.

Die modifizierte Datei wird im Verzeichnis der Originaldatei mit dem Dateinamen der Originaldatei mit dem Zusatz "_mpl" ausgegeben. Aus "Original.GED" wird "original_mpl.ged".

Eine log-Datei "original_mpl.log" (deutsch) und optional "original_mpl_en.log" (englisch) wird erstellt mit den Fehlern und der Anzahl der jeweiligen Veränderungen entsprechend den gewählten Optionen und ist im Ordner der ged-Datei gespeichert.

1.3.4 PLAC Referenzdatei

Diese ist eine normale Textdatei im **tsv**-Format (mit Tabulator als Separator) und enthält die nachfolgenden Daten. Sie ist, entsprechend des persönlichen Bedarfs, selbst zu erstellen und besteht aus 7 Spalten, mit Kopfzeile (in dieser 1. Zeile muss am Textanfang "Alt" stehen!):

1. PLAC Text alt – enthält den Originaltext der ged-Datei von PLAC und ist das Schlüsselfeld. Es wird zum Vergleich des PLAC Textes der ged-Datei verwendet. Es sind nur solche Texte einzutragen, die korrigiert bzw. ergänzt werden sollen.
2. PLAC Text neu – ersetzt den alten PLAC Text (aus Spalte 1) der ged-Datei.
3. _GOV Kennzeichen – wenn gefüllt, wird nach PLAC als "3 _GOV xxx" eingefügt.
4. LATI Text – wenn gefüllt, wird nach PLAC als "3 MAP" mit "4 LATI xxx" eingefügt.
5. LONG Text – wenn gefüllt, wird nach PLAC als "3 MAP" mit "4 LONG xxx" eingefügt.
6. Auslassen Kommentare – wird für unbekannte Ortsangaben verwendet:
Eine Reihe Ortsangaben ist wie überliefert erfasst. Eine Prüfung zum GOV führte zu keiner Lokalisierung. Dies wird hier kommentiert. Wenn hier ein Eintrag vorkommt, wird diese Zeile ausgelassen. Damit ist protokolliert, dass man sich um diese Zeile bereits bemüht hat. Der PLAC taucht in der log-Datei unter "Folgende PLAC Texte der ged-Datei wurden in der PLAC Referenzdatei nicht gefunden" nicht mehr auf.
7. Spezielle Texte, wie z.B. Tippfehler
Hier sind Tippfehler festgehalten, die im Datenbestand aus dem die ged-Datei exportiert wurde, nachgebessert werden müssen. Eine Merkliste also. Interessant, wenn mehrere an einem Bestand arbeiten und sich abstimmen müssen.

Die Spalten 1+2 müssen gefüllt sein, der Rest kann leer sein. Beispiele dazu im Anhang.

1.3.5 Optionen

Verfügbar sind für:

- **PLAC/_GOV/MAP** Verarbeitung
 - **Liste fehlende PLAC in Ref-Datei**
Listet die PLAC Texte, die nicht in der Ref-Datei angegeben sind, mit Datensatznummer, bei INDI mit NAME, und mit dem betreffenden Stufe-1 Tag.
 - **Unterdrücke PLAC Texte ...**
Unterdrückt in der "*Liste fehlende PLAC in Ref-Datei*" alle PLAC Texte, die im Textfeld angegeben sind. Die Texte können in Kleinschrift angegeben werden, mehrere durch Semikolon trennen.
 - **Liste nicht verwendete Einträge**
Listet die Zeilennummer und die Spalte "1" der Referenzdatei, deren Daten nicht verwendet wurden.
 - **Liste ausgelassene Zeilen**
Listet die PLAC Texte, sofern unter Spalte "6" ein Eintrag ist, die nicht in der Ref-Datei angegeben sind, mit Datensatznummer, bei INDI mit NAME, und mit dem betreffenden Stufe-1 Tag.



