

Sukututkimustiedon käsittely

Kaarle Kaila

Espoo 4.3.2002

Pro gradu tutkielma

HELSINGIN YLIOPISTO

Tietojenkäsittelytieteen laitos

Sisällys

| | | |
|-----|--|----|
| 1 | Johdanto | 4 |
| 2 | Sukututkimustieto | 5 |
| 2.1 | Jälkipolviraportti | 5 |
| 2.2 | Esipolviraportti | 11 |
| 3 | Sukututkimustiedon mallintaminen | 14 |
| 3.1 | GEDCOM 5.5 | 14 |
| 3.2 | THE GEDCOM STANDARD (Future Direction) | 16 |
| 4 | Sukututkimusohjelmat | 20 |
| 4.1 | Sukujutut | 21 |
| 4.2 | Juuret | 23 |
| 4.3 | LifeLines | 25 |
| 4.4 | Brother's Keeper | 26 |
| 4.5 | Genus | 28 |
| 4.6 | Sukuohjelmisto | 29 |
| 4.7 | Personal Ancestral File | 32 |
| 5 | GEDCOM-standardi | 33 |
| 5.1 | GEDCOM-tietoformaatti | 34 |
| 5.2 | GEDCOMin tietuerakenteet | 36 |
| 5.3 | GEDCOM-standardin alirakenteet | 39 |
| 5.4 | GEDCOM-standardin puutteita | 41 |
| 5.5 | Käytetyt merkistöt | 42 |
| 6 | XML:n käyttö sukututkimustiedon tallennus- ja siirtoformaattina..... | 42 |
| 6.1 | GedML | 43 |
| 6.2 | Sukuohjelmisto 2000:n varmuuskopio XML-muodossa | 44 |
| 6.3 | GEDCOM/XML | 46 |
| 6.4 | XSL-tekniikan hyödyntäminen | 46 |
| 7 | Sukututkimuspäätelyn mallintaminen..... | 48 |
| 7.1 | Tietomallin ”luonteenpiirteet” | 48 |
| 7.2 | Mallin suhde GEDCOM-standardiin | 50 |
| 7.3 | Sukututkimusprosessin eteneminen | 50 |
| 7.4 | Sukututkimustiedon yhtenäisteoria | 52 |
| 7.5 | Tietomallille asetetut vaatimukset | 55 |
| 7.6 | GENTECHin sukututkimuksen tietomalli | 56 |
| 7.7 | DeadEnds –tietomalli | 67 |
| 7.8 | Tekstitiedon esittäminen | 70 |
| 8 | Yhteenveto | 71 |

Esimerkit, taulukot ja kuvat

| | |
|--|----|
| Esimerkki 1. Jälkipolviesitys Bergholmin sukukirjassa | 6 |
| Esimerkki 2. Jälkipolviesitys Kankarin Kailat nro 6 sukukirjassa | 7 |
| Esimerkki 3a. Register-jälkipolviesitys osa 1 | 8 |
| Esimerkki 3b. Register-jälkipolviesitys osa 2 | 8 |
| Esimerkki 4. Jälkipolviraportti Genealogia Sursilianan mukaan | 10 |
| Esimerkki 5. Esipolviraportti numeroitu Stradonizin mukaan | 12 |
| Esimerkki 6. Esivanhemmat esitettynä sukukunnittain | 13 |
| Esimerkki 7. Yksilötietue | 38 |
| | |
| Taulukko 1. Kyselyn mukaan käytetyt sukututkimusohjelmat | 21 |
| Taulukko 2. GEDCOM-standardin tietuetyypit | 37 |
| Taulukko 3. Yksilötietueen rakenne | 37 |
| Taulukko 4. Perhetietue | 39 |
| | |
| Kuva 1. GEDCOM 5.5 yksinkertaistettu malli | 15 |
| Kuva 2. Perherakenne | 16 |
| Kuva 3. Henkilörakenne | 16 |
| Kuva 4. Sukututkimustiedot | 17 |
| Kuva 5. Bibliografiset tiedot | 18 |
| Kuva 6. Sukujutut-ohjelman keskusikkuna | 22 |
| Kuva 7. Juuret-ohjelman henkilöikkuna | 24 |
| Kuva 8. LifeLines ohjelman Windows version henkilönäyttö | 25 |
| Kuva 9. Brother's Keeper 6:n Muokkaa-näyttö | 27 |
| Kuva 10. Genus-Senior ohjelman päänäyttö | 29 |
| Kuva 11. Sukuohjelmisto 2000:n tietokannan tärkeimmät rakenteet | 30 |
| Kuva 12. Sukuohjelmisto 2000:n tietokanta- ja perheikkuna, henkilöikkunan sukulaiset | 31 |
| Kuva 13. Personal Ancestral File ohjelman perhenäyttö | 32 |
| Kuva 14. Sukutiedon esittäminen XML muodossa | 45 |
| Kuva 15. XSL-esimerkki | 47 |
| Kuva 16. XML-tiedosto esitetty XSL-tyylitiedostoa käyttäen | 48 |
| Kuva 17. Superlausetyypin yleisimmät yhdisteet | 54 |
| Kuva 18. Todisteet osamalli | 58 |
| Kuva 19. Johtopäätökset osamalli | 59 |
| Kuva 20. DeadEnds-tietomallin luokkakaavio | 68 |
| Kuva 21. DeadEnds tietomallin olioiden välisiä suhteita | 70 |

1 Johdanto

Sukututkimus eli genealogia on sukulaisuussuhteiden selvittämistä, henkilöhistoriallista tutkimusta. Tutkimusta varten sukututkija kerää tietoa eri lähteistä. Tutkimuksen tavoitteena on useimmiten erilaisten esipolvi-[Lut59] ja jälkipolviraporttien [Ber01] tuottaminen. Suomessa suosituin taulumuotoinen jälkipolviesitys on muotoutunut n. 100 vuotta vanhan saksalaisen esikuvan mukaan. Sukututkimusta voidaan tehdä vain itseään ja omaa perhettään varten, mutta usein tulokset pyritään julkaisemaan. Tutkimus voidaan julkaista esimerkiksi sukukirjan muodossa joko painettuna tai monisteena. Nykyään monet tutkijat julkaisevat tuloksiaan myös Internetin välityksellä.

Tutkimuksen yhteydessä syntyvä tieto voidaan jakaa henkilö-, perhe-, lähde- ja tekstitietoihin. Henkilötieto jakaantuu nimi- ja tapahtumatietoihin sekä muihin ominaisuuksiin kuten ammatti, koulutus ja muu tekstimuotoinen elämäkertatieto. Perhetieto koostuu perheenjäsenten luettelosta sekä heidän rooleistaan ja esiintymisjärjestyksestään, avioliiton solmimis- ja erotiedoista sekä muista, kuten esimerkiksi adoptiotiedoista. Lähdetieto koostuu viittauksista niihin lähteisiin, joista tieto on saatu. Tutkimuksen yhteydessä syntyvää muuta tietoa voidaan käsitellä tekstitietona. Myös multimediatietoa kuten vanhoja valokuvia käsitellään tutkimuksen yhteydessä.

Sukututkimustieto koostuu arkistoista ja muista lähteistä saaduista tiedoista. Sen käsitteleminen tietokoneella edellyttää tiedon jonkinasteista mallintamista.

Sukututkimusohjelmat mallintavat tiedon omia tarpeitaan silmälläpitäen. Ohjelmien tärkeimpien tietojen tulee olla muunnettavissa GEDCOM-standardin [Ged96] mukaiseen muotoon ja ohjelman käyttämän tietomallin on hyvä olla yhteensopiva GEDCOM-standardin esittämän mallin kanssa. GEDCOM-standardia vastaavaa XML-pohjaista esitysmuotoa on kokeiltu, ja on odotettavissa, että ohjelmat siirtyvät XML-muotoisten tiedostojen käyttöön lähivuosina.

Sukututkimustietoa käsitellään useimmiten sitä varten suunnitelluilla sukututkimusohjelmistoilla. Nykyisillä sukututkimusohjelmilla tallennetaan ja muokataan käsiteltävää aineistoa tiedon järjestelmällistä tallennusta sekä erilaisten tulosteiden tuottamista varten.

Suvun tutkiminen on pitkäaikainen prosessi. Se saattaa kestää jopa usean sukupolven ajan. Näin ollen on tärkeää, että tutkimuksen lähteet ovat löydettävissä jälkeenpäin. Sukututkimusohjelmistot mahdollistavat yleensä lähteiden tallentamisen tietojen yhteyteen. Mikäli tutkija on tunnollisesti tallentanut lähdetiedot järjestelmänsä katsotaan asian olevan hyvällä pohjalla. Luvun 7 käsittelemä sukututkimustiedon päättelyn mallintaminen osoittaa, että myös lähdetietojen pohjalta tehtävät johtopäätökset olisi syytä kirjata. On odotettavissa, että seuraava sukututkimusohjelmistojen sukupolvi tuo mukanaan tähän liittyviä piirteitä.

2 Sukututkimustieto

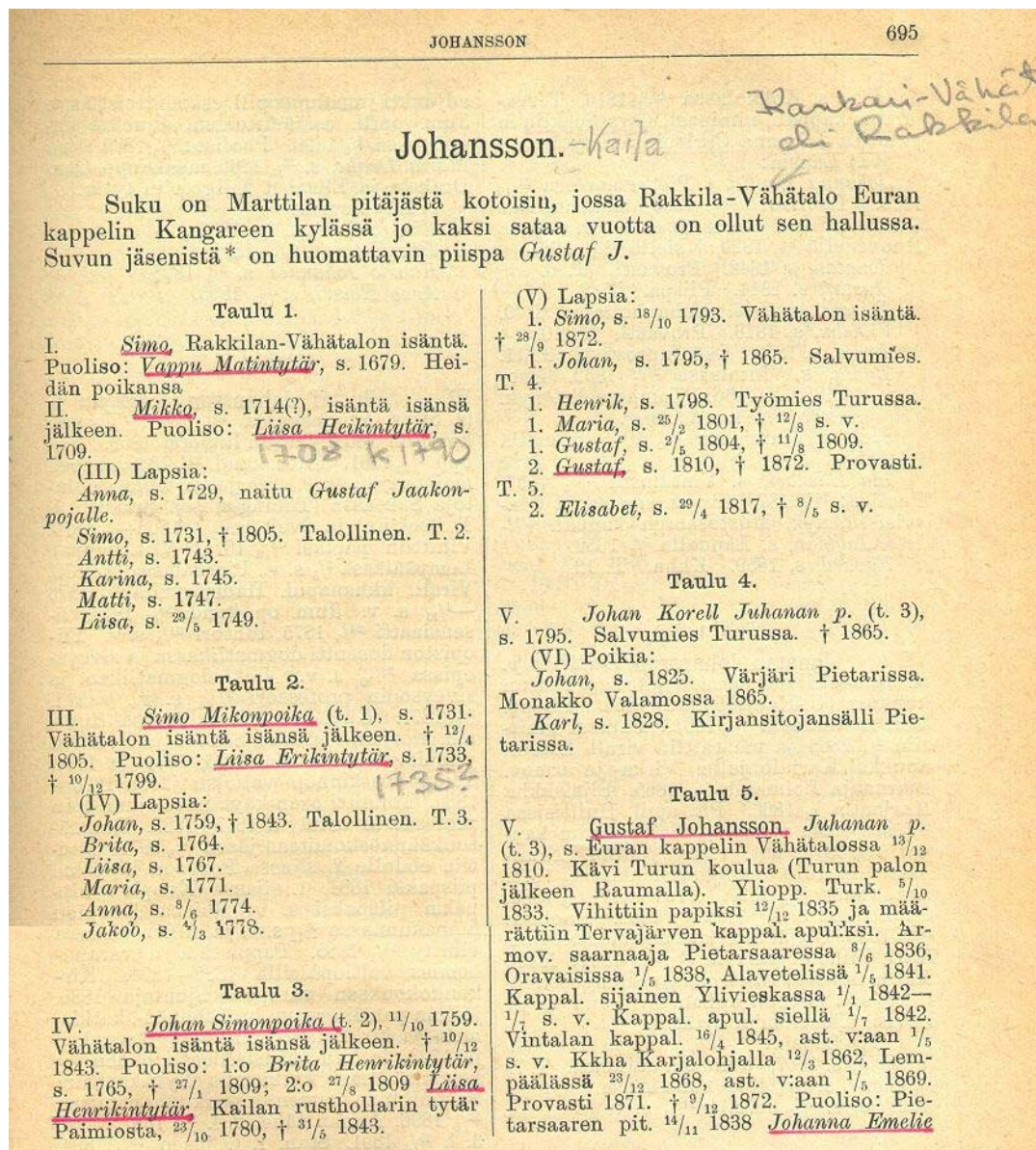
Sukututkimuksen tulokset esitetään useimmiten tekstimuodossa. Jälkipolvirakenteiden suosituin esitysmuoto on kertova esitystapa, jossa pyritään esittämään suvun vaiheet perhekohtaisina tapahtumina. Vastaavasti esipolvirakenne esittää lähtöhenkilön esivanhemmat elämäkertatietoineen.

2.1 Jälkipolviraportti

Suomessa suosituin esitysmuoto jälkipolviraportille on esittää jälkeläiset tauluissa. Taulu sisältää jälkeläisen (tai lähtöhenkilön) puolisoineen ja lapsineen. Taulut numeroidaan juoksevasti esiintymisjärjestyksensä mukaan. Henkilön taulua järjestyksessä seuraavana tauluna tulostuu henkilön vanhimman lapsen taulu. Kun lapsen kaikkien jälkeläisten taulut ovat tulostuneet jatketaan henkilön toiseksi vanhimman lapsen taululla. Taulussa olevien henkilöiden kohdalla viitataan tauluun, jossa hän myös esiintyy esimerkiksi taulun päähenkilönä. Mikäli molemmat puoliset kuuluvat sukuun on kysymyksessä ns. jälkipolvikato. Tällöin on taulussa, jossa hän esiintyy puolisona, viittaus tauluun, jossa hän esiintyy sukuun kuuluvana ja päinvastoin. Lapsettomille suvun henkilöille ei välttämättä haluta muodostaa omaa taulua. Usein riittää, että tämä henkilö on esitetty vanhempiensa taulussa lapsena. Sellaisen henkilön tietoja, jolla on oma taulu, näytetään rajoitetummin vanhempiensa taulussa, kuin henkilön, jolla ei ole lapsia.

Bergholmin [Ber01] sukukirja esitti jälkipolviraportin esimerkin 1 mukaisesti. Georg Luther esitti täydennettynä vastaavat tiedot Kankarin Kailat nro 6 sukukirjassa [Kai00]

esimerkin 2 mukaan. Nämä ovat esimerkkejä käsin tehdyistä jälkipolviraporteista, joita suomalaiset sukututkimusohjelmistot pyrkivät käyttämään esimerkkeinään.



Esimerkki 1. Jälkipolviesitys Bergholmin sukukirjassa

Jälkipolviesitys taulumuodossa nojautuu perhe-olioon perustuvaan perherakenteeseen ja vaatii ohjelmilta, jotka tallettavat tietonsa pääasiassa henkilö-olioon perustuvaan henkilö-olioon perustuvaan välivaihetta, jossa luodaan tilapäinen perherakenne. Nämä rakenteet käsitellään tarkemmin kohdassa 3.1. Tätä perherakennetta ei voi käyttää sellaisenaan, koska tauluun kerätään kaikki perheet, jossa suvun henkilö on puolisona.

Ennen tulostuksen aloittamista tulee kaikki taulut luoda valmiiksi, koska muussa tapauksessa ainakaan taulurakenteessa eteenpäin tehtävät viittaukset eivät onnistu.

Taulu 1

Simo Matinpoika, Kankareen Rakkilan eli Vähätalon isäntä alkuvuodesta 1704, sai syyskuussa vapautuksen ylimääräisistä veroista kolmeksi vuodeksi jotta hän voisi kohentaa rappeutuneen tilan kuntoa. Luopui viimeistään 1720-luvun alussa isännyydestä, mutta asui edelleen Rakkilassa. K. siellä 7.12.1742 ”83-vuotiaana” (tuon ajan ikätiedot eivät ole luotettavia!). Puoliso (1699 tai 1700?) **Vappu Matintytär**, k. Rakkilassa 8.7.1731 ”63 ½ vuoden vanhana”. Vanhemmat: Marttilan Simalan kylän Mattilan isäntä Matti Pertunpoika ja Kirsti.

Lapset:

Heikki Simonpoika, s. noin 1700 (ehkä Simalan kylässä 16.6.1700 syntynyt Heikki, jonka vanhemmiksi kastekirjaan on merkitty Matti Simonpoika ja Vappu Matintytär). Hänen sanotaan olleen Rakkilan isännän, kun venäläiset miehittäjät 1720 veivät hänet sotamieheksi. Vuoden 1723 käräjillä Matti Yrjönpoika, joka oli palannut Venäjältä kotiseudulleen, kertoi tavanneensa Kankareen Heikki Simonpojan Moskovassa 1721.

Mikko Simonpoika, s. 27.9.1704, Rakkilan isäntä, k 1754. **Taulu 2.**

Matti Simonpoika, s. Kankareella 14.2.1708 (kastekirjassa vanhemmat: Simo Yrjönpoika ja Vappu Matintytär) k. siellä 22.8.1709.

Taulu 2

Mikko Simonpoika (taulusta 1), s. Rakkilassa 27.9.1704, oli isäntänä Rakkilassa jo 1722 ensimmäisen isonvihan jälkeisen henkikirjan mukaan. Hänelle kuuluneen pääkunnia Rakkilan kunnostamisesta vauraaksi taloksi ja hänen sanotaan ostaneen sen perintötilaksi 1754. Mikko kuitenkin kuoli jo 6.1.1754, mutta sitä ennen hän lienee aloittanut asian vaatimat toimenpiteet ja hankkinut siihen tarvittavat varat. Puoliso Marttilassa 4.12.1726 **Liisa Heikintytär**, s. Marttilassa, Euran kylässä 6.11.1708, k. Rakkilassa 3.12.1790. Vanhemmat: Turun ja Porin läänin (sodan aikana väliaikaiseksi perustetun) kaksikasratsuväkirykmentin ratsumies Heikki Yrjänänpoika ja Vappu Matintytär, joka ainakin 1740-luvusta lähtien asui Rakkilassa ”ratsumiehen leskenä” ja kuoli siellä 1757 jolloin hänen nimekseen erheellisesti merkittiin Valborg Sigfridsd.r.

