

hcc[!]genealogie

**Gedcom: een oplossing of
een probleem?**

verscheen eerder in Gens Humana 2006

Inhoudsopgave

1. Inleiding	3
2. De elementen van de taal	4
2.1 De componenten.....	4
2.2 De samensteller en andere inleidende gegevens	4
2.3 Gegevens van personen.....	6
2.4 Gegevens van relaties	9
2.5 Bronnen	10
2.6 Betrouwbaarheid.....	13
2.7 Het grote geheel	14
3. De problemen	15
3.1 Hoe standaard is "volgens de standaard"?	15
3.2 Hoe beheersbaar zijn uw gegevens?	17
3.3 Andere (technische) problemen	18
3.3.1 Codepages.....	18
3.3.2 Bestandsformaten	19
4. De "oplossingen"	19
5. Op GEDCOM gebaseerde programma' s	20
5.1 Controleren en repareren.....	21
5.2 Bekijken	21
5.3 Splitsen en Privacy filter.....	21
5.4 Website bouwen	22
5.4.1 Dynamische websites	22
5.4.2 Statische websites	22
5.5 Afdrukken.....	22
5.6 Data-mining.....	22
5.7 Overige bewerkingen	23
6. Meer informatie	23

1. Inleiding

GEDCOM (GEnealogical Data COMmunication standard) is door de leden van de Amerikaanse tak van de Mormonen (Kerk van Jezus Christus van de Heiligen der Laatste Dagen) gedefinieerd. Het was en is bedoeld als hulpmiddel voor het opsturen van de gegevens van de leden van die kerk over hun voorouders. Die gegevens zijn van belang voor hun geloof en worden daarom in Utah centraal bewaard.

Overigens zijn deze gegevens ook via Internet op te vragen. (www.familysearch.org) Hierop kom ik later nog terug.

Zoals bij elke standaard heeft het een aantal jaren en versies geduurd voordat er een stabiele versie was, maar de huidige versie 5.5 gaat al mee vanaf januari 1996. Hoewel er sinds 2002 sprake is van het ontwikkelen van een nieuwe versie, op een geheel andere (XML) grondslag, zit hier weinig voortgang in. Dit artikel gaat dan ook in zijn geheel over de versie 5.5

Het moge duidelijk zijn dat de GEDCOM-standaard ook buiten het beperkte domein van de Mormonen wordt gebruikt. In feite is het een verplicht onderdeel voor elk programma dat genealogische data opslaat om GEDCOM te kunnen inlezen en aanmaken.

De opbouw van het artikel is als volgt: in 'de elementen van de taal' wordt de basis beschreven. De meeste gebruikte delen van de taal komen hier aan de orde, en een korte vermelding van wat er nog meer is, maar eigenlijk nooit gebruikt wordt. Dit zou u kunnen betitelen als het deel dat GEDCOM neutraal beschrijft

In 'De problemen' en 'De oplossingen' wordt aangegeven wat er voor complicaties kunnen zijn en hoe deze soms omzeilt kunnen worden. GEDCOM als probleem dus, met de oplossingen.

De overige hoofdstukken spreken voor zich.

Wim Rozendaal

2. De elementen van de taal

In dit hoofdstuk treft u een beschrijving van de taal aan, toegelicht met eenvoudige voorbeelden. Het is de bedoeling dat u na het lezen van dit hoofdstuk eenvoudige bestanden kunt begrijpen. Maar de echte complexe zaken blijven buiten beschouwing.

2.1 De componenten

GEDCOM is bedoeld om gegevens uit te wisselen. Gegevens van stamboomonderzoekers, liefst goed gedocumenteerd, en zonder onbegrijpelijke 'eigen' manieren van weergeven.

Welke componenten heb je minstens nodig om de bovenstaande beschrijving te voldoen?

1. Gegevens van de samensteller van het bestand
2. Gegevens over de personen in het bestand
3. Gegevens over de relaties tussen die personen
4. Gegevens over de bronnen van het bovenstaande
5. De mogelijkheid om aan te geven hoe betrouwbaar de gegevens zijn (volgens de samensteller van het bestand) Elk van deze punten wordt hieronder in het kort toegelicht. Vooraf nog de vermelding dat een GEDCOM bestand op regel 1, kolom 1 altijd 0 HEAD heeft staan (gevolgd door wat technische regels, die bij de samensteller worden toegelicht) en op de laatste regel, kolom 1 altijd 0 TRLR

2.2 De samensteller en andere inleidende gegevens

Hieronder ziet u de eerste regels van een GEDCOM bestand zoals het wordt aangeemaakt door PAF, het gratis programma om stamboom gegevens vast te leggen van de Mormonen (<http://www.kerkvanjesuschristus.nl/downloads/index.php>).

```
0 HEAD
1 SOUR PAF
2 NAME Personal Ancestral File
2 VERS 5.2.18.0
2 CORP The Church of Jesus Christ of Latter-day Saints
3 ADDR 50 East North Temple Street
4 CONT Salt Lake City, UT 84150
4 CONT USA
1 DEST Other
1 DATE 6 Jul 2006
2 TIME 11:44:33
1 FILE vb.ged
1 GEDC
2 VERS 5.5
2 FORM LINEAGE-LINKED
1 CHAR UNICODE
1 LANG English
1 SUBM @SUB1@
0 @SUB1@ SUBM
1 NAME Wim Rozendaal
```

1 ADDR straat 12
2 CONT woonplaats
2 CONT provincie
2 CTRY Nederland
1 PHON 0123456789
1 EMAIL rozendaa@xs4all.nl

Dit stukje code geeft al direct weer hoe een bestand is opgebouwd. Elke regel bevat een nummer (maximaal 99), gevolgd door ofwel een label (startend en eindigend met @) ofwel een code voor het type gegevens (ook wel Tag genoemd). En als laatste de gegevens. Een regel met het nummer 0 zit op het hoogste niveau in de hiërarchie en is een zelfstandige eenheid. Een regel op nummer 1 zegt iets naders over de eenheid op nummer (ook wel level genoemd) 0, 2 geeft een extra detail over 1 etc. Er zitten dus twee zelfstandige eenheden in het bovenste blok: 0 HEAD en 0 @SUB1@ SUBM 0 HEAD beschrijft het bestand, maar de meeste gegevens die hier kunnen staan worden in de praktijk weggelaten. In dit geval staan er dus gegevens over het programma waarmee het bestand is aangemaakt (SOUR met alle details), de bestemming van het bestand (DEST), de datum en tijd van het aanmaken van het bestand en de naam waaronder het is aangemaakt. De regel 1 GEDC geeft aan dat het om GEDCOM gaat en de volgende welke versie van de standaard is gebruikt. 1 CHAR geeft aan dat de lettertekens in het bestand zijn gecodeerd volgens het Unicode-systeem De laatste regel van dit blok zegt de samensteller (SUBM) gedetailleerd staat beschreven onder het label @SUB1@.

Het tweede zelfstandige blok bevat een eerste regel met een label (@SUB1@) en de aanduiding dat het om informatie over de inzender van het bestand gaat (SUBM). Dit lijkt niet alleen dubbel op; het is dubbel op. Want in de regel er boven staat al dat het label @SUB1@ over SUBMitter data gaat. Toch is dit de consequent toegepaste werkwijze bij GEDCOM, die er voor zorgt dat een bestand slechts eenmalig hoeft te worden gescand bij het verwerken, hoewel het niet is te voorkomen dat soms pas nadat het label is gedefinieerd wordt aangegeven waar het gebruikt wordt. Met de huidige computers is een tweede scan geen groot probleem, maar met de apparatuur uit b.v. 1990 toen de standaard in ontwikkeling was, maakte dat nog een heel groot verschil.