Gedcom Map Programm – "ModPlac"

- **Liste "Spezielles"**
Listet für gefundene Orte, sofern unter Spalte "7" ein Eintrag ist, die PLAC Texte und Spezielles Texte mit Datensatznummer, bei INDI mit NAME, und mit dem betreffenden Stufe-1 Tag.
- **Optionen**
 - **Log auch "englisch"** erstellt die Logdatei mit der Zusammenfassung und den Listentiteln zusätzlich in englischer Sprache.
 - **Erstelle csv-Datei (INDI Id, NAME, BIRT.DATE)** erstellt die csv-Datei "*_mpl.csv" im Ordner der ged-Datei mit den 3 angegebenen Spalten. Die erste Zeile ist eine Kopfzeile und enthält die Titel.

1.3.6 Spezielle Eigenschaften

Das Programm ist vorbereitet für das OFB/GSP Dienstprogramm "BatchRun". Es kann mit der Command Line Angabe für die "ini" Datei und dem Schalter "/,q" von diesem aufgerufen werden.

Die entsprechende Zeile der Textdatei mit den Programmaufrufen für BatchRun muss enthalten:

```
"Programmname_mit_kompletten_Pfad"[,"ini-Dateiname_mit_kompletten_Pfad"][,/q]
```

ohne jegliches Leerzeichen.

Beispiel:

```
"C:\Program Files (x86)\OFB\ModPlac.exe","C:\Users\xxx\AppData\Roaming\OFB\ModPlac.ini",/q
```

```
"C:\Program Files (x86)\OFB\ModPlac.exe",/q
```

In Zeile 1 werden beide Parameter übergeben und es wird die angegebene ini-Datei verwendet.

In Zeile 2 ist keine ini-Datei angegeben und es wird die im Programm gespeicherte ini-Datei verwendet.

1.3.7 Schaltknöpfe

- [Symbol mit geöffnetem Ordner] Laden einer ged-Datei bzw. Referenzdatei.
- [Edit] mit linker Maustaste öffnet die ged-Datei, die Referenzdateien im Text-Editor bzw. die csv-Datei im Standardprogramm für csv-Dateien. Bei ged-Dateien startet "GedShow" bei Klick mit der rechten Maustaste mit der ged-Datei.
- [?] Aufruf Programmbeschreibung.
- [Neu] löscht die Einträge und setzt die Systemwerte.
- [Info] zeigt den Speicherort der Handbücher, der Lizenzdatei und der ini-Datei (enthält die Einstellungen vom letzten "Schließen").
- [Start] Lesen der ged-Datei, Starten der Analyse und Veränderung der Daten, Schreiben der geänderten ged-Datei.
- [Edit Log] Editieren der Log-Datei(en) im Standard Texteditor.
- [Schließen] Schreibt die ini-Datei mit den Einstellungen (diese wird beim nächsten Start des Programms automatisch geladen) und beendet das Programm.

1.4 Fix kodierte Umwandlungen

Fest im Programm kodierte Umwandlungen oder Streichungen sind, unterteilt nach Datensatztyp:

1.4.1 Allgemein

- Fehlerhafte DATE Texte in der Form "ABT dat1 TO ABT dat2" werden umgewandelt in das Format "FROM dat1 TO dat2".

1.4.2 INDI + FAM

- Noch keine.

1.5 Erklärung des Ablaufs

Folgender Ablauf ist bei jeder Übertragung neuer Daten erforderlich:

1. Export ged-Datei aus dem Genealogieprogramm, z.B. PAF.
2. Bei fehlerhaften Datum Werten DATE Korrektur mit dem GSP-Konverter mit entsprechender Steuerdatei (Bsp. "date.gcsd") und der original ged-Datei.
3. Nach Abschluss der Konvertierung starten "ModPlac.exe" mit der korrigierten "*_u. ged". Ist keine Korrektur erforderlich, starten mit der original ged-Datei.



Gedcom Map Programm – "ModPlac"

4. Ggf. prüfen der Datei mit GSP-Validierer auf logische Datum Fehler.
5. Nach Abschluss umbenennen "*_u_mpl.ged" in beliebigen Dateinamen und Weitergabe.