Esimerkki 2. Jälkipolviesitys Kankarin Kailat nro 6 sukukirjassa

Taulujen numerointi voi muuttua, jos henkilön sukulaisuussuhteisiin tehdään muutoksia.

Jos esimerkiksi henkilölle syntyy lapsi (joka aikaisemmin ei ole ollut tiedossa) eikä hänellä aiemmin ollut lapsia, muodostetaan henkilölle oma taulu oikeaan paikkaan.

Tulostuksessa täytyy huolellisesti noudattaa tutkijan asettamaa puolisoitten ja lasten järjestystä, joka osaltaan vaikuttaa taulujen numerointijärjestykseen. Sukututkimustietokannassa saattaa olla suuri joukko henkilöitä. Yli 20 000 henkilön tietokannat eivät ole Suomessa harvinaisia. Suurimmissa tietokannoissa lienee yli 90 000 henkeä.

Taulurakenteen muodostaminen voi siis viedä aikaa ainakin, jos tietoja ei säilytetä keskusmuistissa. Henkilörakennemuotoinen tieto voidaan muuttaa perherakenteiseksi siten, että luetaan lapsisuhteet lapsen mukaisessa järjestyksessä. Muodostetaan kustakin

lapsesta olio, jossa hänet yhdistetään tässä vaiheessa esiintyneisiin vanhempiinsa. Tämän jälkeen lajitellaan näin saadut oliot isän ja äidin mukaisessa järjestyksessä. Luetaan puolisosuhteet miehen ja naisen mukaisessa järjestyksessä. Yhdistäminen tehdään niin, että kustakin isä-äiti parista muodostetaan ”perheolio” johon liitetään kaikki siihen sopivat ”perheen lapsioliot”. Lisätään perheolioiksi lapsettomat puoliset. Mikäli huomioidaan myös adoptiotapaukset, tarvitaan edellä mainittuun lisälogiikkaa.

Jälkipolvi- ja taulut numeroidaan alkaen annetusta lähtöhenkilöstä. Taulujen numerointi onnistuu esimerkiksi rekursiivisella metodilla. Jo käsiteltyjen henkilöiden tauluja ei tässä luoda uudestaan vaan niihin muodostetaan jälkipolvikatoviittaukset. Koska ennen tätä vaihetta henkilö- ja rakenteisesta tietokannasta on muodostettu tilapäinen perherakenne tapahtuu taulujen numerointi samantapaisella kaavalla kummassakin tapauksessa.

Esimerkkien 3a ja 3b esittämä RootsWeb [Roo02] verkkosivulta tulostettu esitys jälkipolviraportista Yhdysvalloissa suosituksessa Register-esitysmuodossa poikkeaa numerointimenetelmien ja järjestyksen suhteen suomalaisesta esitysmuodosta. Tässä esityksessä taulut järjestetään sukupolvittain. Ensimmäiseksi esitetään lähtöhenkilön taulu, jota seuraavat hänen lastensa taulut tutkijan asettamassa järjestyksessä, yleensä ikäjärjestyksessä. Lasten taulujen jälkeen esitetään lastenlasten taulut jne.

Descendant Register, Generation No. 1

1. [Simo Matinpoika](#) died 7 DEC 1742 in Rakkila. He married [Vappu Matintytär](#) BET 1699 AND 1700. She died 8 JUL 1731 in Rakkila.

Children of Simo Matinpoika and Vappu Matintytär are:

- 2 i. [Heikki Simonpoika](#) was born ABT 1700.
- + 3 ii. [Mikko Simonpoika](#) was born 27 SEP 1704 in Rakkila, and died 6 JAN 1754 in Rakkila.
- 4 iii. [Matti Simonpoika](#) was born 14 FEB 1708 in Kankare.

Descendant Register, Generation No. 2

3. [Mikko Simonpoika](#) (Simo Matinpoika¹) was born 27 SEP 1704 in Rakkila, and died 6 JAN 1754 in Rakkila. He married [Liisa Heikintytär](#) 4 DEC 1726 in Marttila. She was born 6 NOV 1708 in Marttila, Euran kylä, and died 3 DEC 1790 in Rakkila.

Children of Mikko Simonpoika and Liisa Heikintytär are:

- 5 i. [Anna Mikontytär](#) was born 1 DEC 1727 in Rakkila, and died 27 MAR 1790. She married [Kustaa Jaakonpoika](#) 6 JUN 1749 in Marttila. He was born 3 AUG 1722 in Marttila, Karvelan Isotalo.
- 6 ii. [Maria Mikontytär](#) was born 5 AUG 1730 in Rakkila, and died 15 JAN 1765 in Tapalan Postissa. She married [Erkki Yrjönpoika](#) 24 SEP 1749 in Marttila. He was born 1714, and died 1 AUG 1779 in Posti.
- + 7 iii. [Simo Mikonpoika](#) was born ABT 1731 in (ei löydy Marttilan kastekirjasta), and died 1805 in Rakkila.
- 8 iv. [Jaakko Mikonpoika](#) was born 2 MAR 1733 in Rakkila, and died 20 JUL 1733 in Rakkila.
- 9 v. [Liisa Matintytär](#) was born 20 MAY 1734 in Rakkila, and died 9 MAR 1737 in Rakkila.
- 10 vi. [Vappu Matintytär](#) was born 13 JAN 1739 in Rakkila, and died 16 AUG 1740 in Rakkila.
- 11 vii. [Jaakko Mikonpoika](#) was born 7 APR 1740 in Rakkila, and died 13 JUN 1745 in Rakkila.
- 12 viii. [Antti Mikonpoika](#) was born 14 NOV 1741 in Rakkila.
- 13 ix. [Pirkko Mikontytär](#) was born 4 AUG 1743 in Rakkila.
- 14 x. [Kaarina Mikontytär](#) was born 19 JAN 1745 in Rakkila, and died 20 NOV 1801 in Tapalan Postissa. She married [Mikko Erkinpoika](#) ABT 1765. He was born 31 OCT 1742 in Posti, and died 20 OCT 1785 in Posti. She married [Heikki Rekonpoika](#) 29 DEC 1787 in Koski. He was born 17 MAY 1748 in Pertteli, Romsilan kylä.
- 15 xi. [Matti Mikonpoika](#) was born 30 JAN 1747 in Rakkila, and died 19 AUG 1780 in Marttilan Prunkilan kylässä. He married [Liisa Juhontytär](#) 13 NOV 1768 in Marttila. She was born 1750.
- 16 xii. [Liisa Mikontytär](#) was born 25 SEP 1749 in Rakkila. She married [Heikki Juhonpojan](#) 4 AUG 1776. He was born 1746, and died 10 FEB 1805.
- 17 xiii. [Vappu Mikontytär](#) was born 19 DEC 1751 in Rakkila, and died 7 MAY 1820 in Euran kylän Kirstin Vanhatalo. She married [Matti Matinpoika](#) 9 OCT 1774 in Tarvasjoki. He was born 18 FEB 1748 in Kirsti, and died 17 DEC 1810.

Descendant Register, Generation No. 3

- 7 [Simo Mikonpoika](#) (Mikko Simonpoika², Simo Matinpoika¹) was born ABT 1731 in (ei löydy Marttilan kastekirjasta), and died 1805 in Rakkila. He married [Liisa Erkintytär](#) 22 JUN 1755 in Marttila. She was born 14 FEB 1735 in Marttila, Ruskolaisten kylän Neuvo, and died 14 DEC 1799 in Rakkila.

Children of Simo Mikonpoika and Liisa Erkintytär are:

- 18 i. [Anna Simontytär](#) was born 27 NOV 1756 in Rakkila, and died 20 AUG 1761 in Rakkila.
- + 19 ii. [Juho Simonpoika](#) was born 10 NOV 1759 in Rakkila, and died 22 MAY 1843 in Rakkila.
- 20 iii. [Heikki Simonpoika](#) was born 10 JUN 1762 in Rakkila, and died 4 APR 1764 in Rakkila.

Register esitysmuodossa henkilöt numeroidaan juoksevasti. +-merkki henkilön numeron edessä osoittaa, että henkilöllä on oma taulu seuraavan sukupolven kohdalla. Numerointi voidaan vaihtoehtoisesti tehdä niin, että vain henkilöt, joilla on jälkeläisiä numeroidaan ja lapsettomat jälkeläiset jätetään ilman numeroa. Lasten järjestysnumero osoitetaan (roomalaisilla) numeroilla. Taulun päähenkilön kohdalla voidaan luetella hänen sukuun kuuluvat esivanhempansa varustettuna sukupolvinumerolla.

Esimerkin 3 henkilöiden tiedot ovat varsin vaatimattomia. RootsWeb sivujen kohdalla on mahdollista hyperlinkkien kautta siirtyä henkilön omalle sivulla jossa henkilön kaikki tiedot ovat nähtävissä. Sukukirjaan tulostettaessa yksittäiset taulut voivat olla samantapaisia kuin esimerkin 2 tapauksessa poiketen lähinnä register esityksen erilaisella numerointitavalla.

Genealogia Sursilliana sukukirjan [Sur71] tapaa esittää jälkipolvet esimerkin 4 tapaan on havaittu vaikeammin luettavaksi eikä ole yhtä suosittu kuin taulumuotoinen esitys. Mm Kankarin Kailat sukukirjojen aikaisemmat painokset noudattivat tätä luettelomaista esitystä.

| | | |
|------|--|------|
| XI | <u>Maria Charlotta Johanintytär Hydén</u> , s. 1805. | 8766 |
| X | <u>Anna Elisabet Johanintytär Hydenius</u> , s. Kälviällä 14. jouluk. 1768, k. siellä 10. jouluk. 1829. — Puoliso: Kälviällä 21. helmik. 1786 <u>Christoffer Carlenius</u> , s. Oulussa 14. kesäk. 1751. Ylioppilas Turussa (pohj.) 1771. Vihittiin papiksi 1778. Kälviän kirkkoherra 1808. K. Kälviällä 19. kesäk. 1813. Vanhemmat: Kemin varanimismies ja siltavouti <u>Isak Carlenius ja Beata Liljebäck</u> . | 8767 |
| XI | <u>Johan Christoffer Christofferinpoika Carlenius</u> , s. Kälviällä 21. heinäk. 1788, k. siellä 5. syysk. 1792. | |
| XI | <u>Catharina Elisabet Christofferintytär Carlenius</u> , s. Kälviällä 18. helmik. 1792, k. Kaarlelassa 26. syysk. 1832. — Puoliso: Kälviällä 25. tammik. 1810 Kokkolan pitäjänapulainen <u>Fredrik Stenhagen</u> (N:o 7972), s. Munsalassa 1. toukok. 1782, k. Kaarlelassa 27. tammik. 1835. Vanhemmat: Kokkolan kirkkoherra <u>Fredrik Stenhagen ja Anna Granroth</u> . | 8768 |
| XI | <u>Brita Johanna Christofferintytär Carlenius</u> , s. Kälviällä 16. heinäk. 1794. Oli tehokkaiden lääkevoiteiden tuntija. K. Teerijärvellä 10. heinäk. 1854. — Puoliso: Kälviällä 27. kesäk. 1815 <u>Lars Schalin</u> , s. Vaasassa 18. elok. 1790. Ylioppilas Turussa (pohj.) 1809. Vihittiin papiksi 1813. Teerijärven kappalainen 1840. Varapastori 1841. K. Teerijärvellä 3. elok. 1847. Vanhemmat: Vaasan sairaalan taloudenhoitaja, ravintoloitsija <u>Lars Schalin ja Elisabet Bering</u> . | 8769 |
| XII | <u>Johanna Emelie Larsintytär Schalin</u> , s. Kokkolassa 11. maalisk. 1816, k. Mustasaarella 1. huhtik. 1898. — Puoliso: Pietarsaaren maaseurakunnassa 14. marrask. 1838 <u>Gustaf Johansson</u> , alkuaan Kustaa Vähätalo eli Lillgård, s. Marttilaan kuuluneessa Euran kappelissa (myöhemmällä Tarvasjoella) 13. jouluk. 1810. Ylioppilas Helsingissä (turk.) 1833. Suoritti jumaluusopillisen tutkinnon 1835. Vihittiin papiksi s.v. Lempäälän kirkkoherra 1868. Rovastin arvo 1869. K. Lempäälässä 9. jouluk. 1872. Vanhemmat: talollinen Euran kappelissa <u>Juho Simonpoika Vähätalo ja Liisa Heikintytär</u> . | 8770 |
| XIII | <u>Johanna Emilia Gustafintytär Johansson</u> , s. Oravaisissa 21. lokak. 1839, k. Ruovedellä 14. marrask. 1886. — Puoliso: 12. heinäk. 1859 tilanomistaja, kauppias <u>Gustaf Eduard Palmroth</u> , s. Ruovedellä 9. syysk. 1829, k. siellä 3. tammik. 1901. Vanhemmat: kruununvouti <u>Carl Fredrik Palmroth ja Wilhelmina Forselius</u> . | 8771 |
| XIII | <u>Jonathan Gustafinpoika Johansson</u> , s. Ylivieskassa 22. tammik. 1842. Ylioppilas Helsingissä (pohj.) 1861. Suoritti teologisen erotutkinnon ja vihittiin papiksi 1865. Varapastori 1875. Alajärven | 8772 |

Esimerkki 4. Jälkipolviraportti Genealogia Sursilianan mukaan

Tässä mallissa numeroidaan henkilöt esiintymisjärjestyksensä mukaan. Henkilöön liitetään sukupolvea osoittava numero. Henkilön vanhemmat löytyvät siten, että etsitään

edellinen henkilö, jonka sukupolvi on pienempi kuin henkilön oma sukupolvi. Pienissä esityksissä, joissa sukupolvien lukumäärä on vähäinen, voi tämä olla varsin havainnollinen tapa. Sukupolvet on tällöin mahdollista sientämällä havainnollistaa selkeämmin. Haittapuolena tällä esitystavalla on ensisijaisesti se, että perheen jäsenet hajoavat luettelossa hyvinkin kauaksi toisistaan. Tässä suhteessa tätä esitystapaa voidaan verrata kohdassa 2.2 esitettyyn Stradonizin mukaan numeroituun esipolviraporttiin jossa esivanhemmat voivat olla hyvin kaukana toisistaan.

2.2 Esipolviraportti

Suomessa on esipolvitulosteet tehty perinteisesti siten, että kullekin esivanhemmalle annetaan oma numeronsa, joka on johdettavissa hänen (vanhimmasta sukuun kuuluvasta) lapsestaan. Stefan Kekule von Stradoniz:n vuonna 1898 esittämällä numerointijärjestelmällä [Lut59] henkilön isä saa numeron, joka on henkilön numero kerrottuna kahdella ($n*2$). Hänen äitinsä saa seuraavan numeron ($n*2+1$). Tässä järjestelmässä lähtöhenkilön numero on 1, hänen isänsä on 2, isän isä 4 ja isän äiti 5 jne. Julius Oskar Hagerin vuonna 1905 esittämä järjestelmä poikkeaa tästä siten, että kussakin sukupolvessa numerointi aloitetaan 1:stä ja numeroon lisätään sukupolvinumero, jolloin lähtöhenkilön isä saa numeron 1/1 ja äiti 1/2, isän isän saa numeron 2/1 ja isän äiti 2/2 isän isän isä saa vastaavasti numeron 3/1 jne. Lähtöhenkilölle ei Hagerin järjestelmän mukaan anneta numeroa. Hager numero on muunnettavissa Stradoniz numeroksi (ja päinvastoin) yksinkertaisella laskutoimituksella. Esimerkki 5 esittää edellisessä kohdassa esitettyjä henkilöitä esipolviraporttimuodossa Stradonizin kehittämää numerointijärjestelmää käyttäen. Sukupolvi esitetään tässä roomalaisilla numeroilla ennen henkilön numeroa.

Kummassakin esitystavassa taulunumerot kasvavat sukupolveen nähden 2:n potenssissa. Yli kuuden numeron lukuja on vaikea muistaa, joten tämä esitystapa ei ole kovin helppolukuinen kun sukupolvia on esityksessä yli 20. Koska näin varhaisia esivanhempia todennäköisesti on vain vähäinen määrä ei näitä moninumeroisia lukuja esitykseen tule suurta määrää, joka puolestaan tekee lukemisen helpommaksi.

Näiden esipolviesityksen lukemista haittaava asia on se, että esivanhempien vanhemmat saattavat olla esityksessä hyvin kaukana toisistaan. Kukin esiintyy yksin, mahdollisesti

korkeintaan puolisonsa kanssa yhdessä. Vertaa kohdassa 2.1 esitettyä jälkipolviraportin luetteloesitystä.

. 1

Juho Simonpoika. (Isä 2, Äiti 3). Syntynyt 10.11.1759 Tarvasjoella. Kuollut 22.5.1843 Tarvasjoella. Isäntä. Toimi Vähätalon isäntänä vuodesta 1793 aina 1810-luvun lopulle. Meni toisen kerran naimisiin Liisan kanssa 27.8.1809. Suomen sodan ainoat paikalliset väkivaltaisuuDET sattuivat Kankareella, josta venäläiset ryöstivät ruokatarvikkeita. Venäjän hallitus maksoi kuitenkin täyden korvauksen viedyistä tavaroista. Tilan isännistä on kirjassa erinomainen Georg Lutherin laatima kirjoitus.

I. 2

Simo Mikonpoika. (Lapsi 1, Isä 4, Äiti 5). Syntynyt noin 1731 Marttilassa. Kuollut 12.4.1805 Tarvasjoella. Isäntä.

I. 3

Liisa Erkintytär. (Lapsi 1, Isä 6, Äiti 7). Syntynyt 14.2.1735 Ruskolassa. Kuollut 14.12.1799 Tarvasjoella.

II. 4

Mikko Simonpoika. (Lapsi 2, Isä 8, Äiti 9). Syntynyt 27.9.1704 Marttilassa. Kuollut 6.1.1754 Marttilassa. Isäntä.

II. 5

Liisa Heikintytär. (Lapsi 2, Isä 10, Äiti 11). Syntynyt 6.11.1708 Marttilassa. Kuollut 3.12.1790 Tarvasjoella.

Esimerkki 5. Esipolviraportti numeroitu Stradonizin mukaan

Islantilaiset sukututkimuksen harrastajat katsovat yleisesti esi-isikseen muinaiset viikingit, joista heillä on Islannin saagoissa viittauksia. Tämän takia heidän esittämissä esipolviesityksissä on usein yli 30 sukupolvea, jolloin edellä mainittuja numerointi-järjestelmiä käyttäen esivanhempien numerot voivat olla yli satojen miljoonien luokkaa.