De codes in het gegevensblok over de inzender spreken verder voor zich, alleen de code CONT verdient een afzonderlijke vermelding. Deze code kan n.l. niet zelfstandig bestaan. Hij en zijn broertje CONC worden gebruikt om stukken tekst in de GEDCOM op te nemen. Soms alleen vanwege de opmaak, soms ook vanwege de lengte van de gegevens (een regel in GEDCOM mag maximaal 255 tekens lang zijn). Is de tekst langer dan moet die tekst over meerdere regels worden verdeeld. De vervolgregels met de code CONT moeten bij het weergeven dan op een nieuwe regel worden geplaatst, die met CONC moeten direct (zonder extra spatie) achter de tekst van de vorige regel worden geplaatst (dus altijd afbreken midden in een woord)

Bovenstaande vlakke lay out is de gebruikelijke manier om GEDCOM weer te geven. Om de leesbaarheid te vergroten, gebruik ik echter liever het format hieronder, dat met een klein programmaatje vanuit het eerste kan worden gemaakt, en weer kan worden teruggezet. Dat laatste is overigens vaak niet nodig want bijna alle programma's slikken de lege regels en de spaties aan het begin van een regel gewoon in. En dat is ook conform de standaard.

Het desbetreffende programma kan worden opgehaald van <http://www.famrozendaal.net> menukeuze "GEDCOM introductie". En hieronder dus de iets leesbaarder weergave

```
0 HEAD
1 SOUR PAF
    2 NAME Personal Ancestral File
    2 VERS 5.2.18.0
    2 CORP The Church of Jesus Christ of Latter-day Saints

    3 ADDR 50 East North Temple Street
        4 CONT Salt Lake City, UT 84150
        4 CONT USA

1 DEST Other
1 DATE 6 Jul 2006

    2 TIME 11:44:33
1 FILE vb.ged
1 GEDC

    2 VERS 5.5

    2 FORM LINEAGE-LINKED
1 CHAR UNICODE
1 LANG English
1 SUBM @SUB1@

0 @SUB1@ SUBM
1 NAME Wim Rozendaal
1 ADDR straat 12

    2 CONT woonplaats
    2 CONT provincie
    2 CTRY Nederland

1 PHON 0123456789
1 EMAIL rozendaa@xs4all.nl
```

2.3 Gegevens van personen

Laten we eenvoudig beginnen: ik stel mij voor:

```
0 @I1@ INDI

1 NAME /Rozendaal/Willem Hendrik
    2 NICK Wim
    2 NOTE Vernoemd naar zijn overgrootvader

1 SEX M

1 BIRT
    2 DATE 11 MAY 1956
```

- 2 PLAC Ooltgensplaat
 - 3 NOTE Op de oostelijke rand van Goeree-Overflakkee
- 2 SOUR @S1@

- 1 OCCU Manager van diverse rekencentra
 - 2 DATE FROM 1989 TO 2001
 - 2 SOUR Eigen waarneming

- 1 RESN Niet alle privacygevoelige gegevens zijn weergegeven

- 1 EVEN
 - 2 TYPE Hobby
 - 2 NOTE Naast stamboomonderzoek portret en landschapsfotografie

- 1 FAMS @F1@
- 1 FAMC @F2@
 - 2 PEDI Birth
- 1 ASSO @I2@
 - 2 RELA Getuige bij haar huwelijk met ...
- 1 ASSO @I3@
 - 2 RELA Getuige bij zijn huwelijk met ...

Dit simpele voorbeeld laat veel van de rijkdom van de GEDCOM taal zien.

Van boven naar beneden:

Allereerst het label waarmee er naar mijn gegevens kan worden verwezen van elders in het bestand, met de indicatie dat het een INDIVIDUAL (persoon) gaat. Deze indicatie dient ertoe dat het programma dat de GEDCOM leest weet welke betekenis aan de volgende codes moet worden toegekend.

Dan volgt mijn naam (NAME), de achternaam tussen de / tekens en de voornamen dus niet. Met als sub-item mijn roepnaam (NICK), en een aantekening (NOTE) over waar mijn naam vandaan komt.

Het geslacht (SEX) is ook te bevatten, en dan komen zaken die GEDCOM lastig maken: de feiten en gebeurtenissen.

Er zijn 21 voorgedefinieerde feiten voor personen, zoals geboorte (BIRT), doop (BAPM of CHR), beroep (OCCU), en veel gebeurtenissen in de kerk van de Mormonen. Elk van deze feiten heeft zijn eigen structuur en regels voor het vastleggen. Zeker als je tot in details wilt gaan. En daarnaast bestaat er nog de mogelijkheid om 'eigen feiten' (EVEN) te maken, waarbij je het type (TYPE) kunt definiëren en details op de gebruikelijke wijze kunt invoegen

Maar er is wel het een en ander standaard voor alle feiten:

- 1) bij elk feit kan je een datum (DATE) opgeven. En dat kan een enkele dag zijn, of een periode waarbinnen (BETWEEN ... AND ...), of een periode gedurende de welke (FROM ... TO ...). Daarbij is het format DD MMM YYYY. (dag met voorloopnul, de Amerikaanse 3 letter afkorting voor de naam van de maand in hoofdletters en 4 cijfers voor het jaar. Delen van een datum die niet bekend zijn kunnen worden

weggelaten. En dan zijn er nog wat aanduidingen waarmee wordt aangegeven hoe zeker de datum is. (AFT (na), BEF (voor), ABT (ongeveer), EST (geschat), CAL (berekend))

Al met al leidt dat er toe dat volgende data valide zijn:

11 MAY 1956
BETWEEN CAL 1511 AND EST 3 APR 1712
FROM ABT OCT 1984 TO BEF 2 FEB 2002

- 2) Bij elk feit kan een plaats (PLAC) worden meegegeven. En dat kan dan weer met verschillende methodes worden gedaan (Alleen de plaatsnaam, plaatsnaam en provincie/land, geografische coördinaten, etc.)
- 3) Onder elke code (dus die voor het feit, maar ook die voor de datum, en die voor de plaats) kan een notitie of een bronvermelding worden geplaatst die betrekking heeft op het bovenliggende detail (dus een notitie bij een bron vermelding, met een bronvermelding voor die notitie is niet verboden). De tekst van de notitie of de bronvermelding kan direct worden opgenomen, of met een label, als verwijzing naar een speciale constructie. (Die dan wel ergens moet staan in het bestand.)
- 4) Er is geen GEDCOM code om het tijdstip van een gebeurtenis vast te leggen.

Tussen de feiten staat nog een code die veelal wordt weggelaten, maar die je zou kunnen gebruiken om aan te geven dat iemand niet alle gegevens heeft gekregen (RESN., restriction notice). Ik heb die code daar neergezet omdat het spontaan in me opkwam tijdens het intikken. En bij GEDCOM maakt de volgorde van de codes op hetzelfde level, onder dezelfde hogerliggende code niet uit. Dus mijn geboortegegevens had ik ook achteraan kunnen zetten als:

```
1 BIRT
  2 SOUR @S1@
  2 PLAC Ooltgensplaat
```

```
3 NOTE Op de oostelijke rand van Goeree-Overflakkee
2 DATE 11 MAY 1956
```

De 7 laatste regels zeggen iets over mijn relaties met anderen. FAMS (altijd gevolgd door een label!) verwijst naar een gezinsverband/ouderpaar waar van ik een van de partners ben. FAMC (ook altijd gevolgd door een label) naar een gezinsverband/ouderpaar waar ik als kind bij betrokken was. Hoe die betrokkenheid was geeft de code PEDI weer. De gegevens over de gezinsverbanden worden bij het label weergegeven en staan verderop beschreven

De ASSO code is een buitenbeentje. Hij moet worden gebruikt om alle andere relaties mee weer te geven, en zou dus een heel krachtig hulpmiddel kunnen zijn bij het vastleggen van de gegevens. Maar in de standaard staat hij nauwelijks genoemd en in de praktijk worden de gegevens veelal als notities vastgelegd. Een gemiste kans.