Beachte: Es erfolgt keine Prüfung der ged-Datei auf Struktur- oder Datenfehler! Hier helfen die "GSP Analyse- und Validier-Programme".



1.6 Logdatei

1.6.1 Beispiel

Erstellt am 03.09.2018 14:09:06 - ModPlac-Version: 0.4.0
Basis: "C:\Ofb\1\paf\riehl\krasna gedcom_u.ged" vom 02.09.2018 17:53:42
Ergebnisdatei: "c:\ofb\1\paf\riehl\krasna gedcom_u_mpl.ged"

03.09.2018 14:09:06: Start lesen der Ref Plac-Datei
2557 Einträge in PLAC Referenzdatei gefunden
03.09.2018 14:09:06: Start lesen der Ref Sour-Datei
695 Zeilen mit 693 Einträge in SOUR Referenzdatei gefunden
03.09.2018 14:09:06: Start lesen der ged-Datei

81.444 Änderungen:
40.364 alt/neu Austausch
39.935 _GOV eingefügt
429 MAP.LATI/LONG eingefügt
6 OCCU eingefügt
21 NOTE eingefügt
689 SOUR eingefügt

Listeneinträge:
770 PLAC nicht in Ref-Datei (1.971 Einträge in ged-Datei)
135 kein Text in PLAC Zeile
127 PLAC Zeilen ausgelassen (161 Einträge in ged-Datei)
39 Spezielles in PLAC Zeilen
20.466 INDI ohne zusätzliche NOTE OCCU SOUR

Folgende PLAC Texte der ged-Datei wurden in der PLAC Referenzdatei nicht gefunden:

```
-----  
1: "Old" St Philomena Cem,Selfridge,Sioux ND,USA -> @I12920@ [Johanna /Heidrich/]: BURI  
2: ,,CA,USA -> @I13119@ [Martin /Kuntz/]: DEAT, @I15735@ [Nickolaus /Stocke/]: DEAT ...  
3: ,,LATV -> @I9318@ [Otto /Paul/]: BURI  
...
```

#####

Folgende Zeilen der PLAC Referenzdatei wurden nicht verwendet:

```
-----  
1: 49: ,,USA  
2: 65: ,Bender,BESS,RUSS  
...
```

#####

Folgende Zeilen der PLAC Referenzdatei wurden ausgelassen:

```
-----  
1: 139: Acpenar,Tuftscha,DOB,ROMN - "nicht gefunden" -> @I9188@ [Kartella /Schreiber/]: BIRT  
2: 148: Alexandrowka,Bender,BESS,RUSS - "kann nicht zugeordnet werden" -> @I11515@ [Theodor /Kalinicenko/]:  
BIRT  
3: 155: Alt-Grabonaka - "nicht gefunden" -> @I9738@ [Karl /Hermann/]: BIRT  
...
```

#####

Folgende Zeilen der PLAC Referenzdatei beschreiben Spezielles/Tippfehler:

```
-----  
1: 200: Akkermann,Akkermann,BESS,RUSS - "Akkerman" -> @I819@ [Anton /Chmotowski/]: BIRT, @I819@ [Anton  
/Chmotowski/]: CHR, @I4645@ [Constancia /Chmotowski/]: BIRT, @I4645@ [Constancia /Chmotowski/]: CHR, @I14060@  
[Johannes /Pantschenko/]: BIRT  
2: 275: Aschaffanburg,BAV,GERM - "Aschaffenburg" -> @I12907@ [Katharina /Wuitschik/]: DEAT  
3: 606: Cetata Alba,Akkermann,BESS,RUSS - "Akkerman" -> @I6098@ [Reinhold /Brudet/]: BIRT  
...
```

#####

Folgenden INDI Id's wurden keine NOTE, OCCU oder SOUR der SOUR Referenzdatei zugefügt:

```
-----  
1: @I1@ LIVING  
2: @I5@ Dorothea /Dirk/  
3: @I10@ Martin /Riehl/  
4: @I13@ Anton /Kahl/  
...
```

03.09.2018 14:14:12: Abschluss der Konvertierung
326.988 - Zeilen gelesen aus Basis-Datei
373.067 - Zeilen geschrieben in Ergebnis-Datei

Weitere Listenblöcke können bei entsprechenden Fehlern enthalten sein.