Islannissa on yleisessä käytössä tiiviimpi esipolviesitystapa jossa esivanhemmat kootaan mieslinjan mukaisesti sukukuntiin. Tätä esitystapaa on käytetty myös Kankarin Kailat nro 6 [Kai00] sukukirjassa esivanhempien esittämiseksi. Tämä esitystapa (esitetty esimerkissä 6) perustuu siihen, että esivanhemmat jaetaan runkolinjojen mukaisesti sukukuntiin (ruotsiksi stam). Kaikki esivanhempina esiintyvät naiset muodostavat yhdessä isänsä, isänsä isän jne. sukukunnan.

Jos esityksen lähtöhenkilö on mies muodostaa myös hän sukukunnan isänsä, isänsä isän jne. kanssa. Kukin sukukunta esitetään esityksessä omassa taulussaan jälkipolviraportin tapaan.

Taulu. 1

- 1 **Gustaf Johansson** Syntynyt 13.12.1810 Tarvasjoella. Kuollut 9.12.1872 Lempäälässä. Rovasti. Pääsi Turun katedraaliskouluun ja sitten Turun akatemiaan opiskelemaan papiksi. Hän kieritti Turun palaessa 1827 tavaransa ulos silakkanelikossa. Toimi pappina Teerijärvellä, Pietarsaareissa, Oravaisissa ja Alavetelissä. Ilmeisesti Pietarsaareissa hän tutustui Johannaan ja vihittiin 1838. Toimi kappalaisena Vimpelissä, jossa hän korosti mm. lasten kasvatuksen tärkeyttä. Hän kutsui vaivashoitokunnan päättämään kirjastosta, joka hyväksyttiinkin vaikka eräs jäsen purnasi: "loputkiin nuoret lakkaavat tekemästä työtä, kenoveevankirjaa vain luetaan". Tämän jälkeen Gustaf toimi kirkkoherrana Karjalohjalla, jossa hän ystäväystyi naapuripitäjässä asuvan Elias Lönnrotin kanssa. Loppuvuodet hän toimi Lempäälän kirkkoherrana. Hän oli luonteeltaan tarmokas, mutta vaatimaton. Vietyään pojan Edvardin järven rantaan katsomaan Lempäälään tullutta höyrylaivaa hän vei laivan saapuessa pojan sivummalle: "Vi ska gå bort innan förnämiteten kommer".
- 2 **Juho Simonpoika** Syntynyt 10.11.1759 Tarvasjoella. Kuollut 22.5.1843 Tarvasjoella. Isäntä. Toimi Vähätalon isäntänä vuodesta 1793 aina 1810-luvun lopulle. Meni toisen kerran naimisiin Liisan kanssa 27.8.1809. Suomen sodan ainoat paikalliset väkivaltaisuudet sattuivat Kankareella, josta venäläiset ryöstivät ruokatarvikkeita. Venäjän hallitus maksoi kuitenkin täyden korvauksen viedystä tavaroista. Tilan isännistä on kirjassa erinomainen Georg Lutherin laatima kirjoitus. - Vihitty 27.8.1809. Miehen 2., vaimon 1. avioliitto **Liisa Heikintytär Kaila** (Taulu 2). Syntynyt 23.10.1780 Paimiossa. Kuollut 31.5.1843 Tarvasjoella. Kailan rusthollarin tytär. Harrasti liikuntaa juoksemalla hevosen rinnalla pyhäisin kirkkoon pitkän matkan.
- 3 **Simo Mikonpoika** Syntynyt noin 1731 Marttilassa. Kuollut 12.4.1805 Tarvasjoella. Isäntä. - **Liisa Erkintytär** (Taulu 3). Syntynyt 14.2.1735 Ruskolassa. Kuollut 14.12.1799 Tarvasjoella.
- 4 **Mikko Simonpoika** Syntynyt 27.9.1704 Marttilassa. Kuollut 6.1.1754 Marttilassa. Isäntä. - **Liisa Heikintytär** (Taulu 4). Syntynyt 6.11.1708 Marttilassa. Kuollut 3.12.1790 Tarvasjoella.
- 5 **Simo Matinpoika** Syntynyt noin 1659. Kuollut 7.12.1742 Marttilassa. Isäntä. - **Vappu Matintytär** (Taulu 5). Syntynyt noin 1668. Kuollut 8.7.1731 Marttilassa.

Taulu. 2

- 2 **Liisa Heikintytär Kaila** (Taulu 1).
- 3 **Heikki Matinpoika Kaila** Syntynyt 24.11.1739 Paimiossa. Kuollut 12.4.1821 Paimiossa. Isäntä. Oli Kailan ratsutilan isäntä Paimiossa. Talossa toimitettiin isojako, jossa tilan laajuudeksi todettiin 37 ha viljelysmaata ja 33 ha metsää. - **Liisa Eerikintytär** (Taulu 6). Syntynyt 3.1.1743 Paimiossa. Kuollut 1.5.1814 Paimiossa.
- 4 **Matti Heikinpoika Kaila** Syntynyt 1715 Paimiossa. Kuollut 15.1.1782 Paimiossa. Isäntä. Kailan isäntä 1737 lähtien. Tilalta on tehty karjaluetelo, jossa todettiin talossa olevan 2 hevosta, 1 härkä, 6 lehmää sekä lampaat ja siat. Talo oli jo tällöin melko hyvin toimeentuleva. - **Kirsti Simontytär Klemetti** (Taulu 7). Syntynyt 1712 Paimiossa. Kuollut 3.4.1783 Paimiossa.
- 5 **Heikki Ristonpoika** Syntynyt noin 1679. Haudattu 24.4.1753 Paimiossa. Isäntä. Heikki oli tullut Kailan tilalle vävynä. Talon poikien kuollessa ruttoon hän toimi isäntänä 1712-37. Hän eli myös vaikean ison vihan ajan. - **Kaisa Sipintytär Kaila** (Taulu 8). Syntynyt välillä 1668 - 1669 Paimiossa. Haudattu 26.4.1747 Paimiossa.
- 6 **Risto** Isäntä.

Esimerkki 6. Esivanhemmat esitettynä sukukunnittain

Tässä esitysmuodossa henkilöihin liitetään kaksi numeroa – taulun numero ja sukupolven numero. Lähtöhenkilön taululle annetaan tässä esitysmuodossa numeroksi 1. Taulu sisältää mieslinjan esivanhempien lisäksi myös näiden äidit puolisonsa yhteyteen. Mikäli naisen jompikumpi vanhempi on tiedossa luodaan hänelle oma taulu. Taulut numeroidaan vastaavalla tavalla kuin taulumuotoisessa jälkipolviraportissa siten, että esityksessä seuraavaksi esiintyneen (naisen) taulu saa seuraavan numeron.

Tapausta jossa jonkun esivanhemman lasten jälkeläiset ovat saaneet yhteisiä lapsia kutsutaan esipolvikadoksi. Jos esimerkiksi pikkuserkukset ovat menneet naimisiin keskenään tapahtuu esipolvikato heidän isoisovanhempiensa kohdalla. Yleensä esipolvikato esitetään kuten jälkipolvikato siten, että esivanhemman kohdalle tulostetaan hänen kummatkin esivanhempina esiintyvät lapsensa. Ensiksi esiintyneen lapsen kohdalla viittaus tapahtuu numerointijärjestelmän mukaan normaalisti, mutta myöhempien lasten kohdalla viitataan suoraan jo tulostetun vanhemman tauluun.

3 Sukututkimustiedon mallintaminen

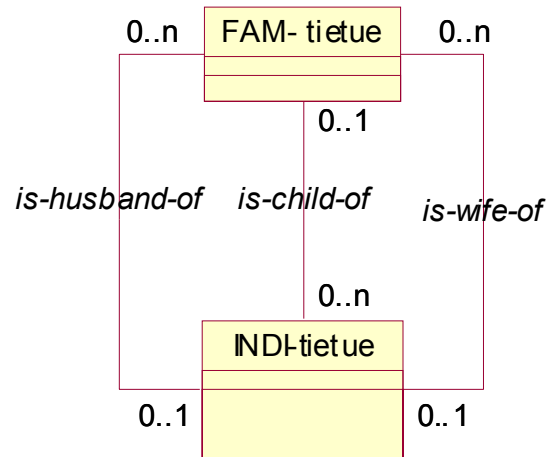
Henkilökohtaisten 16KB muistilla varustettujen 8-bittisten tietokoneiden aikana tyypillinen kysymys tietokoneita käyttävien sukututkijoiden kesken oli: mitä tietokonetta käytät? GEDCOMin ja laitteistoriippumattoman tiedon aikakaudella kysymys oli: mitä ohjelmaa käytät? Tulevaisuudessa kysymys voisi olla: mitä tietomallia käytät [Pri98].

3.1 GEDCOM 5.5

Tammikuussa 1996 julkaistu GEDCOM-standardin versio 5.5 (katso kohta 5) sisälsi kaksi mallikaaviota. Kuva 1 esittää siitä yksinkertaistetun kaavion. Voidaan katsoa, että tämä aloitti tietokoneavusteisen sukututkimuksen toisen sukupolven tai uuden aikakauden – tietomalleihin perustuvan sukupolven.

Useimmat nykyiset sukututkimusohjelmat käsittelevät tietoa, joka on tärkeimmiltä osiltaan muutettavissa GEDCOM-muotoon. Ohjelmat käsittelevät useasti myös tietoa, joka ylittää tämän standardin. Ohjelman uudempi versio sisältää yleensä konversio-rutiinin, jolla voidaan konvertoida ohjelman vanhemman version tietokanta tähän uudempaan muotoon. Ohjelman tekijän kannalta ei ole aina edullista pystyä luomaan kovin täydellistä GEDCOM-tiedostoa, koska tämä voi vaikeuttaa käyttäjäkunnan

pysymistä valitsemassaan ohjelmassa. Toisaalta, jos ohjelma pystyy mahdollisimman hyvin lukemaan GEDCOM-tiedoston, helpottaa se muiden ohjelmien käyttäjien siirtymistä tähän ohjelmaan.

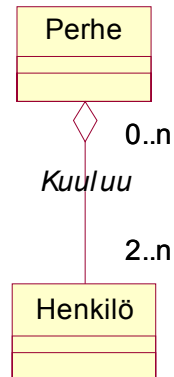


Kuva 1. GEDCOM 5.5 yksinkertaistettu malli

Esimerkkinä asioista, joita GEDCOM-standardilla ei ole mahdollista tallettaa, ovat erikieliset tulostusvaihtoehdot ja tulostuksen ohjaustiedot, kuten esimerkiksi suomenkielen taivutussäännöt paikkakunnan nimille.

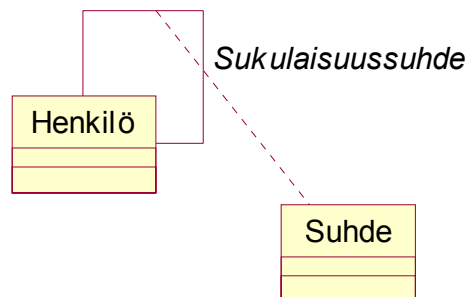
Voidaan kuitenkin todeta, että sukututkimusohjelmat käsittelevät melko homogeenista tietoa. Koska tieto on melko samaa ja sen käyttökin on samantapaista on tämä huolimatta eri teknisistä ratkaisuista mallinnettu pääpiirteittäin samalla tavalla.

On nähtävissä, että sukututkimusohjelmat ovat mallintaneet sukututkimustietoa ainakin perustuen ”perherakenteeseen”, ja ”henkilörakenteeseen”. Kuvassa 2 näkyvä perherakenne koostuu perheoliosta johon perheenjäsenet ovat liitettynä. Perhe koostuu yleensä isästä, äidistä sekä heidän yhteisistä lapsistaan. On olemassa ohjelmia, joissa joudutaan luomaan puuttuvien vanhempien tilalle tyhjiä isä- tai äiti-olioita. Henkilö voi kuulua useaan perheeseen ja perhe sisältää useita henkilöitä. Perherakenne lienee yleisin tapa, jolla sukututkimusohjelmat ovat sukututkimustietonsa mallintaneet. Mm. GEDCOM-tietomalli noudattaa perherakennetta, kuten on havaittavissa.



Kuva 2. Perherakenne

Kuvan 3 henkilö rakenne koostuu puolestaan henkilöolioista, jotka kytketään toisiinsa erilaisten sukulaisuussuhdeolioiden avulla. Perheolioita vastaavat tällöin lapsi-, isä-, äiti- ja puoliso-suhteet. Henkilörakenne on laajennettavissa käsittämään myös muunlaisia suhteita. Esimerkkinä mainittakoon henkilön liittäminen taloon tai joukko-osastoon. Henkilörakenteen haittapuolena on se, että erityisesti jälkipolvitulosteet tehdään henkilöiden perherakenteiden mukaisessa järjestyksessä, joten henkilö rakennetta käyttävän ohjelman täytyy osata tulkitia tietonsa myös perherakenteiden mukaisessa muodossa.



Kuva 3. Henkilörakenne

Sekä perhe- että henkilö rakennetta käyttävän ohjelmiston tulee osata liittää henkilö-, perhe- tai suhdetietoihinsa sekä tapahtuma- että lähdetietoja.

3.2 THE GEDCOM STANDARD (Future Direction)

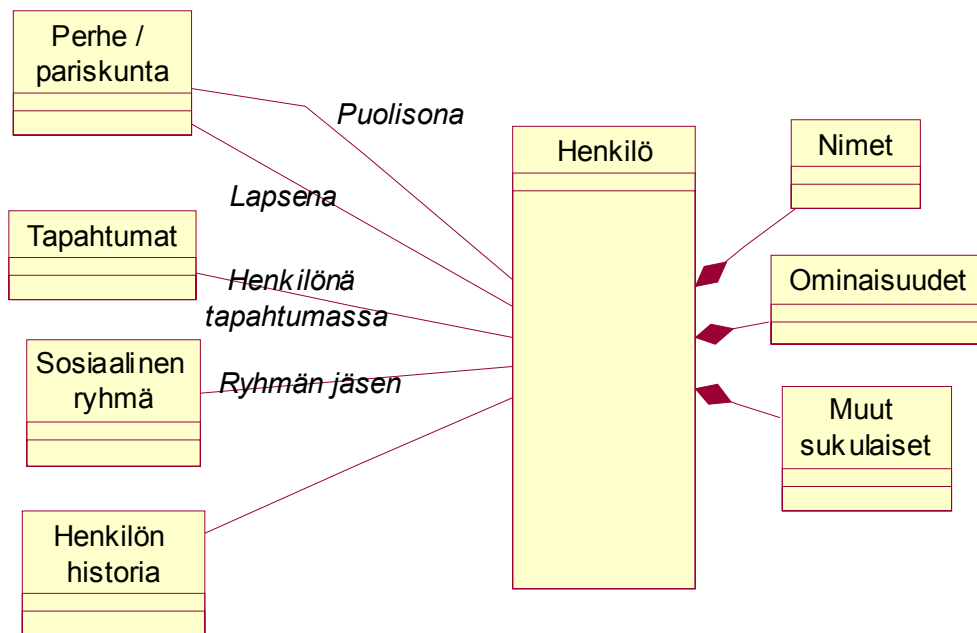
Mormonikirkon henkilöhistoriaosasto julkaisi huhtikuussa 1998 vedoksen uudesta GEDCOM standardista nimellä ”THE GEDCOM STANDARD (Future Direction)”

[Ged98], jossa alaotsikko kertoo, että tämä pohjautuu standardin kuvauksessa esitettyyn tietomalliin. Tämän uusimman GEDCOM version kehitys ei ole edennyt ehdotusastetta pitemmälle, mutta osoittaa osaltaan, että myös GEDCOM-standardin tekijät näkivät tarpeelliseksi sukututkimustiedon nykyistä tarkemman mallintamisen.

Ehdotus pohjautuu laajaan tietomalliin. Malli jakautuu osamalleihin, jotta sen ymmärtäminen olisi helpompaa.

Henkilörakenne (Individual) on keskeisellä sijalla mallin *Sukututkimustiedot osamallissa*. Kuva 4 näyttää, kuinka henkilö rakenne esittää tärkeintä osaa mallin tässä osassa.

Nimitiedot (Genealogical Name) sisältyy henkilö rakenteeseen. Tätä rakennetta käytetään vain nimitiedoissa henkilökohtaisia nimiä varten. Tähän kuuluu kaikki tieto, jota tarvitaan nimen tarkistamista, standardisointia ja dokumentointia varten sekä henkilökohtaisten nimien käsittelyä varten mukaan lukien henkilön käyttämät eri nimet eri aikoina ja eri paikoissa.



Kuva 4. Sukututkimustiedot

Myös *Henkilökohtaiset ominaisuudet* (Individual Characteristic) rakenne sisältyy henkilö rakenteeseen. Tällä rakenteella talletetaan henkilön ominaisuuksia, joiden avulla

hänen henkilöllisyytensä voidaan selvittää tai joista on apua tutkimuksessa tai jotka lisäävät mielenkiintoa häntä kohtaan.

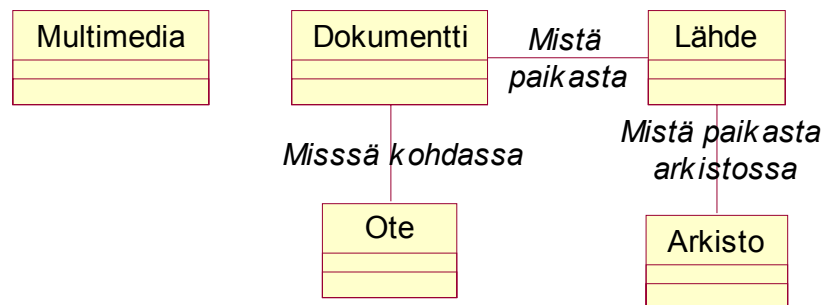
Edellä mainittujen lisäksi *tiedot muista sukulaisista* (Associated Individual Link) rakenne sisältyy henkilörakenteeseen. Tällä rakenteella kuvataan henkilön kytkennät muihin henkilöihin kuin puolisoihin ja vanhempiin/lapsiin. Esimerkiksi voi olla tiedossa, että Paavo Virtanen on Maija Meikäläisen setä, mutta tiedot heitä yhdistävistä henkilöistä puuttuvat.

