

openRGD - sammanslagning och Gedcom-export

Version 1.1

Anders Ardo
Anders.Ardo@gmail.com

February 4, 2017

1 Inledning/Översikt

openRGD använder (och sparar i databasen) endast namn, född, död, gift samt relationer - resten av information sparas i originalform som ett Gedcom-fragment.

Målet för produktions RGD är att behandla nedanstående uppgifter/Gedcom taggar enligt tabellen. Kolumnen 'matchning' indikerar om en uppgift ska användas vid matchningen. Så mycket som möjligt av detta ska implementeras i openRGD. Kolumnen 'openRGD' anger var i processen (jmf UI) detta sker eller är tänkt att ske.

Uppgift	Gedcom	typ	matchning	sammanslagning	openRGD
namn	NAME	event	*	slås samman	7. Sammanslagning
kön	SEX	event	*	slås samman	7. Sammanslagning
född	BIRT	event	*	slås samman	7. Sammanslagning
döpt	CHR	event	BIRT saknas	slås samman	8. Skapa Gedcom
gift	MARR	event	*	slås samman	7. Sammanslagning
död	DEAT	event	*	slås samman	7. Sammanslagning
begravd	BURI	event	DEAT saknas	slås samman	8. Skapa Gedcom
flyttad	-				
inflyttad	IMMI	event		alla unika sparas	8. Skapa Gedcom
utflyttad	EMIG	event		alla unika sparas	8. Skapa Gedcom
levde	RESI	event		alla unika sparas	8. Skapa Gedcom
text	NOTE	text		alla unika sparas	8. Skapa Gedcom
special	NOTE			alla unika sparas	8. Skapa Gedcom
yrke	OCCU			alla unika sparas	8. Skapa Gedcom
skild	DIV	event		alla unika sparas	8. Skapa Gedcom
händelse	EVEN	event		alla unika sparas	8. Skapa Gedcom
	alla andra			alla sparas	8. Skapa Gedcom

1.1 Status

Huvuddelen av det som behandlas i detta dokument (inklusive frågorna nedan och algoritm 2.3.1) är implementerat, men inte uppladdat till DIS server. Ett undatag är kommentaren (av Anders H) om viktning/prioritering i avsnitt 2.3. Koden är inte heller testad speciellt mycket.

2 Frågor

2.1 TYPE i event

Vissa händelser (event) innehåller förutom DATE, PLAC, och SOUR även en TYPE - ska den behandlas som de andra?

Redmine RGD Design 1211 anger att det kan finnas en massa andra taggar också.

Följande Gedcom-taggar har förekommit i BIRT, DEAT, MARR events i filer i RGD:

CONT, _PLC, TEXT, _PRIM, CONC, FILE, DATE, _SIZE, DATA, _CONF_NOTE, ADR1, _PRNT, _SCBK, TITL, _STYLE, LONG, _NAME, CAUS, SOUR, PHON, _TYPE, QUAY, FORM, CITY, _APID, OBJE, PLAC, POST, MAP, ADDR, TYPE, NOTE, LATI, PAGE

Kalle:

TYPE bör absolut inkluderas. Den innehåller förklarande information, t.ex. före eller utom äktenskap vör en relation. Men TYPE kan också innehålla tillägg som släktforskaren gjort för att förtydliga. Taggen EVEN skall alltid ha TYPE för att tala om vilken typ av händelse det är.

En händelse kan bara ha max 1 TYPE

2.1.1 Vid sammanslagning av en event ska taggarna DATE, PLAC, SOUR, och TYPE tas hand om.

2.2 Redmine RGD Design 3514

Vid sammanslagning av händelser kan man komma att välja DATE från en händelse (A) och PLAC från en händelse (B). Vad ska då SOUR för den sammanslagna händelsen sättas till?

1. SOUR från A
2. SOUR från B
3. 2 st SOUR (från A och B)
4. SOUR = RGD med en NOTE som anger SOUR A och SOUR B

Hur generaliserar man detta till sammanslagning av N händelser?

Kalle:

PLAC och SOUR bör hänga samman. SOUR i sig själv ger inte alltid en komplett källa men PLAC och SOUR tillsammans skall alltid göra det. Detta gäller främst när kyrkböcker används som källa.

Datum skulle kunna tas från den andra posten om datumet är mer komplett. Som i ett exempel du tidigare visade att posten med bästa källan bara hade årtal men att den andra posten hade komplett datum.

2.2.1 Se avsnitt 2.3.1

2.3 Redmine RGD Design 1943 + mail - Välja data vid sammanslagning

Källan (SOUR) kan eventuellt, efter maskinell kvalitetsbedömning börja med asterisk och en siffra Maskinellt *1 till *9, förbered för manuellt satt *0 som bästa värde.

Om de poster som skall mergas båda har ett visst data skall:

- källan med lägsta siffra efter asterisken användas
- om bara en källa har asterisk, använd den
- om båda har samma värde efter asterisken används den med mest kompletta texten
- om båda saknar asterisk används den med mest kompletta texten

(hur ska 'mest kompletta texten' tolkas maskinellt? Den längsta?)

Kalle:

Mest komplett text är troligen inte alltid den bästa men det är svårt att sätta en regel. Om uppgifterna i övrigt, datum och ort, är lika kanske förslaget från punkt 4 i 2.2 användas. Är datum och ort i konflikt kanske man kan använda principen att det är den filen man gör sammanslagningen till som "vinner". Datum och plats på matchade individer bör dock inte släppas igenom manuella matchningen.

Anders H:

En ytterligare viktning, som jag tror nämndes någonstans, är att den databas man matchar TILL, har en något högre prioritet vid konflikter.

2.3.1 Algoritm för att välja data vid sammanslagning av händelser (event)

Sammanlagning av databaser B och A till databas A (se också 2.3).

Välj **PLAC** och **SOUR** från den händelse som har högst kvalitet.

Vid flera likvärdiga och händelsen från A tillhör dessa välj **PLAC** och **SOUR** från A.

Annars välj **PLAC** och **SOUR** slumpmässigt bland de händelser som har högst kvalitet.

Välj det mest kompletta **DATE**.

Vid flera likvärdiga välj **DATE** bland de med högst kvalitet (med prioritet för händelsen från A).

Texten från alla unika **TYPE** adderas till en samlad **TYPE** rad.

Övriga taggar ignoreras.

2.4 Vilken källa ska anges

Vi har inte bestämt vilken källa som kan/skall användas i utfilen.

Original källa så som den skrivits i den "bästa" källan (detta är väl normalläget)

Den manipulerade källan som eventuellt innehåller Asterisk och nivåvärde. Den innehåller också vissa förkortningar expanderade på ett standardiserat sätt.

2.4.1 Förslag

Inkludera bäge: Original källa som SOUR och manipulerad källa som NOTE.

2.5 Redmine RGD Design 3511 -Ursprungsmarkering vid sammanslagning

För att poster vid sammanslagning skall kunna identifieras av användaren finns en NOTE tagg.

Denna information bör bli mer informativ för att passa vid sammanslagning av flera filer. Förslag:

Original följt av namnet på filen och som nu identiteten på individen xempel:

1 NOTE Original Kalle 26-10457

1 NOTE Original Gun 14-190

Ger spårbarhet till vem som bidragit till posten och den identiteten i den personens släktforskning. Är det ingen sammanslagen post finns också motsvarande information