De werking van de ASSO is heel simpel. In het voorbeeld verwijst de eerste ASSO van mij naar het label @I2@, waarachter mijn zus schuilgaat. En de onderliggende code

RELA beschrijft vanuit mijn standpunt de relatie: ik (trad op als) getuige bij haar huwelijk met Vanzelfsprekend was ik ook getuige bij het huwelijk van mijn zwager.

Nog een opmerking over het stukje GEDCOM waarin ik mij voorstel: Hoewel het geheel aan de standaard voldoet, zijn de zeven programma's die het ter test heb aangeboden (Aldfaer, Easy Computing StamBoom, Stamboom de Luxe versie 4, FAMYX, Pro-GEN, GEDVIEW en Genealogica Grafica) er geen van allen in geslaagd de gegevens na inlezen weer te geven zoals ik ze bedoelde. Hoe dat komt wordt later besproken, voor dit ogenblik volstaat het om te zeggen dat dat niet betekent dat mijn code fout is, of de programma's niet goed zijn. Je kunt er alleen uit afleiden dat GEDCOM zo zijn beperkingen heeft.

2.4 Gegevens van relaties

Relaties is waar het in het leven om gaat. En ook bij stamboomonderzoek ben je altijd bezig met relaties. De GEDCOM-standaard kent twee soorten relaties: die waarin een ouders/kind(eren) verhouding zou kunnen bestaan, en alle overige. En die 'alle overige' worden in de taal beschreven met de hierboven genoemde ASSO code. Blijft voor dit hoofdstuk dus over de groep waarin er van meerdere generaties sprake is.

Wederom een stukje GEDCOM als voorbeeld:

```
0 @F1@ FAM
1 HUSB @I1@
1 WIFE @I8@
1 CHIL @I4@
1 CHIL @I5@
1 EVEN
2 DATE 6 DEC 1980
2 TYPE Ondertrouw
2 PLAC Purmerend
2 NOTE @FN1@

1 MARR
    2 TYPE Civil
    2 DATE 30 DEC 1980
    2 NOTE Speciaal voor dit huwelijk moest de Purmer
        3 CONC ender markt een half uur vroeger sluiten.
1 NOTE de eerste kennismaking was op 11 april 1980
1 NCHI 2

0 @FN1@ NOTE
1 CONT Er lag sneeuw die dag, een voor
    2 CONC teken voor het weer op de trouwdag zelf.
```

Zoals gebruikelijk begint ook dit blok code met een label en dan de aanduiding van het type gegevens (FAM). Zoals reeds eerder is aangegeven er is geen vaste volgorde waarin de codes moeten staan maar binnen een FAM moet er altijd wel tenminste één HUSB (manlijke partner) of WIFE (vrouwelijke partner) aanwezig zijn, en tevens maximaal één van elke. Het aantal CHIL (kind records) is onbeperkt. En er dienen tenminste

twee personen opgegeven te worden. Dit betekent dus dat het niet mogelijk is om twee broers aan elkaar te koppelen via een FAM-record, tenzij er tenminste 1 ouder bekend is of er een Dummy (Naam Onbekend) wordt ingevoerd.

Het zal de inmiddels duidelijk zijn dat de code HUSB middels het label @I1@ verwijst naar mijn gegevens zoals die in het vorige codefragment staan. En WIFE @I8@ naar het blok in het GEDCOM-bestand waar de gegevens van mijn vrouw worden vermeld, de CHIL codes verwijzen naar de gegevens van mijn kinderen.

Vervolgens komen er wat feiten uit de relatie. Waarbij zich al direct een probleem voordoet want ik wilde de ondertrouw opnemen, maar in de Verenigde Staten (waar de standaard zich hoofdzakelijk op richt) zijn de procedures rond een huwelijk net even anders dan bij ons en er is dus geen code voor onze ondertrouw. MARL (marriage license) wordt er wel voor gebruikt maar dat is toch niet helemaal hetzelfde. De GEDCOM-standaard biedt ook hier de mogelijkheid om (evenals bij personen) eigen types feiten te definiëren en dus heb ik dat gedaan met de EVEN constructie.

Er zijn 11 voorgedefinieerde feiten bij relaties. Waarvan MARR (Huwelijk) waarschijnlijk de meest gebruikte is. Deze code kan bij één relatie meerdere keren voorkomen met verschillen type aanduidingen. (Bijv Church en Civil, voor kerkelijk en burgerlijk huwelijk). In de standaard staat zelfs geen beperking om dezelfde TYPE aanduiding meerdere malen te gebruiken, zodat het in theorie mogelijk is met wat eigen feiten en herhaling van codes een relatie te beschrijven waarin er eerst wordt samengewoond, vervolgens voor de wet getrouwd en pas later voor de kerk. En er na een periode waarin er een echtscheiding (code DIV) was, wordt hertrouwd voor de wet.

Met het vermelden van de code DIV is gelijk de op een na meest gebruikte code genoemd. De overige feiten zal ik hier niet verder benoemen, omdat ze in de praktijk te weinig voorkomen.

In het codevoorbeeld staan voorbeelden van notities op verschillende levels, en die dus steeds alleen slaan op het eerst hogere level. De via een label gekoppelde aantekening over het weer heeft dus in theorie alleen betrekking op de ondertrouw.

De code NCHI staat voor het aantal kinderen. Hier lijkt hij overbodig, maar het expliciet toevoegen van het aantal kinderen als dat bekend is, betekent dat het voor iedereen duidelijk is of alle kinderen zijn vermeld.

De verwijzingen van de personen naar de relatie, en van de relatie naar de personen dienen altijd beiden aanwezig te zijn. Verwijzingen naar andere soorten labels, zijn altijd alleen aanwezig bij het item waar de gegevens achter het label moeten worden ingevoegd. Dus een persoon verwijst wel naar b.v. een aantekening, maar die aantekening verwijst niet terug naar de persoon

2.5 Bronnen

Afhankelijk van het doel van het stamboomonderzoek komt er vroeger of later een ogenblik dat je inziet dat het van belang is om alle bronnen van alle feiten goed vast te leggen. Als je dan twee stamboomonderzoekers bij elkaar plaatst dan heb je al snel kans dat er een verhitte discussie ontstaat over wat een bron is, en wat je van een bron moet opnemen in je genealogie. Ik ga hier niet op die discussie in, maar wil wel aange-

ven dat het bronnen systeem in GEDCOM een van de meest rijke onderdelen van de taal is.

Maar alvorens op de complexe brondefinitie in te gaan, eerst even de opmerking dat een simpele bron kan worden opgenomen als een simpele bronvermelding van een of meer regels. Dit geldt overigens voor veel componenten die als afzonderlijk blok opgenomen kunnen worden (naast de bronnen bij voorbeeld notities en adressen)
Een korte vermelding luid:

1 BIRT

2 DATE 11 MAY 1956

2 PLAC Ooltgensplaat

3 SOUR trouwboekje ouders

Hier staat dus dat de bron waaruit ik mijn geboorteplaats ken het trouwboekje van mijn ouders is. De datum heeft mogelijk een andere bron.