2 Anhang

2.1 Urheberrechte, Lizenzvereinbarungen, Programmfehler und anderes

Diese und weiter relevanten Informationen finden Sie im Anhang der "Map-Start" Beschreibung. Sie gelten hier uneingeschränkt.

2.2 UTF-8 Kodierung und Prüfung

Die Referenzdatei muss **utf-8** kodiert sein. Es gibt die Kodierungen "utf-8 mit BOM²" (Standard) und "utf-8 ohne BOM". Zur Prüfung kann beispielsweise das Programm "**Notepad++**" (siehe Homepage > Hilfsprogramme) verwendet werden.

Zur Prüfung laden Sie die Datei, die Kodierung wird danach unten rechts angezeigt als UTF-8-BOM bzw. UTF-8. Alle anderen Kodierungen führen zu Fehlern bei der Darstellung von Umlauten und anderen sprachspezifischen Zeichen.

2.3 PLAC Referenzdatei – Beispiel

Die Texte der letzten 2 Spalten sind teilweise gekürzt.

Beachte: Die erste Zeile muss mit "Alt" beginnen.

Alte Bezeichnung	Neue Bezeichnung	_GOV	Latitude	Longitud	Auslassen	Speziell / Hinweis
,,,ND	,North Dakota,US-ND	object_276685				
Bitola,Mace,SLVK	Bitola,Mazedonien,MKD	object_376450	41.03333	21.33333		
Borki,Charkiv,UKRA	Borki, Kharkiv Oblast,UA-63		49.71806	36.0552		
Bonn,NRW,GERM	Bonn,DE-NW	BONONNJO30NR				
Bonn,GERM	Bonn,DE-NW	BONONNJO30NR				
Neuwied,Neuwied-Block,RPF,GERM	Block,Neuwied,Mayen-Koblenz,DE-RP	BLOACH_W5451	N50.4325	O7.5075		
Parsenplug,,GERM					Lt Fragebogen von 1941	
Borghorst,RPF,GERM					not in RPF SHO or NRW	
,Wraschko					not found - Gerlinde Stiller's family cards	
Kulertascha,Akkerman,BESS,RUSS	Kulertascha,Akkerman, Kulewtschea,Akkerman,UA-51					Kulewtschea ?
Rural Haynes,Hettinger,ND,USA	Rural Haynes,Hettinger,US-ND	HETGERDN86QA				Haynes or Hettinger
Wosnesenski,Akkerman,BESS,RUSS	Wosnesenski,Akkerman, Wosnessensk,Beresan,UKR					Wosnessensk,Beresan ?

² BOM: Als Byte Order Mark wird eine spezielle Bytefolge am Anfang eines Datenstroms bezeichnet. Diese dient als Kennung zur Definition Kodierungsform in Unicode-Zeichenketten, insbesondere Textdateien.



2.4 Konvertier Steuerdatei: "date.gcsd"

Diese ist beispielhaft für eine ged-Datei des PAF Programms erstellt. Entsprechend den Fehlern der eigenen ged-Datei ist eine eigene Steuerdatei zu erstellen.