Perhe – pariskunta (Family / Couple) sisältää tietoa avioliittoon perustuvista perheistä tai naimattomista pariskunnista, joilla on lapsia. Henkilötietue sisältää viittauksia Perhe – pariskunta tietueeseen sekä perheen vanhempina että sen lapsina.

Tapahtumarakenteeseen (Event) talletetaan tietoa, joka liittyy henkilöihin ja perheisiin ja joka on sukututkimuksen kannalta mielenkiintoista. Tapahtumatietueisiin viittaavat pääasiassa henkilö-, perhe – pariskunta- ja sosiaalinen ryhmä-tietue tai näiden väliset kytkennät.

Sosiaaliset ryhmät (Social Group), kuten naapurit, kerhot, ammatilliset yhteisöt jne voivat olla käyttökelpoisia lähteitä joiden avulla henkilöiden henkilöllisyyttä ja elämäkertatietoja on mahdollista selvittää. Sosiaalisiin ryhmiin voivat viitata henkilötietueet sekä toiset sosiaaliset ryhmät.

Kuvan 5 esittämä *Bibliografiset tiedot osamalli* koostuu viidestä pääosasta.



Kuva 5. Bibliografiset tiedot

Dokumentti (Document) tietue esittää dokumenttia tai tallennettua tietoa, jota voidaan käyttää sukututkimustiedon todistuskappaleena. Esimerkkejä dokumenteista ovat papintodistus, väestölaskentatodistus tai kirjan kappale, joka sisältää tietoa määrätystä

perheestä tai tapahtumasta. Dokumentti on yleensä osa laajempaa lähdeaineistoa, kuten kyseisen pappilan arkisto tai kyseinen väestönlaskenta. Kyseinen lähdeaineisto puolestaan voi löytyä useammastakin paikasta tai arkistosta.

Lähde (Source) on kokoelma sukututkimuksen kannalta mielenkiintoista aineistoa kuten kirja, seurakunnan arkisto jne. Lähteeseen viittaavat tärkeimmät tiedot ovat dokumentit ja lähteen sisältämät arkistot.

Arkisto (Repository) tiedot sisältävät tietoa paikoista, joissa sukututkimuksen lähteitä talletetaan kuten Valtionarkisto tai Tyrvään seurakunta.

Ote (Extracted Detail) esittää yleensä pientä tiedon osaa, joka sisältyy sen viittaamaan dokumenttiin. Otetta käytetään yleensä yhden tai useamman tapahtumamäärän johtopäätösten todistuskappaleena. Esimerkkinä tällaisista johtopäätöksistä ovat henkilön syntymäaika ja paikka. Viittauksia otteeseen voivat tehdä henkilö-, perhe-pariskunta-, tapahtuma- ja sosiaalinen ryhmätieto.

GEDCOM (Future Direction)-dokumentti korostaa tietomallin merkitystä, jota se edellyttää noudatettavan tunnollisesti. Ehdotus sisälsi erikseen osan pelkän sukututkimustiedon ja koko sukututkimustietokannan siirtotiedostosta.

Sukututkimustietokannan siirtotiedoston oletetaan sisältävän kaikki tietokannan tiedot. Malli on varsin laaja ja siinä oletetaan, että mikään yksittäinen ohjelmisto ei tulisi toteuttamaan kaikkia sen piirteitä. Sukututkimustiedon siirtotiedosto on tarkoitettu tiedonsiirtoon kahden eri järjestelmän välillä.

Ehdotus sisältää runsaasti tietoa, jota GEDCOM 5.5 ei tunne. Ehdotuksen fyysinen GEDCOM-rakenne noudattaa samaa GEDCOM-tietuerakennetta kuin GEDCOM 5.5, mutta poikkeaa merkittävästi edeltäjästään yksittäisen rivin kohdalla.

Ehdotus määrittelee selkeämmin kuin edeltäjänsä todiste- ja lähdetiedon tallettamisen sekä erikseen näistä tehtyjen johtopäätösten jälkeen saadun tiedon.

Mormonikirkon tietokanta-arkkitehtuurijaoksen johtaja Robert Booth esitti tämän mallin Denverissä pidettävässä sukututkimuskonferenssissa 6.5.1998. GEDCOM (Future Direction) esitettiin keskustelun avaamiseksi aiheesta. Keskustelua ei tämän esityksen pohjalta kovin paljon kuitenkaan syntynyt. Syynä tähän on ainakin osittain se, että

GENTECH Inc julkaisi Lexicon-nimisen työryhmänsä suunnitteleman mallin saman vuoden syksyllä (katso luku 7). Mormonikirkko lienee vetänyt standardiehdotuksensa myöhemmin takaisin.

4 Sukututkimusohjelmat

Varhaisimmat tiedot kaupallisista sukututkimusohjelmistoista julkaistiin tietokonealan julkaisuissa 1979 [Pri01]. Tätä seurasi seuraavana vuonna kaksi merkittävää artikkelia Atlantin molemmin puolin. Pian tämän jälkeen syntyi tietokoneen käyttöön sukututkimuksessa keskittyneitä yhdistyksiä tärkeimpien sukututkimusyhteisöjen yhteyteen sekä näiden myötä aihetta käsitteleviä julkaisuja. Mainittakoon näistä *Genealogical Computing* (Yhdysvallat 1981), *Computers in Genealogy* (Englanti 1982), *Genealogia & Computer* (Belgia 1983), *Computergenealogie* (Saksa 1984) sekä *Sukutieto* (Suomi 1984). Tietokoneavusteiselle sukututkimukselle oli vuosi 1986 merkittävä virstanpylväs. Tuolloin Mormonikirkon perhehistoriaosasto julkaisi Personal Ancestral File (PAF) ohjelman version 2.0 usealle laitetypille (MS-DOS, Apple II ja CP/M). Tämän ohjelman osana he julkaisivat GEDCOM-siirtoformaatin, jonka merkitys on keskeisellä sijalla tämänkin päivän sukututkimusohjelmistoille. GEDCOM-formaatista julkaistiin PAF-ohjelman seuraavan päivityksen (2.1) yhteydessä vuonna 1987 uusi versio, joka on säilynyt perusteiltaan lähes muuttumattomana näihin päiviin asti.

Useimmat nykyiset sukututkimusohjelmat käsittelevät samantapaisia tietoja. Tämä lienee luonnollista, koska tarve on tallettaa samaa tietoa. Kukin ohjelma on kuitenkin tehty hieman eri lähtökohdista ja eri käyttäjäryhmiä varten, joten detaljitasolla on tiedon käsittelyssä eroja. Tekijän 29.3 – 3.4.2001 Suomen Sukututkimusseuran internet-palvelimen suku- ja sukuohjelmisto- postituslistoilla ja news-ryhmässä sfnet.harrastus.sukututkimus tekemä kysely, jossa tiedusteltiin käyttäjiltä mitä sukututkimusohjelmaa he käyttävät, antoi taulukko 1:n mukaisen tuloksen.

Jäljempänä esitetään eräiden suosituimpien ohjelmien ominaisuuksia. Sukuohjelmiston tiedot perustuvat tekijän tietoihin. Sukujutut-ohjelman tekijä Seppo Luhtasaari kuvasi ohjelmansa toimintoja ja Brother's Keepersin tekijä John Steed luovutti pyydettyä kuvauksen ohjelman tietorakenteista. LifeLines-ohjelman ominaisuuksia on ohjelman tekijä kuvaillut GEDCOM-L- ja LINES-L- postituslistoilla. Rafal Prinke [Pri00] antoi

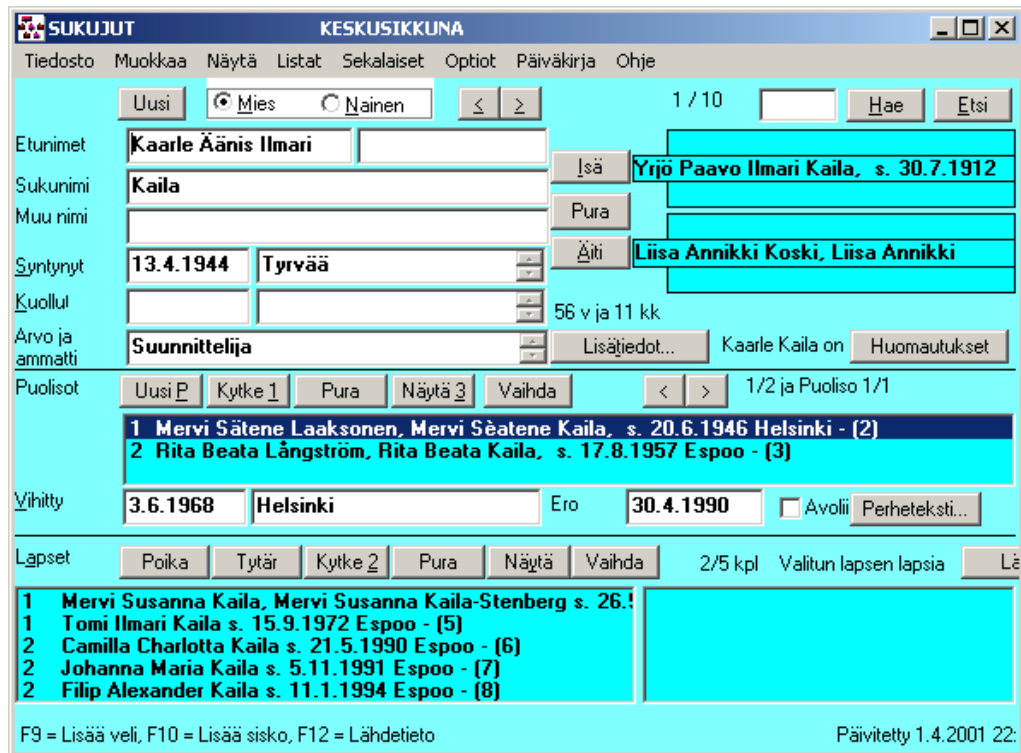
näkemyksensä LifeLines ohjelmasta. Muiden ohjelmien toiminnan selvittäminen perustuu ”reverse engineering” tyyppiseen menetelmään.

| Ohjelma | <i>lkm</i> | <i>%-osuus</i> |
|---|------------|----------------|
| Sukuohjelmisto | 90 | 36,3 |
| Genus | 34 | 13,7 |
| Sukujutut | 32 | 12,9 |
| Brother's Keeper | 27 | 10,9 |
| Juuret | 16 | 6,5 |
| Family Tree Maker | 10 | 4,0 |
| Personal Ancestral File | 10 | 4,0 |
| Tekstinkäsittely- tai muu valmisohjelma | 7 | 2,8 |
| Itse tehty ohjelma | 3 | 1,2 |
| WinFamily | 2 | 0,8 |
| MinSläkt | 2 | 0,8 |
| Disgen | 2 | 0,8 |
| Legacy | 2 | 0,8 |
| Reunion | 2 | 0,8 |
| Gene | 2 | 0,8 |
| LifeLines | 1 | 0,4 |
| Cumberland Family Tree | 1 | 0,4 |
| Suku-Sampo | 1 | 0,4 |
| GedHTree | 1 | 0,4 |
| GeneWeb | 1 | 0,4 |
| TheMasterGenealogist | 1 | 0,4 |
| Holger | 1 | 0,4 |
| Yhteensä vastauksia | 248 | 100 |

Taulukko 1. Kyselyn mukaan käytetyt sukututkimusohjelmat

4.1 Sukujutut

Seppo Luhtasaaren tekemän Sukujutut-ohjelman [Luh00] avulla täytetään kuvan 6 esittämän ohjelman keskusikkunan avulla henkilökortteja ja yhdistellään niitä perhekorteiksi. Henkilöstä talletetaan normaalit tiedot: etunimi, sukunimi, syntymäaika ja -paikka, kuolinaika ja -paikka, ammatti sekä lisätiedot (esim. osoitteet, elämäkerta ja lähdeviitteet). Ohjelmassa on monipuoliset ja helpot tavat tehdä perheitä. Usein päähenkilön tietoihin yhdistetään ensin puoliset ja tämän jälkeen kunkin puolison alaisuuteen yhdistetään heidän yhteiset lapsensa. Puoliset yhdistetään avioliiton solmimisjärjestyksessä. Kun puoliso on yhdistetty, niin tämän jälkeen voidaan tallettaa vihkiaika ja -paikka. Tämän jälkeen yhdistetään puolisoitten lapset siten, että vanhin lapsi yhdistetään ensin. Yhdellä henkilöllä voi olla enintään 10 puolisoa ja 30 lasta.



Kuva 6. Sukujutut-ohjelman keskusikkuna

Henkilöitä lisäämällä saadaan samalle näytölle syntymään perheitä. Lapsella on oltava aina vanhemmat, mutta mikäli toinen vanhemmista ei ole tiedossa, niin jätetään ko. henkilön etu- ja sukunimitiedot tyhjäksi. Mikäli lapset tai puolisot ovat väärässä aikajärjestyksessä, niin valitsemalla hiirellä siirrettävä henkilö ja painamalla Vaihda -painiketta tai hiiren oikeanpuoleista näppäintä kysyy ohjelma uutta järjestysnumeroa ja vaihtaa järjestystä.

Aloittaessaan Sukujutut-ohjelman tekemisen ei Luhtasaarella aluksi ollut tietokantoja käytettävissä, joten hän kehitti vastaavat tietorakenteet suorasaanti- ja binääritiedostojen avulla. Näitä hän on vuosien varrella kehitellyt ja parannellut. Ne ovat varsin nopeita käyttää.

Erityyppiset tiedot talletetaan ohjelman työhakemistoon erillisiin tiedostoihin.

Henkilötiedon kanssa talletetaan samaan tiedostoon myös henkilöön liittyvät lähteet.

Perhesuhteet talletetaan tiedostoon perheet.idx, mutta siihen liittyvät henkilöt syntymä ja kuolintietoineen, teksteineen, paikkakuntien nimineen ja lähteineen talletetaan nimet.idx tiedostoon.

Kaikki yhteen ”sukutietokantaan” liittyvät tiedostot ovat siihen liittyvässä hakemistossa. Hakemistossa oleva tiedosto suku.idx sisältää tietokannan tiedot lyhenteinä ja osoittimina. Tiedot haetaan tarvittaessa muista tiedostoista.

Idx-tiedostoja on tämän lisäksi muutama muu, joihin ohjelma tallettaa tarvitsemiaan osoittimia ym. Ohjelma käsittelee idx-tiedostojaan suora hakumenetelmällä pitäen ajon aikana muistissaan vain tarvitsemansa tietorakenteet. Tämän rakenteen avulla erilaiset haut on saatu optimoitua. Sukujutut-ohjelmasta on olemassa sekä 16- että 32-bittinen versio.

On havaittavissa, että Sukujutut-ohjelma toimii kohdassa 3.1 kuvattua perherakennetta käyttäen.

Sukujutut-ohjelma sisältää noin 20 eri tulostusvaihtoehtoa. Mukana on myös yksinkertaisia tarkistuslistoja. Graafisia taulumuotoisia esityksiä ohjelma tulostaa suoraan kirjoittimelle. Eräät tekstitulosteet tulostuvat yksinkertaiseen tekstitiedostoon mutta osalla raporteista vaihtoehtoina on myös rtf- ja/tai html-muoto. Ohjelmalla on mahdollista tuottaa kohdan 2.1 raporteista taulumuotoinen jälkipolviraportti. Ohjelman esipolviraportti ei noudata kohdan 2.2 mallin mukaisia esimerkkejä. Siinä tulostetaan esivanhemmat taulumuotoisen jälkipolviraportin tapaan perheineen. Nämä taulut numeroidaan juoksevasti kuten jälkipolviraportissa.

4.2 Juuret

Raisiolaisen Consegur Oy:n Juuret-ohjelmassa [Juu00] hakuikkuna sisältää kaikki tiedostossa olevat henkilöt. Näyttöä on mahdollista rajoittaa muutamalla valinnalla. Voidaan esimerkiksi näyttää vain miehet, lapsettomat jne. Hakuikkuna voidaan lajitella ikkunassa olevien kenttien mukaan. Etu- ja sukunimen lisäksi esimerkiksi äidin tai isän etunimen mukaan tai syntymä- tai kuolinajan mukaan.

Henkilöikkunassa on kiinteä joukko kenttiä joihin henkilön tietoja voidaan syöttää. Erityyppisiä kenttiä on varsin monta.

Juuret-ohjelman sukutietokanta koostuu yhdestä tiedostosta, esim Kaila.jrt. Ohjelma pitää tietonsa keskusmuistissa ohjelman suorituksen ajan tekstinkäsittelyohjelman tapaan.

Käyttäjä tallettaa Talleta-käskyllä tiedot tehtyään niihin haluamansa muutokset. Kuva 7

esittää ohjelman päänäyttöä josta on esillä henkilön perustietojen tallennuksen lipuke. Keskusmuistissa Visual Basicilla tehty ohjelmisto käsittelee tietoja todennäköisesti tarpeellisena määränä taulukoita, joihin tiedoston tiedot luetaan kun tiedosto avataan. Tiedonhaku keskusmuistissa olevista taulukoista toimii ainakin kohtuullisen kokoisella tietokannalla varsin ripeästi.

The screenshot shows the 'Henkilö' (Person) window in the Juuret v. 2.0 Senior software. The window title is 'Juuret v. 2.0 Senior - kk6.jrt - [Kaalo, Kaarle Aanis Ilmari, * 1944, Tyrvää]'. The menu bar includes 'Tiedosto', 'Muokkaa', 'Asetukset', 'Toiminnot', 'Ikkuna', and 'Ohjeet'. The toolbar contains various icons for file operations and editing. The main area is divided into tabs: 'Haku', 'Henkilö', 'Sukupuut', 'Vanhemmat', 'Puolisot', 'Lapset', 'Multimedia', and 'Talot'. The 'Henkilö' tab is active, showing a form with the following data:

| | | | | | |
|---------------------------|-----------------------------|--|---|--|-----|
| Sukunimi Kaalo | Etuliite | Etunimi Kaarlo Aanis Ilmari | | <input checked="" type="radio"/> Mies <input type="radio"/> Nainen <input type="radio"/> Ei tietoa | |
| Avionimi | Etuliite | Isän etunimi Yrjö | | | |
| Edellinen sukunimi | Etuliite | Äidin etunimi Liisa | | | |
| Syntymäaika 13.04.1944 | Pitäjä / Kaupunki Tyrvää | Kylä / Kaupunginosa | Talo / Katu | Nro | Maa |
| Kasteaika | Pitäjä / Kaupunki | Kylä / Kaupunginosa | Talo / Katu | Nro | Maa |
| Rippiaika | Pitäjä / Kaupunki | Kylä / Kaupunginosa | Talo / Katu | Nro | Maa |
| Vihkiaika | Pitäjä / Kaupunki | Kylä / Kaupunginosa | Talo / Katu | Nro | Maa |
| Kuolinaika | Pitäjä / Kaupunki | Kylä / Kaupunginosa | Talo / Katu | Nro | Maa |
| Hautausaika | Pitäjä / Kaupunki | Hautausmaa | Kuolinsv | Maa | |
| Lähiosoite | Postinumero | Postitoimipaikka | Maa | | |
| Koulutus | Sotilasarvo | <input checked="" type="checkbox"/> Elää | <input type="checkbox"/> Tiedot tarkistettu | Kieli <input checked="" type="radio"/> Suomi <input type="radio"/> Ruotsi | |
| Ammatti Suunnittelija | Silmien väri | Huusten väri | <input type="checkbox"/> Ei suvunjatkaja | | |
| | | | <input type="checkbox"/> Adoptiolapsi | | |

At the bottom of the window, there are buttons for 'Uusi', 'Poista', 'Edellinen', and 'Seuraava'.