Om de veelheid aan mogelijkheden te laten zien volgt hieronder geen code maar een stukje uit de standaard, en wel dat waarin de structuur van een bronvermelding wordt beschreven. Alle structuren worden op deze wijze in de standaard beschreven, en de uitleg hier kan dus helpen als u zelf verder in de standaard gaat zoeken.

SOURCE_RECORD: =

n @<XREF:SOUR>@ SOUR {1:1}

+1 DATA {0:1}

+2 EVEN <EVENTS_RECORDED> {0:M}

+3 DATE <DATE_PERIOD> {0:1}

+3 PLAC <SOURCE_JURISDICTION_PLACE> {0:1}

+2 AGNC <RESPONSIBLE_AGENCY> {0:1}

+2 <<NOTE_STRUCTURE>> {0:M}

+1 AUTH <SOURCE_ORIGINATOR> {0:1}

+2 [CONT|CONC] <SOURCE_ORIGINATOR> {0:M}

+1 TITL <SOURCE_DESCRIPTIVE_TITLE> {0:1}

+2 [CONT|CONC] <SOURCE_DESCRIPTIVE_TITLE> {0:M}

+1 ABBR <SOURCE_FILED_BY_ENTRY> {0:1}

+1 PUBL <SOURCE_PUBLICATION_FACTS> {0:1}

+2 [CONT|CONC] <SOURCE_PUBLICATION_FACTS> {0:M}

+1 TEXT <TEXT_FROM_SOURCE> {0:1}

+2 [CONT|CONC] <TEXT_FROM_SOURCE> {0:M}

+1 <<SOURCE_REPOSITORY_CITATION>> {0:1}

+1 <<MULTIMEDIA_LINK>> {0:M}

+1 <<NOTE_STRUCTURE>> {0:M}

+1 REFN <USER_REFERENCE_NUMBER> {0:M}

+2 TYPE <USER_REFERENCE_TYPE> {0:1}
+1 RIN <AUTOMATED_RECORD_ID> {0:1}
+1 <<CHANGE_DATE>> {0:1}

Van boven naar beneden staat hier respectievelijk:

- een bron begint altijd met een label (weergegeven door @<XREF:SOUR>@), dat op elk level kan zitten (aangegeven door de letter 'n' aan het begin van de eerste regel) en waarvan er tenminste één en maximaal één moet zijn per bron ({1:1}). De typeaanduiding achter het label is altijd SOUR. Hoewel volgens de standaard een bronvermelding op elk niveau kan beginnen, vindt u in de praktijk alleen het cijfer 0
- een bronvermelding kan een opsomming bevatten van de feiten die in de bron genoemd worden. Deze kan eenmaal voorkomen, en bevat dan 0 tot elk gewenst aantal gebeurtenissen. Die gebeurtenissen worden elk weer met een EVEN structuur beschreven. Verder kan hierbij worden opgenomen welke instelling (AGNC) verantwoordelijk is voor de bron en kan een onbepaald aantal notities worden opgenomen. En elke notitie kan een structuur op zichzelf zijn, met een bronvermelding daarin. Dit alles onder de code DATA, die een level hoger (+1) zit dan het label
- bij een bronvermelding kan worden opgenomen wie de auteur van de bron is (AUTH op 1 level hoger dan het label) en daarbij kunnen weer meerdere regels worden gebruikt om de gegevens goed weer te geven. Er kunnen nul of meerdere auteurs worden genoemd.
- de bron kan een titel hebben, maximaal één. Dit is de titel waaronder de bron gepubliceerd is.
- als de bron in eigen bezit is hier worden opgegeven hoe hij is opgeborgen met behulp van de ABBR code, die ook maar één keer mag voorkomen

Het principe zal nu duidelijk zijn en een verdere opsomming zal ik dan ook niet geven, maar voor details verwijst ik naar de standaard. Indien u dit artikel voor u heeft als computerdocument, kunt u op de verwijzingen klikken om direct bij de juiste hoofdstukken in de standaard te komen. Heeft u het artikel niet als computerdocument dan kunt u het, met het programma voor het formatteren van GEDCOM ophalen van <http://www.famrozendaal.net>, onder GEDCOM introductie.

En bovenstaande geeft u dus zeer veel mogelijkheden om zelf te bepalen wat u van een bron wilt opslaan, en hoe u dat wilt doen. Om dit geheel extra krachtig te maken kunt u nog van het volgende gebruikmaken: bij het verwijzen naar een bron kunt u op de volgende regels extra details toevoegen die alleen van belang zijn op dat punt van verwijzen. Een praktische toepassing daarvan vindt u in het programma GEDCOMpare (zie verderop), waar bij het verwerken van gegevens uit GENLIAS bij elke persoon die in een bron wordt genoemd de verwijzing naar de bron wordt gevolgd door een beschrijving van de rol die die persoon bij de gebeurtenis die in de bron beschreven wordt heeft vervuld. Op die manier hoeven de meeste gegevens van de bron slechts eenmalig vermeld te worden en kan toch specifiek per persoon worden aangegeven wat er aan de hand is. In GEDCOM ziet het er als volgt uit:

0 @I5@ INDI

1 SOUR @S5@

2 ROLE Moeder van de boreling

1 SEX F

1 NAME Anna Barbara/Gorens/

1 BIRT

2 DATE BETWEEN 25 NOV 1800 AND 23 MAY 1802

1 FAMS @F2@

0 @S5@ SOUR

1 NOTE Archief: Drents Archief

2 CONT Burgerlijke stand - Geboorte

2 CONT Toegangsnummer: 0165.033;

2 CONT Inventarisnummer: 1836; Gemeente: Zuidwolde;

2 CONC Soort akte: geboorte; Aktenummer: 53; Aangifdatedatum: 25-11-1836

2.6 Betrouwbaarheid

Als u gegevens verwerkt is het natuurlijk van belang dat u enig idee heeft over de betrouwbaarheid van die gegevens. In GEDCOM kan dat op twee manieren worden aangegeven. Ten eerste impliciet door de volgorde van de gegevens (dat lijkt inderdaad in tegenspraak met de opmerking eerder dat de volgorde er niet toe doet) en expliciet door de betrouwbaarheid van een bron aan te geven.

Om met dat laatste te beginnen: als u een bron heeft beschreven op de wijze zoals die hierboven is weergegeven, moet u op de punten waar u de bron wilt refereren een regel van het type

n SOUR @Snummer@

opnemen. U heeft reeds gezien dat daaronder nadere specificatie van de reden van vermelding kan worden geplaatst (b.v. met ROLE). In deze situatie (en volgens de standaard nergens anders) kunt u de QUAY code gebruiken om aan te geven hoe betrouwbaar u als opsteller van het GEDCOM-bestand de bron van de gegevens acht voor deze vermelding. De waardes lopen daarbij uiteen van 0 tot 3 met (vrij vertaald) de volgende betekenissen:

0 = Onbetrouwbaar bewijsmateriaal of geschatte gegevens

1 = Roddel, achterklap, autobiografieën en andere bronnen die geneigd zijn te "verlevendigen". Twijfelachtige betrouwbaarheid van het bewijs.

2 = Indirect bewijs of officiële vastlegging geruime tijd na de gebeurtenissen

3 = Direct, primair bewijs, of een overdaad aan indirect bewijs.