Erstellt am 10.01.2019 00:49:26 - GedcomConv: 4.1.6
Gedcom-Konvertierung - gewählte Optionen
Datei: "D:\Ofb\1\paf\rieh\date-orig.gcsd" vom: 08.12.2018 17:13:35

Die Verarbeitung erfolgt für jeden Datensatz in der nachfolgenden Reihenfolge

Gruppe: -
Kodierung: Original = ANSI -> UTF-8
Gruppe: 3.1
Deutsche Ungenauigkeiten > Gedcom Form
Gruppe: 3.3
Datum JJ01/02 > BET JJ01 AND JJ02
"/" = Trennzeichen "/" zwischen Jahreszahlen
Gruppe: 5.1
"Verschiebe Text":
"1 DEAT#2 DATE Y" -> "1 DEAT Y"
"1 DEAT#2 DATE y" -> "1 DEAT Y"
"1 DEAT#2 DATE CHILD" -> "1 DEAT Y#2 NOTE deceased as child"
"1 DEAT#2 DATE INFANT" -> "1 DEAT Y#2 NOTE deceased as infant"
"1 DEAT#2 DATE STILLBORN" -> "1 DEAT Y#2 NOTE deceased at birth"
"1 DEAT#2 DATE DECEASED" -> "1 DEAT Y"
"1 CHR#2 DATE Y" -> "1 CHR Y"
Gruppe: 5.2
"Verschiebe Tags":
"FAM#MARR#DATE NOT MARRIED" -> "FAM#_STAT"

Dies ergibt folgende Änderungen (beispielhaft – Datum Fehlerliste gekürzt):

Erstellt am 10.01.2019 00:49:32 - GedcomConv: 4.1.6
Basis: "D:\Ofb\1\paf\rieh\Krasna Gedcom.ged" vom: 12.08.2018 12:00:53
Zeilen gelesen aus Steuerdatei "D:\Ofb\1\paf\rieh\date-orig.gcsd"

7 "Verschiebe Text" + 1 "Verschiebe Tag" Anweisungen

Zeile-Nr neu: DATE bemängelt => DATE verwendet

12151: @I885@ => 1 BIRT => 2 DATE 15 AUG => 1:1 [Rosa /Kosolowski/]
180356: @I13441@ => 1 BURI => 2 DATE Spring 1990 => 1:1 [Barbara /Dirk/]
234755: @I17627@ => 1 DEAT => 2 DATE Betw 1817 and 1830 => 1:1 [Johannes Diebold /Rieh/]
245289: @I18298@ => 1 DEAT => 2 DATE Oct 1996 => 1:1 [John C. /Kosolofski/]
253475: @I18891@ => 1 CHR => 2 DATE 10 OCT => 1:1 [Veronica /Deichert/]
253478: @I18891@ => 1 DEAT => 2 DATE Pre 1969 => 1:1 [Veronica /Deichert/]
263332: @I19676@ => 1 GRAD => 2 DATE ABT 1961 TO ABT 1965 => 1:1 [Steven George /Rieh/]
311825: @F3114@ => 1 MARR => 2 DATE Common-law => 1:1

330.208 Zeilen eingelesen + verarbeitet aus "D:\Ofb\1\paf\rieh\Krasna Gedcom.ged"

109.055 "DATE" Zeilen analysiert:
957 Struktur + 0 "JJJJ AND J" + 952 "JJ01/02" + 0 "JJJ_" geändert
12 Zeilen bemängelt
3.572 "Verschiebe Text" Zeilen geändert
31 "Verschiebe Tags" Zeilen geändert
3305 x 1 DEAT#2 DATE Y |=| 1 DEAT Y
1 x 1 DEAT#2 DATE y |=| 1 DEAT Y
253 x 1 DEAT#2 DATE CHILD |=| 1 DEAT Y#2 NOTE deceased as child
9 x 1 DEAT#2 DATE INFANT |=| 1 DEAT Y#2 NOTE deceased as infant
0 x 1 DEAT#2 DATE STILLBORN |=| 1 DEAT Y#2 NOTE deceased at birth
1 x 1 DEAT#2 DATE DECEASED |=| 1 DEAT Y
3 x 1 CHR#2 DATE Y |=| 1 CHR Y
31 x FAM#MARR#DATE NOT MARRIED |=| FAM#_STAT
Kodierung geändert von "ANSI" in "UTF-8"

326.900 Zeilen neu geschrieben nach "d:\ofb\1\paf\rieh\krasna gedcom_u.ged"
10.01.2019 00:49:45: Abschluss schreiben neue ged-Datei