Kuva 7. Juuret-ohjelman henkilöikkuna

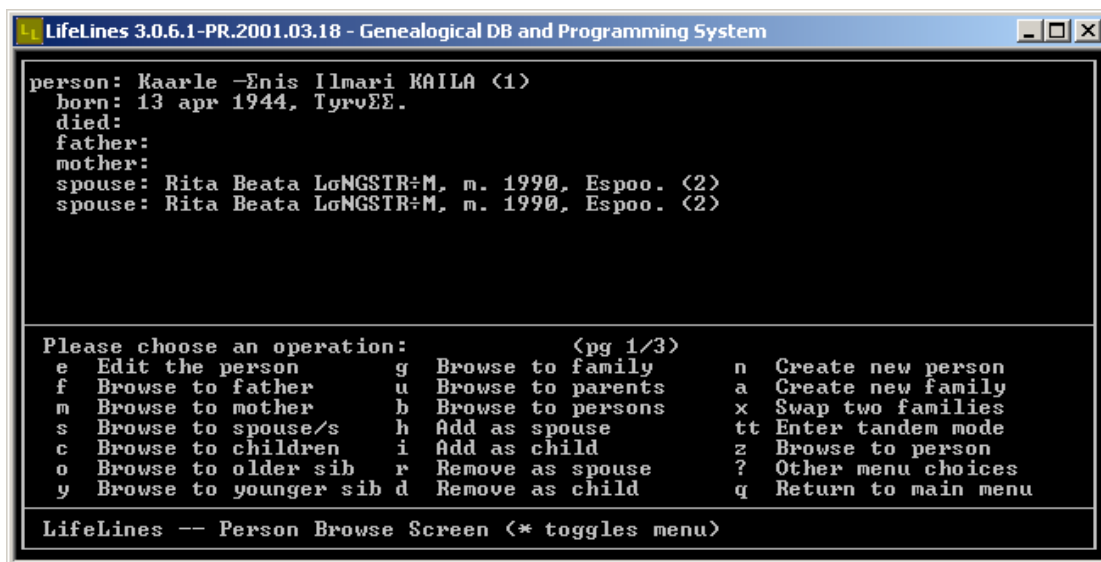
Ohjelman mukana seuraa luettelo historiallisista tapahtumista, ajankohtien hallitsijoista ym, joita ohjelma voi esittää käyttäjälle tallettamiensa henkilöiden yhteydessä.

Ohjelmasta saa käsityksen, että se noudattaa kohdan 3.1 mukaista henkilörakennetta sukulaisuussuhteiden käsittelyssä.

Juuret-ohjelmalla on mahdollista muodostaa kohdan 2.1 mukainen taulumuotoinen jälkipolviraportti. Tulostus voidaan tuottaa rtf- tai html-muotoisena. Ohjelman esipolviraportti ei noudata kohdan 2.2 mallin mukaisia esimerkkejä. Se on toteutettu kuten vastaava raportti kohdan 4.1 Sukujutut-ohjelmalla.

4.3 LifeLines

Thomas T. Wetmore IV:n tekemä LifeLines [Wet91] poikkeaa monessa suhteessa muista tässä esitettävistä sukututkimusohjelmista. LifeLines on ensisijaisesti tehty unix-käyttöjärjestelmää silmälläpitäen. Sillä ei ole graafista käyttöliittymää kuten Windows-ohjelmissa. Kuvassa 8 näkyvä LifeLines-ohjelman Windows-version näyttö esittää henkilön perhesivun. Tämä Windows-versio käynnistää halutun tekstieditorin, esimerkiksi notepad.exe, jolla luvussa 5 esitettävän GEDCOM-standardin tapaisia henkilö- perhe- tai lähdetietoja voi muokata. Monen LifeLines-käyttäjän mielestä



Kuva 8. LifeLines ohjelman Windows version henkilönäyttö

ohjelma on kuitenkin paras kaikista. Tärkein LifeLines-ohjelman ominaisuus on, että se on erittäin joustava. Ohjelmalle voi koodata lähes kaikenlaista tietoa. Tieto koodataan GEDCOM-syntaksin mukaisesti (katso kohta 5). Käyttäjä voi keksiä tarpeen mukaan omia kenttänimiöitä (tag) ja rakenteita. Tietokenttien pituuksia ei ole rajoitettu. Samaa nimiötä voidaan käyttää useampaan kertaan esittämään samantapaista tietoa. Nämä ovat ominaisuuksia, joiden toteuttaminen perinteisissä relaatiotietokannoissa on joko mahdotonta tai ainakin varsin vaikeata, koska rakenne saattaa tulla varsin monimutkaiseksi.

LifeLines-ohjelman ainoa puute on, että se käyttää vain GEDCOM-standardin syntaksia. Eri käyttäjillä voi olla varsin erilaiset rakenteet eikä tämä tieto ole yhtä helposti siirrettävissä kuin muiden ohjelmien kohdalla. Vaikka LifeLines hyväksyy minkä tahansa

GEDCOM-tiedoston, se ei tarkista sen semantiikkaa muuta kuin erittäin rajoitetusti. Toisin sanoen ei ole olemassa LifeLines-tietomallia hyvin väljää GEDCOM-tyylistä mallia lukuunottamatta. Tämä malli sisältää yksilö- ja perhetietueet ja näiden väliset perhesuhteita osoittavat kytkökset.

LifeLines-ohjelman raporttiohjelmilla voi eräässä mielessä pakottaa tietokannan tietyn rajoittavamman mallin mukaiseksi. Verify-raporttiohjelma ajamalla saadaan virheilmoitukset niistä kohdista, joissa tämän raporttiohjelman eheysrajoitteet eivät ole voimassa. LifeLines mahdollistaa näin väljän GEDCOM-syntaksin puitteissa erilaisten tietomallien määrittämisen.

LifeLinesin tiedot säilytetään tekstitietokantana, jota indeksoidaan tehokkaasti B-puu-hakemistojen avulla. LifeLinesin henkilö- ja perhetietoja muokataan käyttäjän valitsemalla tekstieditorilla. Editoitava tieto muokataan GEDCOM-tiedoston syntaksin mukaisina palasina. GEDCOM-standardin osaaminen onkin LifeLines-ohjelman käyttäjälle tärkeitä. LifeLines-ohjelman sukulaisuussuhteet muodostetaan GEDCOM-standardin mukaista perherakennetta käyttäen.

LifeLines-ohjelmistoon kuuluu oma ohjelmointikielensä, joten tulosteita rajoittaa vain käyttäjän kyky määritellä haluamansa ominaisuudet tällä kielellä. Kyseinen kieli on ”oikea” ohjelmointikieli ja sillä kirjoitettuja raportteja on runsaasti tarjolla, eräät niistä ovat varsin pitkälle kehitettyjä.

LifeLines-ohjelman käyttö soveltuu parhaiten tottuneelle UNIX-käyttöjärjestelmän käyttäjälle. Tämä ei ole kovin yleistä sukututkijoiden joukossa, joka selittää sen vähäisen suosion.

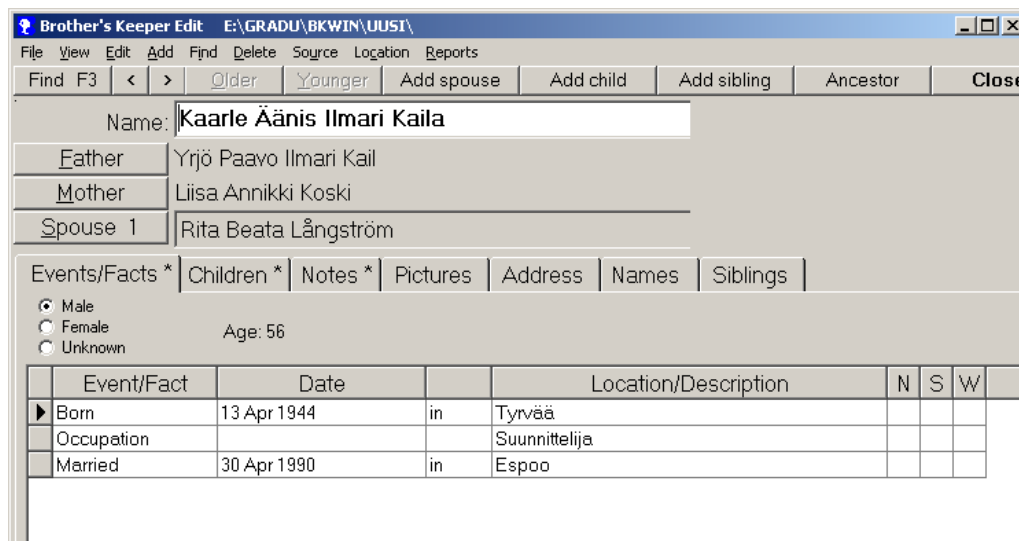
4.4 Brother's Keeper

John Steed aloitti tietokoneiden kanssa työskentelyn 1975. Hän ryhtyi ensimmäisen sukututkimusohjelmansa tekemiseen vuonna 1982. Hän sai valmiiksi ohjelman ennen 1984 tapahtunutta sukukokousta, jolloin hän näytti kokouksen osallistujille heidän sukulaisuussuhteensa muihin juhlan osallistujiin. Myöhemmin hän teki ohjelmaan muutoksia ja ryhtyi jakamaan sitä shareware-periaatteella. Saamansa vähäisen palautteen mukaan hän ryhtyi tekemään ohjelmaan muutoksia. Hän ei ole sukututkija itse, mutta

kuunteli käyttäjien toiveita ja ehdotuksia. Hän työsti ohjelmaa melkein kolme vuotta ennen kuin se oli niin menestyksekkäs, että hän saattoi luopua päivätyöstään. Hän liittyi 1989 yhdistykseen ”Association of Shareware Professionals” ja työskentelee nykyisin täyspäiväisenä ohjelman ylläpidon kanssa.

Brother’s Keeper -sukututkimusohjelmalla [Ste82] käsitellään henkilöitä eri tyyppisine tietoineen kuvan 9 esittämältä ohjelman ylläpito näytöllä. Tältä näytöltä pääsee hiiren napsautuksella siirtymään sukulaiselta toiselle. Ikkunasta pääsee syöttämään monipuolisesti henkilön elämään liittyviä tapahtumia, ominaisuuksia, osoitteita ja nimi-tietoja. Tietoihin on myös lisättävissä teksti- ja lähdetietoja tarpeen mukaan.

Brother’s Keeper -ohjelman 16-bittinen versio BK5 on rajoitetumpi ominaisuuksiltaan kuin 32-bittinen BK6 -ohjelmisto. Esimerkiksi BK5.2:ssa nimen ja paikkakuntakentän pituus on 40 merkkiä kun se BK6:ssa on 100 merkkiä. BK5:ssä henkilön tapahtumilla on käytössä vain 7 tärkeintä tapahtumatyyppeä kuten syntynyt ja kuollut. BK6:ssa on mahdollista esittää yli 100 erilaista tapahtumatyyppeä, joita kutakin voi olla useita kappaleita. Vastaavia laajennuksia on sukulaisuus ja lähdetietojen esittämisessä.



Kuva 9. Brother’s Keeper 6:n Muokkaa-näyttö

Brother’s Keeper tietokantamoottorina toimii Btrieve-tietokantaohjelmisto [Per00]. BK sisältää Btrieve-ohjelmiston ajonaikaisen version, joten sitä ei tarvitse hankkia erikseen. Btrieve-tietokannan taulut sijaitsevat fyysisesti erillisissä tiedostoissa. Esimerkiksi BK5:ssä tiedosto BKPERSON.DT5 sisältää henkilötiedot ja BKMARR.DT5 tiedosto

sisältää perhettä koskevat tiedot. Henkilötietue sisältää osoittimia paikkakunta-, lähde- ja puoliso-tietueisiin ja perhetietueesta löytyy osoittimet lasten tietueisiin kahdesta taulukosta, joihin mahtuu enimmillään 24 lapsen osoittimet.

BK6:ssa tämälntapaiset rajoitukset ovat poistuneet ja BK6 sisältää myös useampia tiedostoja eli Btrieve-tauluja kuin BK5.

Tekstitiedot ja kuvat talletetaan omiin hakemistoihinsa. Ohjelman käyttöliittymän tekstit sijaitsevat erillisessä tekstitiedostossa, joita käyttäjät eri maissa ovat kääntäneet omalle kielelleen. Tietävästi Suomessakin on muutama erillinen käännösversio BK5:n teksteistä. BK5:n tekstit soveltuvat osittain myös BK6:n kanssa käytettäväksi.

Brother's Keeper-ohjelmisto noudattaa sukulaisuussuhteiden käsittelyssä kohdan 3.1 mukaista perherakennetta.

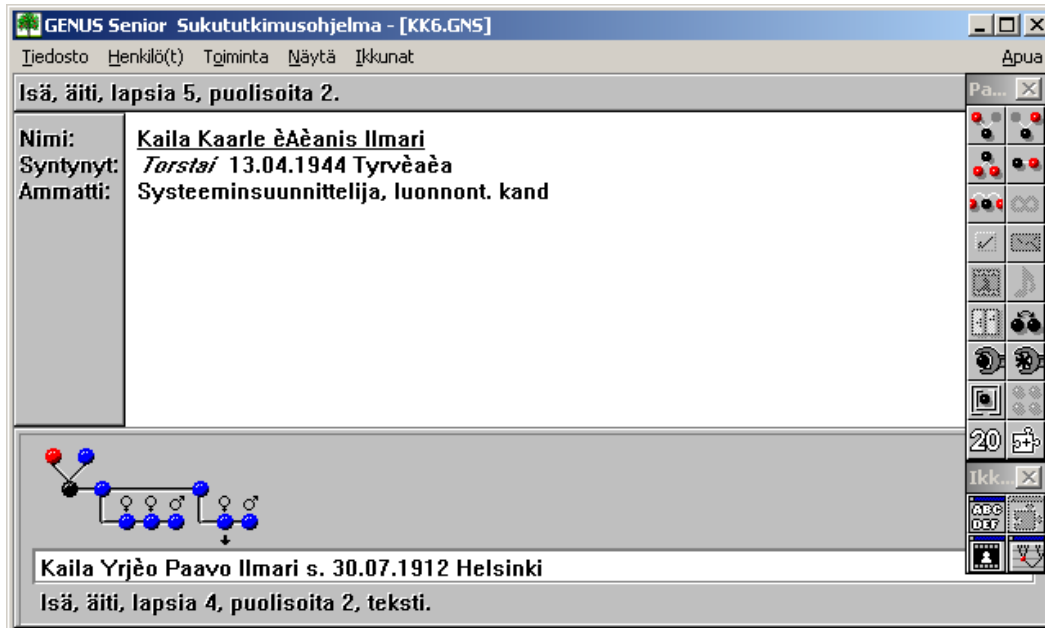
Brother's Keeper-ohjelmalla voi tulostaa kohdan 2.1 mukaisen USA:ssa suositun Register-muotoisen jälkipolviraportin. Esivanhempien tulostusta varten ohjelmassa on joukko graafisia esitysmuotoja.

4.5 Genus

Citius Software, CitiusNet Oy:n myymä Genus- niminen sukututkimusohjelmisto [Gen98] on ollut myynnissä vuodesta 1993 alkaen. Uusin versio Genus Senior 2.21 on vuodelta 1998. Henkilöt syötetään ohjelmaan ns. käsikirjoitusta apuna käyttäen, joka pitää esillä käsittelyn alla olevat henkilöt.

Tiedot talletetaan tiedostoon näyttökohtaisesti. Henkilötietoja varten on oma tiedostonsa. Samoin sukulaisuussuhteita, tekstitietoja ja lähdetietoja varten. Genus esittää kuvassa 10 näkyvällä tavalla käsittelyn alla olevan henkilön perherakenteen. Tämän kuvion avulla voi ohjelmassa vaivattomasti siirtyä perheenjäsenestä toiseen. Useimmissa sukututkimus-ohjelmistoissa on vastaava toiminto, mutta Genus-ohjelmassa tämä on toteutettu tällä tyylikkäällä grafiikalla. Kuvassa esiintyvä epätavallisen näköinen tapa esittää ä- ja ö-kirjaimia johtuu siitä, että ainakaan Genuksen tässä käytetty versio ei osannut tulkita kohdassa 5.5 esitettyä GEDCOM-standardin mukaista ANSEL-merkistöä.

Kuten taulukosta 1 ilmenee on Genus-ohjelma ollut suosittu suomalaisten sukututkijoiden keskuudessa, mutta koska esimerkiksi 32-bittistä versiota ohjelmasta ei ole kehitetty vaikuttaa siltä, että sen kehitys on lopetettu.



Kuva 10. Genus-Senior ohjelman päänäyttö

Genus-ohjelmistosta saa käsityksen, että se käsittelee sukulaisuussuhteita kohdassa 3.1 esitettyä perherakennetta noudattaen.