Het gaat hierbij dus om uw inschatting van de betrouwbaarheid van het bewijsmateriaal. Dit ontheft de ontvanger van de gegevens niet van de plicht om zelf na te gaan of hij het met uw inschatting eens is.

NB: er zijn programma's die deze codes in de omgekeerde volgorde gebruiken (Aldfaer b.v. geeft bij code 3 de betekenis "onbetrouwbaar", maar daar moet bij worden gezegd dat de tag die wordt gebruikt _QUAY is.)

Het andere systeem werkt als volgt: stel u komt twee namen tegen waarvan u aanneemt dat ze op dezelfde persoon betrekking hebben. U kunt dan beide namen opgeven voor deze persoon, waarbij de volgorde zo moet zijn dat degene die u het meest betrouwbaar acht of bij voorkeur gebruikt als eerste wordt genoemd. De volgende naam die u vermeldt heeft minder voorkeur en een eventuele derde weer minder dan de tweede. In dit geval gaat het dus om een oordeel over de juistheid van de gegevens en niet over de juistheid van de bronnen.

En om de verwarring nog groter te maken: u kunt dus de meest betrouwbare naam vooraan plaatsen en daarbij aangeven dat de betrouwbaarheid van de bron minder is dan die van de volgende naam. Terwijl het feit dat het de volgende naam is die aangeeft dat u de naam op zich minder betrouwbaar acht. Een ander niet geheel denkbeeldig feit: u weet van twee mensen dat ze getrouwd zijn geweest want u heeft van beiden de overlijdensakte gevonden. En beide aktes verwijzen naar de andere partner (de een als echtgenoot of echtgenote en ander als weduwe of weduwnaar van). Maar u heeft geen huwelijksakte. In dat geval kunt u dus twee bronnen opgeven voor het bestaan van het huwelijk, maar als u dat doet is er volgens de standaard impliciet een minder betrouwbaar dan de ander. En dat hoeft in de praktijk absoluut niet zo te zijn.

Concluderend: er zijn mogelijkheden om aan te geven hoe betrouwbaar bronnen zijn, maar deze hebben interne tegenstrijdigheden, reden waarom velen ervoor kiezen het oordeel over de betrouwbaarheid als notitie bij de bron te nemen, en dat over een feit bij de gegevens zelf in plaats van bij de bron. En hiervoor de QUAY code misbruiken.

2.7 Het grote geheel

Wat u tot nu toe heeft gelezen betreft slechts een beperkte inleiding in de GEDCOM. De taal kent meer dan 200 vaste codes en een groot aantal per programma verzonden codes. Want dat laatste is volgens de standaard toegestaan: elk programma mag een eigen code, of meerdere eigen codes, invoeren zolang mijn naam maar begint met een teken _.

Andere programma's worden geacht deze codes niet te verwerken, of de verwerking van het bron-programma over te nemen.

Onderdelen die niet genoemd zijn, maar wel uitgebreid worden beschreven in de standaard zijn:

- Multimedia. Het is mogelijk om in een GEDCOM bestand te verwijzen naar multimedia bestanden, die ofwel op internet staan, ofwel los worden meegestuurd, ofwel als onderdeel van het GEDCOM bestand worden meegestuurd.
- Repositories. Dit is een verzamelbegrip voor plaatsen waar bronnen zijn opgeslagen: archieven etc.
- Identificatiecodes voor personen, relaties, etc., en wijzigingsindicaties. Gegevens veranderen tussen verschillende versies van een bestand. Vaak is het wenselijk om niet alle gegevens opnieuw in te lezen, maar alleen de gewijzigde. En dan moet er dus duidelijk zijn wanneer een item: voor het laatst gewijzigd is (daarvoor wordt de CHAN code, met sub-codes gebruikt),
- nieuw is toegevoegd (_NEW, de enige code in de standaard zelf met een underscore begint),
- en over welk item we het hebben.

En dat laatste wordt in de standaard anders beschreven dan door veel programma's in de praktijk wordt gedaan. Het GEDCOM label mag wat de standaard aangaat veranderen tussen verschillende versies van dezelfde gegevens en is dus niet betrouwbaar als identificatie. Toch vertrouwen veel programma's hier wel op.

Al eerder heeft u kunnen lezen dat gegevens tot 99 niveaus diep genest kunnen zijn in een GEDCOM-bestand. In de praktijk komt dat niet voor, maar is 5 of 6 het maximum dat gebruikt wordt. Toch geeft dit niveau van diepgang al een complexiteit die erg moeilijk is te verwerken. De meeste programmamakers beperken zich dan ook tot een subset van de standaard, en proberen het aantal niveaus van de codes tot het minimum te beperken.

3. De problemen

3.1 Hoe standaard is "volgens de standaard"?

De GEDCOM standaard dient om het uitwisselen van gegevens mogelijk te maken. En wel in de richting van de centrale computer van de Mormonen. Dat daarnaast velen hem zijn gaan gebruiken voor eigen doeleinden is prettig voor ons als stamboomonderzoekers, maar heeft het probleem dat de gegevens van de verschillende programma's niet op elkaar aansluiten opeens wel erg zichtbaar gemaakt. Want wat zie je nu: het ene programma kan zelden een GEDCOM van het andere programma foutloos verwerken. Toch beweren alle programma's, meestal geheel terecht, dat zij zich aan de standaard houden. Hoe kan dit?

Om deze vraag te beantwoorden gaan we even een stapje terug. We praten over het uitwisselen van gegevens. En we nemen voor het gemak aan dat elk programma alle gegevens verstaat en begrijpt. Mijns inziens is dat waar het probleem met GEDCOM gegevensuitwisseling ontstaat. Niet elk programma legt immers dezelfde gegevens vast, waardoor niet alle programma's dezelfde gegevens ter beschikking kunnen stellen, en waardoor niet alle programma's behoefte hebben om alle gegevens te begrijpen. Elk programma gebruikt dat deel van de standaard dat noodzakelijk is om de gegevens die in het programma worden vastgelegd te kunnen exporteren en importeren. En soms ook bij de import nog wat extra codes, waardoor bijvoorbeeld de gegevens van andere veelgebruikte programma's ook gesnapt worden. Maar deze gegevens worden dan altijd in het programma omgezet in een vorm die voor het programma zelf wel acceptabel is. Bijvoorbeeld door feiten die het inlezende programma niet kent als notities (meestal zelfs met de GEDCOM-codes er nog bij) weer te geven.

Een heel concreet voorbeeld van niet aansluiten van programma's: HAZA-Data 7.2, een programma dat jarenlang door velen is gebruikt wordt nu door de gebruikers vaak afge-stoten. En dan probeert een deel van de gebruikers of bijvoorbeeld Aldfaer een goed programma zou zijn om als vervanger op te treden. De eerste test die dan plaatsvindt is het importeren van GEDCOM-bestanden in Aldfaer. En dat geeft negen van de 10 keer een grote teleurstelling. Haza (gebaseerd op standaard versie 5.3) gebruikt gewoon te veel eigen codes om door Aldfaer te worden begrepen. Omgekeerd geldt overigens hetzelfde.