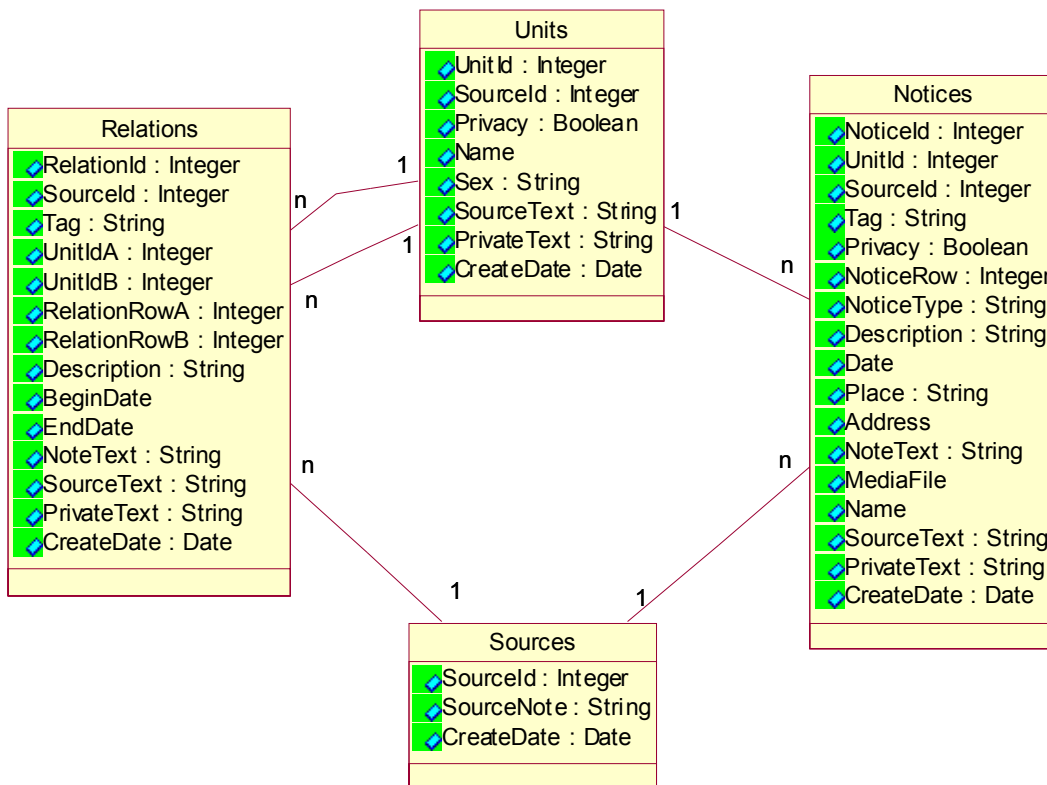
Genus-ohjelmalla voi tulostaa kirjoittimelle kohdan 2.1 mukaisia taulumuotoisia jälkipolviraportteja sekä erinäköisiä luetteloita ja joitakin graafisia taulukkomuotoisia tulosteita.

4.6 Sukuohjelmisto

Sukutietotekniikka ry -niminen yhdistys perustettiin 1984 edistämään tietokoneen käyttöä sukututkimuksen apuvälineenä. Varhaisimmat MS-DOS-versiot Sukuohjelmistosta [Sss00] pohjautuivat ruotsalaisen DIS-yhdistyksen [Dis00] Disgen-ohjelmaan. Ohjelma tehtiin alun perin Ruotsissa suosituille ABC-80-tietokoneelle. Siitä on ollut olemassa myös Commodore-64-versio. Ohjelman viimeinen MS-DOS-versio 3.5 oli myynnissä vuoteen 1994 asti. MS-DOS-version sukutietokanta oli Disgen-ohjelman tapaan jaettu enitään 999 hengen FLK (flock)-tietueisiin. Nämä olivat suorasaantitiedostoja, joiden

avulla jopa isoja tietokantaa oli mahdollista käsitellä pelkillä diskettiasemilla varustetussa tietokoneessa.

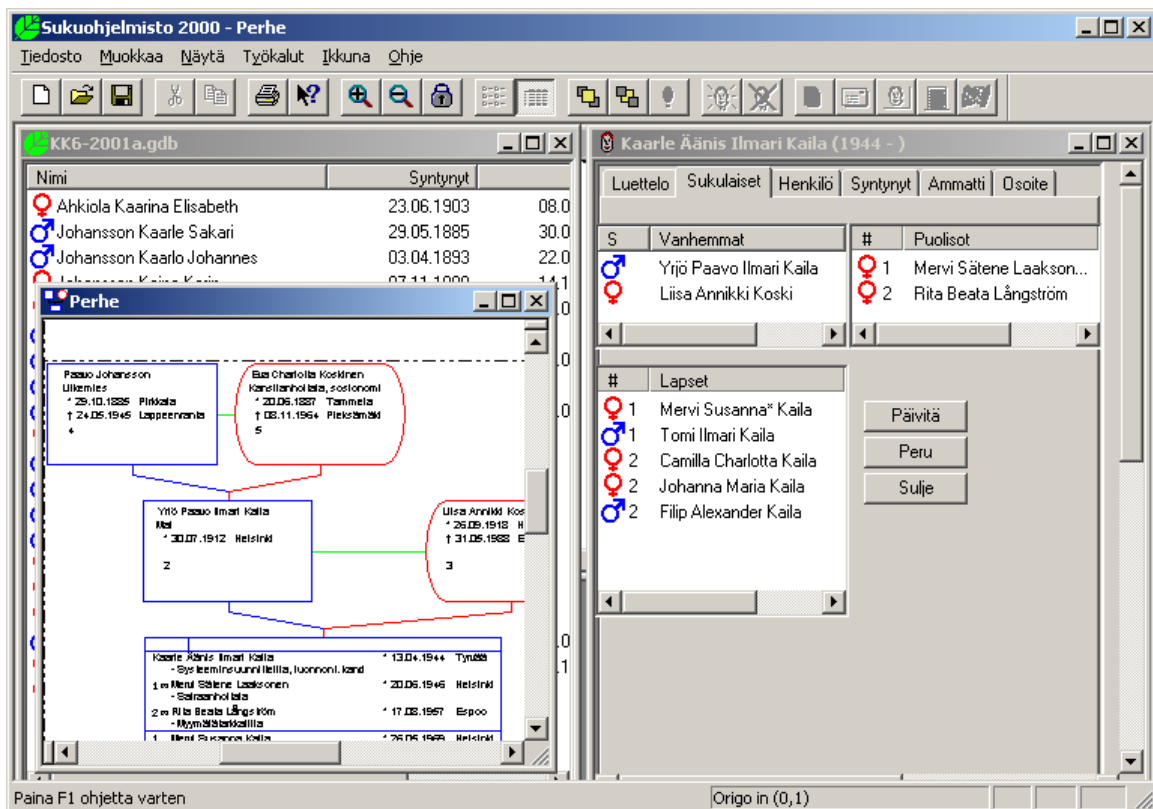
Vuonna 1994 julkaistun ohjelman Windows-versiosta 8.0 alkaen tietokantana on ollut MS-Access-relaatiotietokanta. Tietokanta on kuvattu ohjelman ohjeissa tarkasti, mikä on mahdollistanut tietokannan käytön myös jonkin tietokantasovelluksen, kuten MS-Accessin tai MS-Excelin avulla. Tietokannan tärkeimmät taulut ovat henkilö-, tietojakso- ja relaatiotaulu (kuva 11). Vuoden 2000 keväällä Suomen Sukututkimusseuran tietotekniikkajaos julkaisi ohjelman 32-bittisen version Sukuohjelmisto 2000. Tietokannan taulut ovat tässä versiossa monipuolisemmat ja joitain versiossa 8.0 havaittuja puutteina pidettyjä ominaisuuksia on siinä toteutettu uudella tavalla. Tietokanta on kuvattu tarkasti ja sitä pääsee tietokantaohjelmistojen ja esimerkiksi JDBC-liittymän avulla käsittelemään Java-kielisillä ohjelmilla. Syksyllä 2000 Sukututkimusseura luopui sukututkimusohjelmien tuottamisesta, jolloin Sukuohjelmisto 2000 myynti siirtyi ohjelman tekijän (Kaarle Kailan) tehtäväksi [Kai01].



Kuva 11. Sukuohjelmisto 2000:n tietokannan tärkeimmät rakenteet

Sukuohjelmisto 2000:ssa päivämäärätiedot esitetään tekstimuodossa, eikä tietokannan tuntemassa päivämäärämuodossa. Tämä on yleistä useimmille sukututkimusohjelmille siitä syystä, että tietokoneiden tuntemat päivämääräformaattit vaativat yleisesti ainakin päivän tarkkuuden. Sukututkimustiedoissa ei tätä tarkkuutta aina saada ja tieto on voitava esittää sillä tarkkuudella kuin se on saatavana. Aina ei edes vuoden tarkkuutta saavuteta, joten erityyppiset tarkentimet, kuten ”ennen”, ”jälkeen” jne ovat tarpeen. Sukuohjelmisto 2000 sallii useimmat GEDCOM-standardin mukaiset päivämäärätarkentimet.

Sukuohjelmisto 2000 mahdollistaa yksinkertaisten, mutta laajojenkin kuvassa 12 näkyvien graafisten esitysten tekemisen. Sukuohjelmisto 2000 on toteutettu VisualC++-ohjelmointikielillä alunperin yliopiston eri kurssien harjoitustyönä. Ohjelman raporttien tekemiseen päädyttiin käyttämään tekstinkäsittelyohjelmaa, joksi valittiin MS Word 97 (tai myöhempi). Ohjelmalle on myös tehty Java-kielillä XML-kielisen varmuuskopion vedostus- ja palautusohjelma sekä html-muotoisia raportteja tekevä ohjelmisto.



Kuva 12. Sukuohjelmisto 2000:n tietokanta- ja perheikkuna, henkilöikkunan sukulaiset

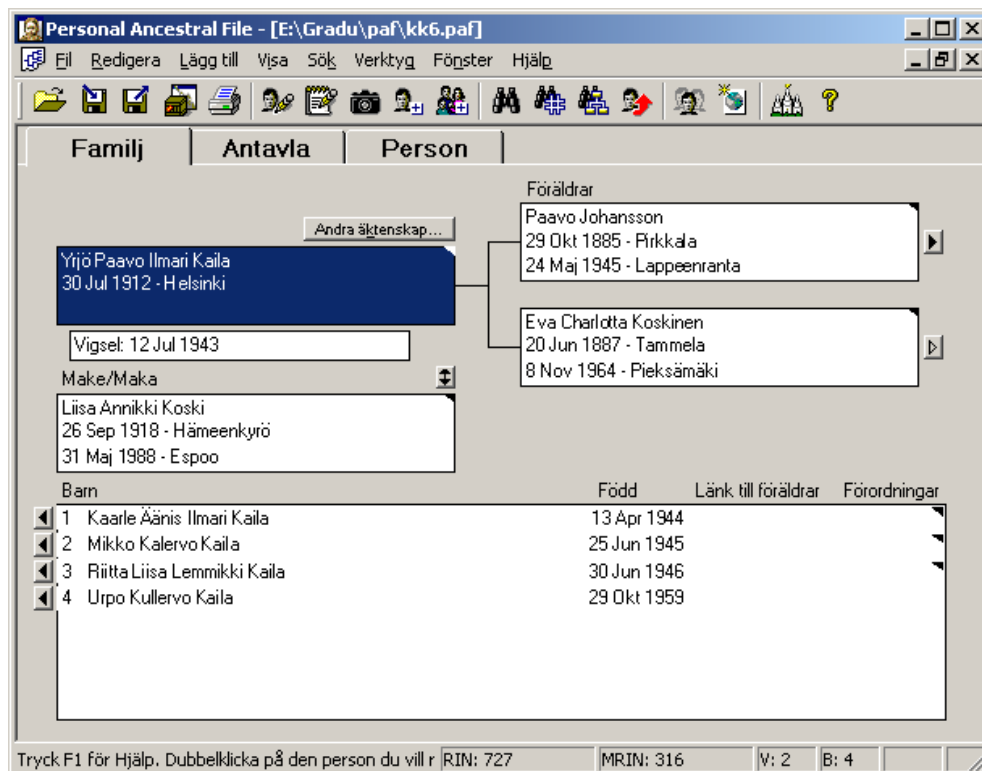
Sukuohjelmisto 2000:n raporttiohjelmien mukana seuraavat ohjelmien lähdekoodit (VBA ja Java/Swing). Java-kielinen raporttiohjelmisto on toteutettu niin, että Java-ohjelmoija voi lisätä siihen omia raporttiohjelmiä siinä olevan liitännän kautta.

Sukuohjelmistossa sukulaisuussuhteiden käsittely on toteutettu kohdan 3.1 mukaista henkilörakennetta käyttäen.

Sukuohjelmisto 2000 Word-raporttiohjelmistolla voidaan tulostaa jälkipolviraportti taulurakenteella. Esipolviraportin Word-ohjelmisto voi tehdä Stradoniz- ja Hager-numerointia käyttäen sekä taulumuotoisena sukukunnittain. Java-kielinen jReport-ohjelmisto soveltuu ensisijassa html-muotoisten tulosteiden tekemiseen. Ohjelmassa on valmiiksi ohjelmoituna edellämainitut Word-tekniikalla tuotetut raportit normaaleina tulosteina mutta myös siten, että kukin taulu tulostuu erilliseen tiedostoon, jolloin tiedostot soveltuvat paremmin internetissä julkaisemista varten.

4.7 Personal Ancestral File

Kuvan 13 esittämä Personal Ancestral File (PAF) on mormonien ilmaisjakelussa oleva



Kuva 13. Personal Ancestral File ohjelman perhenäyttö

sukututkimusohjelma [Fam00]. Tämä ohjelma soveltuu erityisesti heidän omaan käyttöönsä. Ohjelma sisältää sivut, joiden kautta erilaiset mormoniuskonnolle tärkeät tiedot syötetään tietokantaan. Ohjelman näytöistä mainittakoon tietokantanäyttö, joka näyttää tietokannan sisältämät henkilöt, esivanhemmat-näyttö, joka näyttää valitun henkilön 5 esipolvea, sekä perhenäyttö, joka näyttää henkilön perhesuhteen. Miltä tahansa näytöltä pääsee muokkaamaan henkilöiden henkilö- ja perhetietoja.

PAF-ohjelmisto on ollut läheisessä yhteydessä GEDCOM-standardin kanssa. Alun perin GEDCOM olikin PAF-ohjelmiston tuottama siirtotiedosto mormonien tietokantaa (Ancestral File) varten josta on pääteltävissä, että ohjelma käsittelee sukulaisuussuhteita kohdan 3.1 mukaista perherakenteen mukaisesti.

Ohjelma sisältää runsaan valikoiman erilaisia yksinkertaisia raportteja sekä rtf- että html-muotoon tulostettuna.

5 GEDCOM-standardi

Ensimmäisen (ja toistaiseksi ainoan) laajassa käytössä olevan sukututkimustiedon siirtoformaatin kehitti Mormonikirkon perhehistoriaosasto sekä julkaisi sen lokakuussa 1987 nimellä GEDCOM 3.0 (versiot 1.0 ja 2.0 olivat luonnoksia). Tämä merkitsi tietokoneavusteisen sukututkimuksen alkua, sillä sitä ennen olleilla ohjelmistoilla ei yleensä ole ollut pidempiaikaista merkitystä [Pri98]. Tietokoneavusteisen sukututkimuksen ensimmäisen vuosikymmenen hallitsi GEDCOM.

Mormonikirkon perhehistoriaosasto kehitti GEDCOM-standardin luodakseen joustavan ja yhtenäisen muodon sukututkimustiedon välittämistä varten. GEDCOM on akronyymi sanoista **G**enealogical **D**ata **C**ommunication. Sen tarkoitus on edistää sukututkimustiedon siirtoa ohjelmasta toiseen sekä ohjelmistojen kehittämistä suuntaan, joka helpottaa suku-, historian- ja muiden tutkijoiden työtä. Vuonna 1996 julkaistu standardin versio 5.5 [Ged96] on tätä kirjoitettaessa yleisessä käytössä.

GEDCOM-standardi jakautuu GEDCOM-tietoformaattiin (GEDCOM data format) sekä GEDCOM-muotoon (GEDCOM form).

5.1 GEDCOM-tietoformaatti

GEDCOM-tietoformaatti on yleinen tiedon esittämismuoto, jolla on mahdollista esittää mitä tahansa rakenteellista tietoa peräkkäistiedostona. Siinä käsitellään rakenteellisen tiedon syntaksia ja tunnistamista yleisesti ottamatta kantaa, minkä tyyppisestä tiedosta on kysymys. Ajatuksena standardin tekijöillä on täten ollut tarjota tämä rakenne myös muuhun kuin sukututkimustiedon käsittelyyn.

GEDCOM-tiedosto on merkkijonomuodossa oleva tietokanta, joka sisältää keskenään riippuvaisuussuhteessa olevia tietueita. Tietue esitetään joukkona peräkkäisiä, nimiöillä (tag) hierarkkisesti järjestettyjä vaihtuvanmittaisia rivejä. Rivi koostuu aina hierarkian tasoa osoittavasta numerosta, nimiöstä sekä mahdollisesta arvosta. Rivi voi myös sisältää ristiviittaus-tunnisteen tai -osoittimen. Rivi loppuu rivinvaihtoon <cr>- tai <lf>- merkkiin tai mihin tahansa näiden merkkien yhdistelmään. Rivin pituus voi olla enintään 255 merkkiä.

Nimiö, jonka merkitys voi riippua sen sijainnista hierarkiassa, määrittelee rivillä olevan tiedon samalla tavalla kuin tietokannan kentän nimi määrittelee kentän sisältämän tiedon. Rakenne sallii myös uudenlaisten kenttien ottamisen mukaan GEDCOM-tiedostoon aiheuttamatta epäyhteensopivuutta, koska vastaanottava järjestelmä hylkää tiedon, jota se ei ymmärrä.

Hierarkiatason osoittaa rivin alussa oleva tasonumero. Riville alisteisilla riveillä on yhtä korkeampi tasonumero. Hierarkia sallii, että rivillä on alisteisia rivejä, joilla puolestaan voi olla alisteisia rivejä jne. Rivi ja sen alisteiset rivit kuvaavat samaa kokonaisuutta. Uusi tietue aloitetaan rivillä, jonka tasonumero on 0.

Hierarkkisten suhteiden lisäksi GEDCOM määrittelee tietueiden väliset suhteet. Näitä suhteita kuvaa kaksi valinnaista rivin osaa: ristiviittaus-osoitin ja tunnistin. Ristiviittaus-osoitin osoittaa tietueeseen, jonka tunnistimena on vastaava yksiselitteinen ristiviittaus-tunnistin.

GEDCOM-kielioppi sisältää tarkat säännöt GEDCOM-formaatille. Esimerkkinä esitetään yksinkertainen yksilötietue, joka sisältää henkilön syntymä- ja kuolinajan.

```
0 INDI
1 NAME Jaakko Elmeri/Kaila/
```

Sukututkimustiedon käsittely

```
1 SEX M
1 BIRT
2 DATE 12 MAY 1920
1 DEAT
2 DATE 1960
```

Tässä esimerkissä rivi 2 DATE 12 MAY 1920 sisältyy yksilötietueeseen (individual, INDI) BIRT (birth) -riviin esittäen henkilön syntymäpäivää. Toinen DATE sisältyy yksilötietueen kuolinkenttään (DEAT) tietoon. DATE-rivin täydellinen merkitys riippuu siitä, mihin kokonaisuuteen se kuuluu.

GEDCOM-tietoformaattia ei tiettävästi ole sen monipuolisuudesta huolimatta otettu käyttöön kuin sukututkimustiedon käsittelyn yhteydessä. Tämän jälkeen tässä esityksessä tarkoitetaan GEDCOM:illa käytössä olevaa sukujohteiskielioppia (Lineage-Linked Grammar).

Standardin tämä (pää)osa määrittelee ne erityiset nimiö-, arvo- ja osoitinyhdistelmät, joita käytetään perhepohjaisen sukuperään sidotun sukututkimustiedon esittämiseen GEDCOM-formaatissa. Sukuperään sidottu tieto koostuu henkilöistä, jotka ovat sukua keskenään useammassa sukupolvessa.

GEDCOM-formaatti on myös tekijöiden, eli mormonien ainoa hyväksymä tapa välittää uskontokunnan tietokantoihin tietoa.

GEDCOM-tiedostoon kirjoitettavien GEDCOM-rivien järjestyksen määrittelee konteksti ja tasonumero. Tietueelle tai kentälle suoraan alisteisten, samalla tasonumerolla ja eri nimiöllä varustettujen rivien keskinäinen järjestys ei ole merkitsevä. Saman tasonumeron ja saman nimiön omaavien rivien esiintyminen samassa kontekstissa käsitetään määritelmän mukaan siten, että tiedosta on olemassa useita mielipiteitä tai useita arvoja. Järjestyksen merkitys tulkitaan näissä tapauksissa tekijän asettamana etusijana. Ensiksi esiintyvää pidetään etusijalla. Esimerkiksi tutkija luettelee todennäköisimmän syntymäajan ensimmäisenä ja vähiten todennäköisen viimeisenä. Joissakin tapauksissa tosin saman nimiön omaavilla riveillä voidaan esittää esimerkiksi henkilön ammatit kronologisessa järjestyksessä. Eräille rivityypeillä, kuten *tekstitieto* (NOTE), joiden tieto ei aina mahdu yhdelle riville on määritelty jatkorivin käyttömahdollisuus CONT- ja CONC- rivien avulla. Jatkorivin tasonumero on yhtä suurempi kuin varsinaisen tietoa osoittavan rivin tasonumero. CONT-rivi alkaa tekstissä uuden rivin, kun CONC-rivi

jatkaa ilman välimerkkejä. Jatkorivit on katsottava osaksi niiden pääriviä ja luonnollisesti jatkorivien järjestys on merkittävä vaikka nimiöt voivatkin olla toisistaan poikkeavat.

Vastaanottava järjestelmä, joka ei hallitse kuin yhden kentän esiintymisen tulisi vastaanottaa vain ensimmäisen kentän. Suositeltavaa on, että kentät, joita vastaanottava järjestelmä ei hyväksy sijoitetaan tekstikenttään. Esimerkiksi GEDCOM-rakenteessa voidaan esittää henkilön ystäviä tai naapureita <ASSOCIATION_STRUCTURE>-rakenteen avulla jota harvat ohjelmat osaavat käsitellä muuten kuin tekstinä.

5.2 GEDCOMin tietuerakenteet

GEDCOM tiedosto koostuu seuraavista pääosista:

```
0 <<HEADER>> {1:1}
0 <<SUBMISSION_RECORD>> {0:1}
0 <<RECORD>>{1:M}
0 TRLR
```

Aaltosuluissa {} olevat luvut osoittavat elementin esiintymiskerrat. Otsikkotietue (header) sisältää ainakin rivit:

```
0 HEAD
1 SOUR <Lähetettävän järjestelmän tunnus>
1 SUBM @XREF:SUBM@
```

Tämän lisäksi Header-tietue sisältää koko tiedostoa koskevaa tietoa. Lähetettävän järjestelmän tunnus on alun perin rekisteröity järjestelmässä määrättyssä osoitteessa, joka ei liene enää käytössä. Tiedoston tehneen ohjelmiston lyhytnimeä on yleensä tässä käytetty tunnuksena. SUBM-rivin osoittama ristiviittaus osoittaa tiedoston tekijän tiedot sisältävään tietueeseen.

GEDCOM-standardi olettaa, että tiedosto sisältää ANSEL-merkkejä (katso alakohta 5.5) ellei otsikko sisällä riviä:

```
1 CHAR <CHARACTER_SET>
```

jossa sallitut vaihtoehdot ovat [ANSEL|UNICODE|ASCII]. Käytännön toteutuksissa tällä kohdalla on kirjavuutta. Erityisesti monet ohjelmat olettavat, että ellei merkistöä ole määritelty, ei konversiota ANSEL – koodista ohjelman käyttämään merkistöön (kuten Latin-1, IBM-PC) suoriteta.

SUBMISSION_RECORD-tietue liittyy mormonien järjestelmiin eikä sillä ole muuta merkitystä. Se jätetään sukututkimustietojen tapauksessa usein pois. Muut tietueet

Sukututkimustiedon käsittely

sisältävät varsinaisen sukututkimustiedon. Standardi määrittelee taulukko 2:n esittämät tietuetyypit.

| | |
|-------------------|------------------|
| FAM_RECORD | PERHETIETUE |
| INDIVIDUAL_RECORD | YKSILÖTIETUE |
| MULTIMEDIA_RECORD | MULTIMEDIATIETUE |
| NOTE_RECORD | TEKSTITIETUE |
| REPOSITORY_RECORD | ARKISTOTIETUE |
| SOURCE_RECORD | LÄHDETIETUE |
| SUBMITTER_RECORD | LUOVUTTAJATIETUE |

Taulukko 2. GEDCOM-standardin tietuetyypit

Näistä ovat merkittävimmät tiedoston rakenteen kannalta perhetietue, yksilötietue sekä lähdetietue.

Taulukossa 3 on lueteltu *Yksilötietueen* (INDIVIDUAL record) tärkeimmät osat.

| | |
|---|--------------------------------------|
| 0 @XREF:INDI@ INDI | 0 @YKSILÖTUNNUS@ INDI |
| 1 <<PERSONAL_NAME_STRUCTURE>>{0:M} | 1 <<HENKILÖNIMIRAKENNE>>{0:M} |
| 1 SEX <SEX_VALUE>{0:1} | 1 SEX <SUKUPUOLI>{0:1} |
| 1 <<INDIVIDUAL_EVENT_STRUCTURE>>{0:M} | 1 <<YKSILÖN_TAPAHTUMARAKENNE>>{0:M} |
| 1 <<INDIVIDUAL_ATTRIBUTE_STRUCTURE>>{0:M} | 1 <<YKSILÖN_OMINAISUUSRAKENNE>>{0:M} |
| 1 <<CHILD_TO_FAMILY_LINK>>{0:M} | 1 <<LAPSI_PERHESUHDE>>{0:M} |
| 1 <<SPOUSE_TO_FAMILY_LINK>>{0:M} | 1 <<PUOLISO_PERHESUHDE>>{0:M} |
| 1 <<SOURCE_CITATION>>{0:M} | 1 <<LÄHDEVIITTAUS>>{0:M} |
| 1 <<MULTIMEDIA_LINK>>{0:M} | 1 <<MULTIMEDIAVIITTAUS>>{0:M} |
| 1 <<NOTE_STRUCTURE>>{0:M} | 1 <<TEKSTIRAKENNE>>{0:M} |

Taulukko 3. Yksilötietueen rakenne

TUNNUS on yksilötietueen yksilöivä tunnus. Kaikki GEDCOM-tunnukset alkavat kirjaimella ja usein loppuosa on numeerinen. Yleistä on, että yksilötietueiden tunnuksat ovat muotoa I1, I2, I3 jne, perhetietueiden tunnuksat muotoa F1, F2, F3 jne ja lähdetietueiden tunnuksat muotoa S1, S2, S3 jne.

Henkilön nimirakenne koostuu useimmiten pelkästä NAME-rivistä, jossa sukunimi erotetaan muusta nimestä //-merkeillä. Näiden perässä esiintyy vielä USA:ssa yleisessä käytössä oleva jr, sr yms. lyhenteet. Erillistä paikkaa aatelinimissä olevaa etuliitettä varten, kuten von ja af, ei rakenteessa ole. GEDCOM 5.5:ssä on määritelty erikseen nimien osien luettelointi erillisillä riveillä, mutta tätä ominaisuutta ei useimmissa ohjelmissa ole.

Sukupuoli voi olla joko M tai F. Mikäli sukupuolta ei tiedetä, jätetään rivi pois.

Yksilötietueen *tapahtumat* (events) ovat henkilön elämänvaiheita kuvaavia tapahtumia, kuten syntynyt, kastettu, kuollut, adoptoitu (BIRT, CHR, DEAT, ADOP) tai vapaasti määriteltävä tapahtuma (EVEN).

Yksilötietueen *attribuutit* ovat ammatti, oppiarvo, osoite, titteli, kasti jne. Sekä tapahtumatieta että attribuuttia täydennetään TAPAHTUMAN_TARKENNUS-rakenteella.

LAPSI_PERHESUHDE ja PUOLISO_PERHESUHDE rakenteet sisältävät osoittimen perherakenteeseen, jossa henkilö on jäsenenä. Ensiksi mainittua voidaan tarkentaa adoptio- tai kasvatuslapsitiedolla.

MULTIMEDIAVIITTAUS-osa voi GEDCOM 5.5:ssä olla GEDCOM-rakenne, joka sisältää multimediatiedon kuvaavat otsikkotiedot sekä tiedot siitä, missä erillisessä tiedostossa multimediatieto sijaitsee. GEDCOM 5.5:n rakenne sallii multimediatiedon koodaamisen osana GEDCOM-tiedostoa. Tätä mahdollisuutta eivät ohjelmat ole yleensä käyttäneet, ja se on poistettu GEDCOM 5.5-1 ehdotuksesta.

TEKSTIRAKENNE sisältää tekstimuotoista tietoa henkilöstä. NOTE rivi voidaan jatkaa joko CONT tai CONC riveillä mahdollisine lähdetietoineen.

Esimerkin 7 esittämä henkilö on puolisona perherakenteissa F1 ja F2 sekä lapsena perherakenteessa F3. Puolisoiden järjestys määräytyy FAMS rivien järjestyksen mukaan.

```
0 @I1@ INDI
1 NAME Kaarle/Kaila/
1 SEX M
1 BIRT
2 DATE 13 APR 1944
2 PLAC Vammala
1 OCCU Systeemin suunnittelija
1 NOTE Kaarle Kaila kirjoittaa Pro
2 CONC Gradu tutkielmaa Helsingin yliopiston tietojenkäsittelytieteen
2 CONC laitokselle (CONC tulee samalle riville )
2 CONT Hän aloitti opiskelun vuonna 1968 (Tämä tulee uudelle riville)
1 FAMS @F1@
1 FAMS @F2@
1 FAMC @F3@
```

Esimerkki 7. Yksilötietue

Perhetietueen (FAMily record) tärkeimmät osat on esitetty taulukossa 4.

| | |
|-----------------------------------|----------------------------------|
| 0 @<XREF:FAM>@ FAM | 0 @<PERHETUNNUS>@ FAM |
| 1 <<FAMILY_EVENT_STRUCTURE>>{0:M} | 1 <<PERHETAPAHTUMARAKENNE>>{0:M} |
| 1 HUSB @<XREF:INDI>@{0:1} | 1 HUSB @<YKSILÖOSOITIN>@{0:1} |
| 1 WIFE @<XREF:INDI>@{0:1} | 1 WIFE @<YKSILÖOSOITIN>@{0:1} |
| 1 CHIL @<XREF:INDI>@{0:M} | 1 CHIL @<YKSILÖOSOITIN>@{0:M} |
| 1 <<NOTE_STRUCTURE>>{0:M} | 1 <<TEKSTIRAKENNE>>{0:M} |
| 1 <<SOURCE_CITATION>>{0:M} | 1 <<LÄHDEVIITTAUS>>{0:M} |

Taulukko 4. Perhetietue

Yleisimmät perhetapahtumat ovat avioliiton solmimis- ja purkamistapahtumat. Lapset pyritään esittämään syntymäajan mukaan kronologisessa järjestyksessä.

```
0 @F1@ FAM
1 HUSB @I1
1 WIFE @I2
1 MARR
2 DATE 13 APR 1943
2 PLAC Äänislinna
1 DIV
2 DATE 1978
1 CHIL @I3
1 CHIL @I4
1 CHIL @I5
```

Esimerkki on perhetietueesta jossa on kuvattuna isä, äiti, heidän avioliiton solmimis- ja ero- ajankohtansa sekä kolme lasta. Lasten järjestys määräytyy perhetietueen CHIL rivien mukaan.

Sekä henkilö- että perhetietueiden *LÄHDEVIITTAUS* (SOURCE_CITATION) osa voi olla lähdetiedon sisältävä SOUR-rivi mahdollisine jatkoriveineen tai osoitin lähdetietueeseen, jota voi tarkoittaa jopa varsin monitasoinen tarkenninosa.

Sukututkimusten tulosta arvioitaessa on lähteiden merkitys suuri. Sukututkimuksessa tulisi olla lähteet esitetty ainakin kaikista niistä tiedoista, joissa lähettä ei voi päätellä muuten. Usein ollaan sitä mieltä, että kirkonkirjoja ei tarvitse mainita erikseen, mutta erikoisemmat lähteet täytyy mainita. Lähdetietojen perusteella tulisi toisen tutkijan myöhempien tutkimustensa yhteydessä olla mahdollista päästä samaan johtopäätökseen.

5.3 GEDCOM-standardin alirakenteet

GEDCOM-tietue sisältää, kuten edellä on havaittu, rivejä joihin puolestaan voi liittyä alisteisia rivejä. Alisteiset rivit noudattavat standardissa määrättyjä muotoja. Standardissa

on määritelty n 20 erilaista alisteisista riveistä muodostunutta alirakennetta, jotka useimmissa tapauksissa voivat sisältää toisia alirakenteita. Esimerkiksi henkilön ominaisuusrakenne voi sisältää tapahtumarakenteen, joka puolestaan voi sisältää päiväyksen, paikkarakenteen, osoiterakenteen ym. Rakenteisiin on useimmiten liitettävissä ainakin vapaamuotoinen tekstirakenne ja lähderakenne.

Päiväys on liitettävissä useisiin alirakenteisiin. Päiväyksen esittämiseen GEDCOM antaa varsin monipuoliset mahdollisuudet. Monet sukututkimusohjelmat eivät tosin hallitse kuin peruspäiväyksiä. Peruspäiväyksen voi esittää tuntemallaan tarkkuudella esimerkiksi

```
1 DATE 13 APR 1944
1 DATE APR 1944
1 DATE 1944
```

Päiväys kirjoitetaan aina yllä olevan esimerkin antamassa muodossa. Tämä on yksiselitteinen tapa, johon eri kielissä ja eri maissa käytetyt päivämääräformaattit on konvertoitava. Mikäli päiväystä ei tiedetä tällä tarkkuudella, voidaan päiväys kuvata myös päivämäärävälinä tai likimääräisenä päiväyksenä esimerkiksi

```
1 DATE BET 13 APR 1944 AND 15 APR 1944
1 DATE ABT APR 1944
1 DATE BEF MAY 1944
```

Sallitut avainsanat on lueteltu GEDCOM-standardissa. Mikäli lähdetiedossa päiväys on annettu tekstimuodossa, joka ei ole muunnettavissa päiväysmuotoon, esitetään se siten, että kyseinen teksti kirjoitetaan sulkujen sisään.

GEDCOM-standardilla päiväyksiä on mahdollista kirjata laajalta aikaväliltä sekä eri puolilta maailmaa. Vuosiluvun perässä olevalla B.C.-tekstillä voidaan kirjata vuosia ennen Kristuksen syntymää. Päiväykseen voidaan myös liittää kalenterityypin määrittely. Oletuskalenteri on nykyisin käyttämämme gregoriaaninen kalenteri. Ennen gregoriaaniseen kalenteriin siirtymistä oleville päiväyksille voidaan merkitä käytetyn juliaanista kalenteria. Ranskan vallankumouksen jälkeen käytetyille päiväyksille voidaan käyttää ranskalaista kalenteria ja juutalaisia päiväyksiä merkittäessä on mahdollista käyttää heprealaista kalenteria.

5.4 GEDCOM-standardin puutteita

GEDCOM-standardilla on mahdollista kuvata monipuolisesti sukututkimusaineistoa, mutta sen käyttö rajoittuu yleensä sukututkijan tekemien johtopäätösten esittämiseen. Nykyisillä sukututkimusohjelmistoilla talletetaan lähinnä tutkijan tekemiä johtopäätöksiä, eikä kaikkea sitä tutkimusmateriaalia, jonka tutkija on selvittänyt, joten tätä ei yleensä pidetä rajoituksena.

Eräänä GEDCOM-standardin vakavimpana puutteena pidetään sitä, että tietojen järjestystä ei ole standardissa määritelty. Yksilötiedoissa tiedon järjestyksellä on standardin mukaan merkitystä vain, jos tiedot ovat samalla tasolla ja niillä on sama nimiö. Yleensä myös eri nimiön omaavien tietojen järjestyksellä voidaan katsoa olevan merkitystä standardin määrittämisestä huolimatta. Tämä riippuu luonnollisesti käytetystä sukututkimusohjelmasta. Mikäli eri tyyppisille tiedoille on ohjelman tietorakenteissa kiinteä paikka, ei kyseisen tiedon esiintymispaikalla GEDCOM-tiedostossa ole merkitystä.

Rafal Prinke kirjoitti artikkelissaan ”Quo Vadis, GEDCOM?” [Pri96] ongelmasta pian GEDCOM 5.5 -standardin julkaisemisen jälkeen. Hänen mukaansa GEDCOM-standardin mukaisesti ei voi FAM-tietueilla tyylipuhtaasti kuvata perheitä joista toinen vanhemmista on tuntematon. Rafal Prinke esitti ratkaisuksi erillisten SPOU (SPOUse) ja USPO (Unknown SPOuse) -rakenteiden avulla, joilla henkilön sukulaisiin tehtäisiin viittaukset pelkästään INDI-tietueista ja FAM-tietueilla esitettäisiin vain avioliittoon liittyvät tiedot. Näitä GEDCOM-standardin ulkopuolisia rakenteita ei liene otettu käyttöön GEDCOM-yhteensopivissa sovelluksissa.