Laten we de volgende denkbeeldige persoon eens nemen:

@I1@ INDI

1 NAME /Rozendaal/Willem Hendrik

2 NICK Wim
 1 BIRT
 2 NOTE @N1@
 2 SOUR @S1@
 2 DATE ABT 20 MAY 1926
 2 PLAC Dordt

 1 DEAT Y
 2 PLAC Gorcum
 2 DATE 12 AUG 2005
 1 _ALDFAER_TBS
 1 FAMS @F1@
 0 @N1@ NOTE
 1 CONT Terwijl zijn moeder op visite was in Dordrecht

 0 @S1@ SOUR
 1 CONT Familie-overlevering

Het moet na de vorige hoofdstukken voor iedereen te volgen zijn wat hier staat. Alleen de code _ALDFAER_TBS zal voor een aantal mensen onbegrijpelijk zijn. Dat is zo'n code die slechts door één programma wordt gebruikt, in dit geval Aldfaer, en die inhoudt dat er geen begrafenis is geweest maar dat het lichaam ter beschikking is gesteld van de wetenschap. Vergelijk dat eens met de persoon hier onder:

0 @I1@ INDI

 1 NAME Willem Hendrik/Rozendaal/
 2 NICK Wim

 1 BIRT
 2 NOTE Volgens de familie-overlevering is hij geboren toen z
 3 CONC ijn moeder op visite was in Dordrecht
 2 DATE EST 22 MAY 1926
 2 PLAC Dordrecht

 1 DEAT Y
 2 PLAC Gorinchem
 2 DATE 12 AUG 2005

 1 BUR I N
 2 NOTE Het lichaam is aan de wetenschap geschonken

 1 FAMS @F1@

Voor degenen die niet bekend zijn in de omgeving van Dordrecht en Gorinchem: de namen voor beide plaatsen kennen een korte vorm en een lange vorm. Het gaat dus om dezelfde plaatsen.

Met die wetenschap is het te zien dat het om dezelfde persoon gaat, maar toch is meer dan de helft van de regels verschillend. Let ook op de details als de volgorde van de naam: voornaam en achternamen staan verschillend.

Dit voorbeeld maakt wel duidelijk dat de standaard weliswaar bepaalt hoe je dingen mag opschrijven, maar dat binnen die regels de menselijke geest van de vrijheid heeft om verschillende notaties voor dezelfde gegevens te gebruiken. Het is dus geen standaard manier om dingen op te schrijven, maar de GEDCOM-standaard geeft aanwijzingen over de mogelijke wijze waarop de gegevens voor andere computers leesbaar kunnen worden weergegeven.

In dit voorbeeld staan de gegevens niet ver van elkaar en was u gewaarschuwd dat er iets aan zat te komen, dus zal de overeenkomst u zijn opgevallen. Maar voor een computer staat hier echt tweemaal iets verschillends.

Waarmee maar gezegd is, de standaard helpt wel bij het uitwisselen van gegevens, maar er is meer nodig dan alleen een standaard voor de te gebruiken constructies om een taal te maken die door anderen begrepen wordt.

3.2 Hoe beheersbaar zijn uw gegevens?

Kijkt u nog eens naar de eerste drie code-segmenten bij dit artikel. We hebben het dan over 80 regels code. En het enige dat u dan weet is dat ik op 11 mei 1956 ben geboren, 30 december 1980 ben getrouwd en twee kinderen heb. En daarbij is er nog een aantal ongedefinieerde labels, waardoor de code nog gaat groeien bij het volledig maken

GEDCOM-bestanden worden al snel erg groot. En dus hoeft niemand zich te verbazen als hij/zij een GEDCOM-bestand van 50 Megabytes of zo'n 3 miljoen regels krijgt opgestuurd. Daar zitten dan meestal 100.000+ personen in.

En u was zelf ook ijverig bezig geweest en had al zo'n 30,000 namen in uw bestanden.

Hoe vind je nu uit welke namen er dubbel in het systeem komen als de bestanden worden samengevoegd? Of (hypothetisch bij uw bestand, maar niet bij dat van een ander) wie er al dubbel in één van de bestanden zit.

En hoe gaan we na wat de kwaliteit van de gegevens is? Of het wel allemaal personen zijn die voor u van belang zijn, of er lussen in de gegevens zitten, of het bestand technisch correcte GEDCOM bevat, of de gegevens logische samenhang vertonen?

Dat kan met dergelijk grote hoeveelheden gegevens dus alleen met computerprogramma's. Maar dan gaat het zo goed als de programmeur die het programma heeft gemaakt is. En dat is over het algemeen ook maar een mens, die fouten maakt. Anders gezegd: GEDCOM biedt de mogelijkheid om snel veel gegevens uit te wisselen, maar introduceert daarmee ook de mogelijkheid dat uw zorgvuldig opgebouwde gegevensverzameling, door het importeren van vervuiling aangeleverd door anderen, waardeloos en onbruikbaar wordt.

Mijn schatting is dat van de gegevens die enkel als GEDCOM worden verzonden, een groot deel klakkeloos wordt overgenomen, waarbij het ene programma de betrouwbaarheidsaanduidingen van het andere programma niet snapt en die dus maar overboord gooit. En dan worden gegevens vaak doorgespeeld van de een naar de ander, met als gevolg dat een zorgvuldig opgebouwd bestand met een deel van mijn stamboom met veel commentaar en waarschuwingen in het origineel, zonder enige notities, kwaliteitsaanduidingen etc in Oost-Azië op een website stond, binnen enkele dagen nadat ik het

een ver familielid had toegestuurd (overigens was de tikfout die ik bij een naam had gemaakt bewaard gebleven).

Is dat erg, dat mijn gegevens daar staan? Ja en Nee.

Nee, het gaat over mensen die al eeuwen dood zijn, dus niemand heeft direct last van. Ja, want er staan zaken als hele waarheden gepresenteerd terwijl er in mijn bestand bijstond "is het op zijn mist speculatief te noemen om aan te nemen dat om ging". En door die weglating wordt een mogelijkheid die ik voor mezelf als zodanig (een mogelijkheid en niet meer dan dat) had genoteerd als de heilige waarheid door anderen overgenomen. En maakt het gebruik van GEDCOM het mogelijk om snel veel fouten verspreiden.

3.3 Andere (technische) problemen

3.3.1 Codepages

En dan zou de techniek de techniek niet zijn geweest als die van iets simpels als lettertekens niet iets onmogelijk complex had weten te maken. Want wat is het geval? Computers onthouden niet de grafische weergave van de letters, maar een getal dat voor een letter staat. Simpel toch: Overall waar je het getal 65 tegenkomt druk je een hoofdletter A af. En zo ook voor alle ander letters en cijfers. Een kwestie van met elkaar afspreken welke code voor welke letter staat. En dat afspreken is dan ook gebeurd: per taal, per operating system, soms zelfs per lettertype.

Dat is geen enkel probleem als je je gegevens altijd van mensen krijgt die met dezelfde instellingen van hun computer werken als je zelf doet, maar leidt anders bijna altijd tot problemen bij tekens 'met een dode toetsaanslag', é, ä, ñ, ç.

Wat is hier exact aan de hand:

Om te beginnen doen we steeds minder zuinig bij het toekennen van ruimte voor de getallen die de letters weergeven. In de glorie-dagen van de telex waren dat 7 bits per teken. Daarmee waren er 128 tekens mogelijk en dat was voldoende om de wereld te runnen. 26 Hoofdletters, 26 kleine letters, 10 cijfers, 30 stuurcodes (nieuwe regel, BELL etc) lieten zelf nog wat ruimte voor speciale tekens (die ook weer niet helemaal vast-lagen).

Met de komst van computers was het opeens niet meer zo logisch om 7 bits te gebruiken voor het opslaan van tekens, want computers doen van oudsher alles in veelvoud van 8 Bits. Dus werd dat 8ste bit ook voor de codering gebruikt. Dat gaf de mogelijkheid om 128 extra tekens toe te voegen. Vaak is dat per taal gebeurd en moet u dus de vertaaltabel (Eng: codepage) van de juiste taal gebruiken om iets zinnigs te kunnen lezen

Vervolgens kwam de Amerikaanse overheid met ANSEL (American National Standard for Extended Latin Alphabet Coded Character Set for Bibliographic Use, ofwel een uitbreiding aan destijds bekende codes om de tekens uit het Spaans een plekje te geven), een coderingstechniek die soms twee bytes (16 bits) gebruikt voor een teken, soms één byte (8 bits). Deze volgens velen inmiddels achterhaalde standaard was het neusje van de zalm toen de eerste versies van de GEDCOM-standaard werden opgesteld. En dus werd die codering tot de standaard codering van GEDCOM-files verklaard. Alle van het Arabische alfabet afgeleide letters hebben een plek gevonden in deze code, maar Grieks schrift, Russisch schrift en dergelijke worden niet ondersteund.

Om deze tekortkomingen weg te nemen zijn nieuwe systemen bedacht waarbij altijd van 16-bits (2 bytes) gebruik wordt gemaakt. Al met al een overdaad aan systemen om letters te coderen, en al die systemen worden gebruikt ook.

Als het goed is staat in het begin van elk GEDCOM-bestand welke codering wordt gebruikt, maar in de praktijk gaat dit vaak fout met als gevolg dat uw bestand na inlezen allerlei vreemde namen met onleesbare tekens weergeeft. Op zich is dat voor u wel te herstellen, maar een computerprogramma zal er problemen mee hebben. Want hoewel er tweemaal Daniël bedoeld wordt staat het er eenmaal niet en andermaal wel. En dus herkent de computer het niet als dezelfde naam.

3.3.2 Bestandsformaten

Een computerbestand is vaak meer dan het oog ziet. Er zitten soms b.v. een aantal tekens in de start van het bestand die worden verzwegen door Windows, maar waar het wel uit kan lezen dat er b.v. Big-endian 8 codering is toegepast. Handig voor Windows, maar lastig voor ons want wij willen als we het bestand openen als allereerste lezen '0 HEAD'. En jawel: sommige programma's zijn zo assertief dat ze niet van de Windows routines gebruik maken (b.v. omdat ze op Linux lopen) en dus de voorloopbytes gewoon gepresenteerd krijgen. Met als gevolg een foutmelding, want het bestand begint niet met de juiste eerste regel.

Een ander principieel verschil dat vaak tot ellende leidt is de gewoonte van Windows om een Carriage Return en een Line Feed aan het einde van een regel te zetten (stelt u zich even voor dat alles met een automatische schrijfmachine wordt uitgetikt, en u begrijpt wat deze codes doen). Linux, Unix en andere systemen gaan er van uit dat er altijd aan het begin van de volgende regel moet worden begonnen en geven de Carriage Return niet. En ook dat kan als een bestand van het ene naar het andere operating system wordt overgezet tot problemen leiden. En vergist u zich niet, want u werkt waarschijnlijk vaker met UNIX / LINUX dan u weet: bijna alle websites draaien op een van deze operating systemen.

4. De "oplossingen"

Als u tot hier bent gekomen en alle problemen heeft gelezen dan zult u zich misschien afvragen waarom er überhaupt nog met GEDCOM wordt gewerkt: doodeenvoudig, er is niets dat minder slecht is. En we slagen er nog altijd in om veel gegevens uit te wisselen. En met verstand en het opvolgen van een paar regels is het heel wel mogelijk om toch plezier te hebben van het gebruik van GEDCOM voor het uitwisselen van gegevens.

Hieronder volgen die regels:

- 1) Bewaar alle gegevens die u als GEDCOM heeft ontvangen in dat formaat. Dan is de originele set gegevens altijd bij de hand als er moet worden overgestapt naar een ander systeem, en hoeft er niet tweemaal een vermindering op te treden doordat GEDCOM niet vlekkeloos wordt verwerkt. Bovendien kunt u dan altijd terugzoeken wat de originele gegevens waren.
- 2) Vraag altijd om GEDCOM in het pure 5.5 standaard formaat. Veel programma's hebben speciale opties om bij het aanmaken van de GEDCOM alleen de tags te gebruiken die in de standaard staan. Dat levert soms iets gegevensverlies op, maar meestal staat dan gewoon alles in een iets andere notatie in

het GEDCOM-bestand. PAF, Legacy, FTW, GensDataPro, Aldfaer en een groot aantal andere programma's maken standaard GEDCOM (of iets dat daar op lijkt) aan met veel eigen trucs, en een ander programma kan dat over het algemeen niet lezen. Maar allemaal hebben ze optie om zuivere 5.5 GEDCOM op te leveren.

- 3) Vraag indien mogelijk altijd om GEDCOM in een 'neutrale' codering. In het vorige hoofdstuk heeft u kunnen lezen over de verschillende coderingen en codepages die voor een meer dan Babylonische spraakverwarring kunnen zorgen. De meeste programma's kunnen tegenwoordig zuivere Unicode encoding verwerken, en deze encoding (die altijd 16 bits voor een letter gebruikt) is in principe taal onafhankelijk. Als zuivere Unicode niet mogelijk is (b.v. omdat de gegevens uit een wat ouder (Dos?) programma komen, en het gaat alleen om westerse schrifttekens, vraag dan als iemand u een GEDCOM-bestand wil opsturen om het in ANSEL aan te maken. Dat is in principe ook 'taalneutraal', en bij mijn weten ondersteunt bijna elke programma dat GEDCOMs kan inlezen deze manier van coderen. En mocht u toch niet de gewenste taal neutrale encoding ontvangen, dan kunt u met het programma Textpad (van Helios software, officieel moet hier voor betaald worden, maar u kunt de volledige versie als niet commerciële gebruiker oneindig lang 'evalueren') het aangeleverde bestand inlezen (Textpad herkent bijna altijd de juiste invoer codering) en direct via 'opslaan als' in Unicode opslaan. U heeft bij dit programma n.l. de mogelijkheid om zelf op te geven welke encoding bij het wegschrijven moet worden gebruikt.
- 4) Vraag u af wat uw primaire dataopslag is. Tenzij u het ideale programma heeft gevonden dat alle overzichten maakt zoals u ze hebben wilt, zult u regelmatig met uw gegevens naar andere programma's gaan. Slept u uw gegevens op de manier steeds van het ene programma naar het andere dan is de kans groot dat u op het einde uw gegevens niet meer herkent. Het is dus van het grootste belang dat u één centrale plaats heeft waar u uw gegevens bewaart en aanpast. En dan steeds van daaruit naar een ander programma brengt om een speciaal rapport te maken of b.v. uw website te vullen. Maar zet de gegevens niet terug van het tweede programma naar het eerste.
- 5) Pas een goed privacy filter toe voordat u uw gegevens uit handen geeft. Veel programma exporteren alle gegevens zonder rekening te houden met of een persoon nog leeft, of er notities staan die u alleen voor eigen gebruik heeft, etc. Gebruik dan een programma dat de mogelijkheid biedt of achteraf alsnog een privacyfilter toe te passen. Doet u dat niet dan krijgt u in het beste geval een boze achterneef aan de telefoon en bent u in het slechtste geval verdachte van het overtreden van de privacy wetgeving.
- 6) Houdt er rekening mee dat GEDCOM bestanden overlappings hebben. Voeg geen nieuwe gegevens via GEDCOM toe aan uw huidige gegevens zonder eerst te kijken of er overlappings zijn tussen de twee verzamelingen. De meeste programma's voelen zich er goed bij om drie personen te hebben die Wim Rozen-daal heten en op 11 mei 1956 in Ooltgensplaat (een klein dorpje!) zijn geboren.

5. Op GEDCOM gebaseerde programma' s

In dit hoofdstuk treft u een lijstje van programma's aan. Onderverdeeld naar de functie van het programma. Ik heb me daarbij niet aan een kwaliteitsoordeel gewaagd. Immers

wat ik prettig vind, hoeft u niet prettig en te vinden. En een functie die ik nuttig vind, zult u misschien nooit gebruiken.

De filters die ik heb toegepast zijn:

1. De taal. Als er in een bepaalde categorie een Nederlandstalig programma voorhanden is staat dat bovenaan in de lijst.
2. Windows interface
3. Gratis verkrijgbaar

5.1 Controleren en repareren

Controleren van een GEDCOM-bestand kan op twee manieren: Wat staat er (juistheid van de gegevens) en Hoe staat het er (de juiste vorm van het GEDCOM-bestand). Beide functies vindt u in:

- Kstableau (<http://www.home.zonnet.nl/kstableau/index.html>), en zijn opvolger Genealogica Grafica (<http://home.versatel.nl/genealogicagrafica/>). Deze programma's hebben naast het standaard opgestarte Engelstalige interface ook een Nederlandstalige versie en zij controleren tijdens het inlezen van de gegevens voor hun hoofdfunctie, het grafisch weergeven van de stamboom. Alleen een technische toets, maar dan gecombineerd met het repareren van het bestand treft u in
- GEDCOMpare (www.aldfaer.nl). Dit programma is volledig Nederlandstalig. Alleen een technische toets:
- GEDCHK (http://www.familysearch.org/Eng/Home/FAQ/faq_gedcom.asp?color=green), van de makers van de standaard. Engelstalig

5.2 Bekijken

Kstableau en Genealogica Grafica, hierboven al genoemd hebben veel verschillende mogelijkheden om de stamboom te bekijken.

GEDViewer (www.ebotsoft.nl) is een Nederlandstalig programma om een GEDCOM te bekijken.

Gedrelation (www.ebotsoft.nl). Dit programma vraagt in het Nederlands om de invoer en geeft dan in het Engels aan hoe twee personen in de GEDCOM met elkaar verwant zijn.

5.3 Splitsen en Privacy filter

GEDCOMpare. Dit reeds eerder genoemde programma heeft een scala aan mogelijkheden voor het toepassen van filters en het maken van splitsingen.

Gedsplit (<http://www.rootsweb.com/~gumby/ged.html>), maakt een splitsing in een GEDCOM die u zelf kunt opgeven tot op de persoon. Engelstalig.

Gedclean (<http://www.raynorshyn.com/gedclean>), verwijdert alle nog levende personen uit het GEDCOM-bestand. Pas wel op want men wil graag dat u de betaalde versie download.

5.4 Website bouwen

5.4.1 Dynamische websites

Een dynamische website heeft geen vaste pagina's staan, maar elke pagina die u wilt zien wordt op dat moment opgebouwd en toont dus altijd de laatste gegevens die u aan het systeem heeft opgegeven. Er zijn veel pakketten in de handel die aan de hand van een GEDCOM-bestand een dynamische website bouwen, maar ik ken slechts een pakket dat deze functie voor niets levert:

PHPGedView (www.sourceforge.net) Met dit gratis pakket dat elke westerse taal spreekt, haalt u een zeer complete website binnen zonder dat u zelf hoeft te programmeren. Wel is een Internet hostingprovider met PHP en MYSQL vereist.

5.4.2 Statische websites

Een statische website bestaat uit vastgelegde pagina's. Bij elke wijziging van de gegevens is het nodig de hele website opnieuw te genereren, terwijl bij dynamische websites het voldoende is om de gegevens opnieuw op de website te zetten. De twee programma's hieronder volgen een totaal verschillende benadering om de statische websites te genereren.

DFTCOM (<http://www.dftcom2.co.uk/>) Met dit programma bouwt u een enkele eenvoudige grafische module met een interface in meerdere talen. Met deze module kunt u snel en eenvoudig door de gegevens bladeren, maar als uw stamboom wijzigt moet u de module helemaal opnieuw maken.

HuMogen (www.huubmons.nl/software) Dit programma genereert een verzameling HTML-bestanden, die bij elkaar een complete website vormen

5.5 Afdrukken

Voor het afdrukken van gegevens in een GEDOM-bestand bestaan er legio mogelijkheden. Ten eerste kunt u gebruik maken van de mogelijkheden van het programma of de programma's waarin u uw gegevens bijhoudt. Al deze programma's bieden tenminste een tekstuitvoermogelijkheid. En vaak bieden ze dan ook nog fraaie grafische mogelijkheden. Mocht u extra afdrukmogelijkheden zoeken, dan bieden de programma's die hierboven zijn genoemd vaak opties.

5.6 Data-mining

Data-mining, het zoeken naar veel gegevens en daar met behulp van programma's uit halen wat er van je gading is, is een nog redelijk onontgonnen gebied. Vermoedelijk is dit deels veroorzaakt door het feit dat je pas wat kunt doen met veel data als je veel data hebt. En pas de laatste jaren komt er meer data beschikbaar. De tweede reden waarom het geen grote vlucht heeft genomen is m.i. de haast aan het hele wezen van het genealogische onderzoek strijdige karakter van data-mining. In plaats van de archieven af te speuren naar die ene akte zet je een computer een nacht aan het werk en de volgende dag heb je je resultaat (als het meezit).

De grootste bronnen voor GEDCOM-data in Nederland zijn de digitale stamboom-monitor en de GenliasMonitor (beide te bereiken via www.genliasmonitor.nl). Deze diensten leveren geselecteerde gegevens op als PDF, Spreadsheet, HTML en GED-COM.

Maar evenals bij de data van een van de grootste bronnen ter wereld (www.family search.com, de site van de Mormoonse kerk) worden de gegevens als kleine samenhangende brokjes afgeleverd.

Meestal krijg je een ouderpaar met een kind (bij een huwelijk gecombineerd met een ander ouderpaar met hun kind). En hetzelfde ouderpaar kan (zelfs met hetzelfde kind) meerdere keren in de gegevens zitten

De twee genoemde diensten harken de gegevens uit de Genlias database en uit een aantal andere bronnen met uw vragen als invoer en sturen regelmatig de wijzigingen aan u toe.

Het hierboven al genoemde programma GEDCOMpare heeft een functie om de gegevens zoals ze worden gedistribueerd door o.a. de Genlias Monitor verder te interpreteren en in een veel gedetailleerder GEDCOM om te zetten dan er door de 'harksystemen' wordt opgestuurd. En vervolgens kan het de dubbelen in deze gegevens opzoeken, leden van één gezin bij elkaar plaatsen en dergelijke.

5.7 Overige bewerkingen

Overige bewerkingen met GEDCOM die in programma's zijn vastgelegd, zult u niet veel vinden. Een uitzondering is het programma GEDCOMpare, dat behalve de reeds genoemde functies ook de mogelijkheid geeft om:

1. GEDCOM het van het ene programma beter op het andere te laten aansluiten.
2. GEDCOM bestanden met elkaar te vergelijken
3. GEDCOM bestanden samen te voegen, met het wegwerken van doublures en
4. Geselecteerde gegevens op leveren als spreadsheet.

6. Meer informatie

De eigenlijke standaard is op vele plekken op het internet te vinden. Bijvoorbeeld: <http://homepages.rootsweb.com/~pmcbride/gedcom/55gctoc.htm>