Seppo Sippu esitti artikkelissaan ”On the Meaning of FAMILies in GEDCOM” [Sip97] miten GEDCOM-standardin mukaisella FAM tietueilla voidaan esittää myös epätavalliset perherakenteet. Tämän mukaan miehen ja naisen sekä heidän yhteiset lapsensa tulisi ensisijaisesti esittää yhdessä FAM-tietueessa riippumatta heidän avioliittojensa määrästä. Mikäli puolisoilla on eri avioliittojensa välissä ollut muita puolisoita saattaa olla tarpeen jakaa puolisoitten avioliitot useampaan FAM-tietueeseen, jotta puolison lasten järjestys saadaan halutuksi. Isän kaikki ne lapset, joiden äiti on tuntematon, voidaan esittää yhdellä FAM-tietueella. On myös mahdollista käyttää useampaa FAM-tietuetta. Jos esimerkiksi

tiedetään, että osalla lapsista on yhteinen äiti ja muiden lasten äidistä ei ole tietoa, voidaan käyttää kahta FAM-tietuetta. Myös lasten iät tai muut seikat, kuten haluttu lasten luettelointijärjestys, saattavat osoittaa sopivan tavan hajottaa lapset katraiksi (FAM-tietueiksi). Tämän esittämiseen ei tarvita Rafal Prinken esittämiä SPOU/USPO nimiöitä. FAM-tietueiden määrä ei millään lailla kiinnitä puolisoiden parisuhteiden todellista määrää. Parisuhteiden määrä kuvataan FAM-tietueiden sisältämällä MARR-tiedoilla. Seppo Sippu osoitti artikkelissaan, että GEDCOM-standardin versioon 5.5 ei ole tarvetta tehdä muutoksia erityyppisten perherakenteiden kuvaamista varten.

5.5 Käytetyt merkistöt

GEDCOM-standardin ANSEL-merkistöllä on mahdollista kuvata Länsi-Euroopan kielialueen nimet ja sanat. USASCII merkistö on ANSEL-merkistön osajoukko. ANSEL-merkistö American National Standard Z39.47-1985 on käytössä myös ”MARC” kirjastojärjestelmässä. GEDCOM-standardi määrittelee käytettäväksi vaihtoehtoisesti UNICODE-merkistön, mutta vain harvat ohjelmat osasivat sitä käyttää ajankohtana, jolloin GEDCOM-standardi julkaistiin. Tämän lisäksi on markkinoilla runsaasti ohjelmistoja, jotka käyttävät merkistöjä, kuten ANSI, WINDOWS, IBM_PC jne, joita standardin mukaisesti ei saa käyttää.

ANSEL-merkistön merkit 0-127 ovat samat kuin 7-bittinen USASCII merkistö. Merkistöllä esitetään Länsi-Euroopassa esitetyt diakriittiset merkinnät erillisellä merkillä (alueelta 128-255) ennen varsinaista kirjainta siten, että esimerkiksi Tyrvää kirjoitetaan Tyrvææa koska ANSEL-merkistön ä-pisteet esitetään Latin-1-merkistön è-merkillä. Osa merkeistä, kuten Æ ja Ð esitetään yhdellä ANSEL-merkillä.

6 XML:n käyttö sukututkimustiedon tallennus- ja siirtoformaattina

XML:n [XML98] käyttö on valtaamassa alaa. Sukututkimustiedon tallennukseen XML sopii erityisen hyvin teknisenä ratkaisuna. Eri tyyppiset tietokanta- ja tiedostoformaattit vanhenevat ohjelmaversioiden ja tallennusvälineiden vanhetessa. Sukututkimustietoa kerätään useimmiten harrastusluonteisesti ja tiedon keräilyajankohtien välillä voi joskus kulua vuosia. Ei ole epätavallista, että tiedon keräilyä ja kokoamista jatkaa tutkijan poika tai muu sukulainen vuosien päästä.

XML:n tulevaisuutta on mahdotonta arvata, mutta dokumentoidun tekstimuotonsa takia voidaan päätellä, että se on puretavissa pitkienkin aikojen kuluttua.

Ongelmana XML:n käytölle sukututkimustiedon tallennusformaattina on tällä hetkellä se, että ei ole vielä olemassa GEDCOM:in tapaista standardia (DTD:tä, skeemaa), jolla sukututkimustiedon tallennus tulisi tehdä.

Ainakin kaksi toteutettua sovellusta sukututkimustiedon tallettamiseksi XML-muotoon on julkaistu. Ensimmäinen julkaistu XML-formaatti oli Michael Kayn esittämä GedML: Genealogical Data in XML [Kay99]. Toinen malli XML:n käytölle on Kaarle Kailan syksyllä 2000 toteuttama XML-muotoinen Sukuohjelmisto 2000-tietokannan varmuuskopio XML-muodossa [Kai01].

XML soveltuu ainakin pienempien tietokantojen tapauksessa myös varsinaiseksi tiedoston tallennusformaatiksi. XML-tiedosto puretaan tiedoston avauksen yhteydessä ohjelman muistirakenteisiin. Tiedon käsittely voidaan toteuttaa joko DOM-kirjasto-luokkien avulla [Dom02] tai ohjelman sisäisillä luokilla, jonne tiedosto on luettu SAX ohjelmistoa [Sax02] käyttäen. Vastaavasti tiedoston talletuksen yhteydessä luodaan uusi XML-tiedosto.

6.1 GedML

XML-standardin suositus 1.0 julkaistiin helmikuussa 1998. (XML-standardin ensimmäinen vedos julkaistiin marraskuussa 1996.) Vuonna 1998 löytyi varsin vähän ohjelmistoja, jotka hyödynsivät XML-tekniikkaa. Englantilainen Michael Kay oli näin ollen aikaisin liikkeellä, kun hän julkaisi huhtikuussa 1998 GedML-ohjelmistonsa. Hän jatkoi ohjelmistonsa kehittämistä ainakin toukokuuhun 1999 asti.

GedML-ohjelmistoa voidaan pitää kokeena, jolla haluttiin osoittaa, kuinka XML-standardi soveltuu sukututkimustiedon käsittelyyn ja erityisesti GEDCOM-muodossa olevan tiedon käsittelyyn.

GedML koostuu Java-kielisestä GEDCOM – XML- sekä XML – GEDCOM-muunnosohjelmistosta sekä GedML-muodon määrittelemän Document Type Declaration (DTD)-tiedostosta.

GedML-formaatti vastaa rakenteeltaan GEDCOM-formaattia. GedML:ssä esiintyy sama tieto samassa järjestyksessä kuin GEDCOM:issa. GedML:n tarkoituksena on tekijällä ilmeisesti ollut tutustua XML:n ohjelmistojen käyttöön. GedML-muodosta saatuja hyötyjä ovat tämän tutustumisen lisäksi arveltu olevan mahdollisuus hyödyntää XML-tiedostoja käsitteleviä ohjelmistoja ja luokkakirjastoja.

GEDCOM tiedosto kuvataan DTD-muodossa lauseella:

```
<!ELEMENT GED ( HEAD, (SUBMITTER)? , ( INDI | FAM | OBJECT | NOTE |  
REPOSITORY | SOURCE | SUBMISSION )* ) >
```

Tämän jälkeen kuvataan tässä esiintyvät elementit ja näiden attribuutit.

GedML-tiedoston rakenne selviää parhaiten kuvan 14 esittämän kahden vanhemman ja yhden lapsen perheen hieman yksinkertaistetun esimerkin avulla.

6.2 Sukuohjelmisto 2000:n varmuuskopio XML-muodossa

Sukuohjelmisto 2000:n varmuuskopion ottaminen XML-muotoiseen tiedostoon otettiin käyttöön syksyllä 2000. Sukuohjelmisto 2000:n apuohjelmalla jBackup voidaan tallettaa Sukuohjelmisto 2000:n tietokanta XML-muotoisena tiedostona suoraan zip-tiedostoon.

Tiedoston formaatti selviää parhaiten kuvan 14 esittämän esimerkin avulla. XML-tiedosto vastaa sisällöltään kuvan 11 esittämää tietokannan rakennetta. Tietokannan tietojen purkaminen myöhemmässä vaiheessa ja mahdollisesti toisenlaisessa laiteympäristössä ei välttämättä onnistu yhtä hyvin MS Access-muotoisesta tietokannasta kuin XML-tiedostosta.

Tiedoston muoto määritellään DTD-tiedostolla. Esimerkkinä esitetään tiedoston juuri-elementin määrittelevä lause:

```
<!ELEMENT genealog (owner?,units,relations,groups,sources,conversions?,  
checkouts?,reports?,views?) >
```

Tämän jälkeen kuvataan tässä esiintyvät elementit ja näiden attribuutit.

| GEDCOM 5.5 | GedML | Sukuohjelmisto 2000 varmuuskopio |
|--|---|--|
| <pre> 0 HEAD 1 SOUR Suku 3 ADDR www.iki.fi/kaila 1 GEDC 2 VERS 5.5 2 FORM LINEAGE-LINKED 1 CHAR ANSEL 0 @I1@ INDI 1 NAME Yrjö /Kaila/ 1 SEX M 1 BIRT 2 DATE 30 DEC 1912 2 PLAC Helsinki 1 OCCU majuri evp 1 FAMS @F1@ 0 @I2@ INDI 1 NAME Liisa /Koski/ 1 SEX F 1 BIRT 2 DATE 26 SEP 1918 2 PLAC Hämeenkyrö 1 DEAT 2 DATE 31 MAY 1988 2 PLAC Espoo 1 OCCU Ompelija 1 FAMS @F1@ 0 @I3@ INDI 1 NAME Kaarle /Kaila/ 1 SEX M 1 BIRT 2 DATE 13 APR 1944 2 PLAC Tyrvää 1 OCCU Suunnittelija 1 FAMS @F1@ 0 @F1@ FAM 1 MARR 2 DATE 1943 2 PLAC Viipuri 1 HUSB @I1@ 1 WIFE @I2@ 1 CHIL @I3@ 0 TRLK </pre> | <pre> <GED> <HEAD> <SOUR>Suku</SOUR> <ADDR>www.iki.fi/kaila </ADDR> </HEAD> <INDI ID="I1"> <NAME>Yrjö <S>Kaila</S></NAME> <SEX>M</SEX> <EVEN EV="BIRT"> <DATE>30 JUL 1953</DATE> <PLAC>Helsinki</PLAC> </EVEN> <EVEN EV="OCCU"> majuri evp</EVEN> <FAMS REF="F1"/> </INDI> <INDI ID="I2"> <NAME>Liisa <S>Koski</S></NAME> <SEX>F</SEX> <EVEN EV="BIRT"> <DATE>26 SEP 1918</DATE> <PLAC>Hämeenkyrö</PLAC> </EVEN> <EVEN EV="OCCU"> Ompelija</EVEN> <FAMS REF="F1"/> </INDI> <INDI ID="I3"> <NAME>Kaarle<S>Kaila</S> </NAME> <SEX>M</SEX> <EVEN EV="BIRT"> <DATE>13 APR 1944</DATE> <PLAC>Tyrvää</PLAC> </EVEN> <EVEN EV="OCCU"> Suunnittelija</EVEN> <FAMS REF="F1"/> </INDI> <FAM ID="F1"> <HUSB REF="I1"/> <WIFE REF="I2"/> <CHIL REF="I3"/> <EVEN EV="MARR"> <DATE>1943</DATE> <PLAC>Viipuri</PLAC> </EVEN> </FAM> </GED> </pre> | <pre> <genealog> <units> <unit unitid="I1" tag="INDI" sex="M" > <name> <givenname>Yrjö</givenname> <surname>Kaila</surname> </name> <notices> <notice tag="BIRT" row="1" > <date>19120730</date> <place>Helsinki</place> </notice> <notice tag="OCCU" row="2" > <description>majuri evp</description> </notice> </notices> </unit> <unit unitid="I2" tag="INDI" sex="F" > <name> <givenname>Liisa</givenname> <surname>Koski</surname> </name> <notices> <notice tag="BIRT" row="2" > <date>19180926</date> <place>Hämeenkyrö</place> </notice> <notice tag="DEAT" row="3" > <date>19880531</date> <place>Espoo</place> </notice> <notice tag="OCCU" row="4" > <description>Ompelija</description> </notice> </notices> </unit> <unit unitid="I3" tag="INDI" sex="M" > <name> <givenname>Kaarle</givenname> <surname>Kaila</surname> </name> <notices> <notice tag="BIRT" row="1" > <date>19440413</date> <place>Tyrvää</place> </notice> <notice tag="OCCU" row="2" > <description>Suunnittelija</description> </notice> </notices> </unit> </units> <relations> <relation unitida="I1" unitidb="I2" tag="MARR" rowa="32" rowb="32" > <begindate>1943</begindate> </relation> <relation unitida="I3" unitidb="I1" tag="CHIL" rowa="1" rowb="64" /> <relation unitida="I3" unitidb="I2" tag="CHIL" rowa="2" rowb="64" /> </relations> </genealog> </pre> |

Kuva 14. Sukutiedon esittäminen XML muodossa

Sukuohjelmisto 2000:n tietosisältö vastaa suurin piirtein vastaavan GEDCOM-tiedoston sisältöä, mutta täysin yhteensopiva se ei ole. jBackup-ohjelmalla tuotettuun XML-tiedostoon talletetaan sitä vastoin kaikki oleellinen tieto tietokannasta. XML-tiedoston nimiöiden pitkät nimet kasvattavat xml-tiedoston kokoa, mutta tiivistyvät hyvin zip-tiedostoon, koska samat nimet esiintyvät tiedostossa jatkuvasti.

Mikäli Sukuohjelmistosta jossain vaiheessa päätetään tehdä versio, joka ei hyödynnä Microsoftin Access-tietokantarakennetta Windows- tai muussa ympäristössä, voi XML-muotoisesta varmuuskopioista olla merkittävää hyötyä. Muuta tekniikkaa hyödyntävän ohjelmiston ei näin tarvitse pystyä lukemaan Access-kantaa tietojen konvertoimiseksi toiseen muotoon.

6.3 GEDCOM/XML

Mormonikirkon tietotekniikkaosasto on ilmoittanut tavoitteena siirtyä käyttämään XML-tekniikkaa tietojen lähettämässä heidän suuriin tietokantoihinsa. Dick Eastman [Eas01] kirjoitti näistä suunnitelmista verkkolehdeksään helmikuussa ja huhtikuussa 2001.

Mormonikirkko on näiden tietojen perusteella kertonut, että he tulevat jatkossa keskittymään omien tarpeittensa ylläpitämiseen, jolloin laajemman sukututkimuksen huomioonottaminen saattaa jäädä vähemmälle huomiolle. Heillä on tavoitteena siirtyä käyttämään tarkemmin määrittelemätöntä GEDCOM/XML-formaattia tietojen vastaanotossa. GEDCOM/XML-mallin parantamisessa he ovat sanoneet käyttävänsä pohjana GEDCOM Future Direction -mallia.

Mahdollisen GEDCOM/XML-formaatin käytöstä sukututkimustiedon yleisenä standardina ei ole suunnitelmia. Nämä mormonikirkon ilmoitukset onkin nähtävissä lähinnä siten, että myös heillä on tarkoitus hyödyntää XML-tekniikkaa jatkossa.

6.4 XSL-tekniikan hyödyntäminen

XML tekniikan yhteyteen on kehitetty useita aputekniikoita. Näistä on edellisissä kappaleissa hieman käsitelty XML-tiedoston määrittely DTD:n avulla. Kehitteillä olevan Xschema-tekniikan [Xsc02] avulla voidaan tiedosto määritellä tarkemmin kuin mitä DTD:n avulla on mahdollista tehdä.

Myös XML-tiedoston esittämistä varten on olemassa joukko tekniikoita. Tässä esitetään yksinkertaisen esimerkin avulla XSL [Xsl99].

XML-tiedosto on varsinaisesti tarkoitettu ohjelmistolla käsiteltäväksi olkoonkin, että sen lukeminen (selkeän) tekstimuotonsa takia onnistuu sellaisenaankin. XML-tiedosto voi sisältää runsaasti tietoa ja käyttäjän haluaman oleellisen tiedon löytäminen voi sen takia olla hankalaa. XSL on eräs tekniikka, jolla voidaan esittää XML-tiedosto halutussa muodossa. Kuvan 15 esittämän esimerkin XSL-tiedosto esittää Sukuohjelmisto 2000:n XML-muotoisen varmuuskopion niin, että vain henkilöiden nimet, syntymäajat ja paikkakunnat luetellaan näkyviin.

| | |
|---|---|
| <pre> <xsl:stylesheet xmlns:xsl="http://www.w3.org/TR/WD- xsl"> <xsl:template match="/"> <html> <body> <xsl:apply-templates/> </body> </html> </xsl:template> <xsl:template match="genealog"> <h1>Genealog</h1> <xsl:apply-templates/> </xsl:template> <xsl:template match="units"> <h2>Persons</h2> <xsl:apply-templates/> </xsl:template> </pre> | <pre> <xsl:template match="unit"> <xsl:apply-templates/> </xsl:template> <xsl:template match="name"> <xsl:value-of select="."/> <xsl:apply-templates/> </xsl:template> <xsl:template match="notices"> <xsl:apply-templates select="notice[@tag='BIRT']"/> </xsl:template> <xsl:template match="notice"> Born: <xsl:value-of select="."/> </xsl:template> </xsl:stylesheet> </pre> |
|---|---|

Kuva 15. XSL-esimerkki

Esimerkissä luodaan html-tiedoston kehys (<html><body>). Genealog-elementti aiheuttaa otsikon kirjoittamisen (Genealog). Units-elementti aiheuttaa aliotsikon (Persons) ja unit-elementistä kirjoitetaan uudelle riville henkilön nimi, syntymäaika ja paikka

Kuvan 14 esittämä varmuuskopio XML-muodossa näkyy käyttäjälle IE5-selaimella käyttäen kuvan 15 mukaista XSL-tiedostoa kuvan 16 mukaisesti.

Genealog

Persons

Yrjö Kaila Born: 19120730 Helsinki

Liisa Koski Born: 19180926 Hämeenkyrö

Kaarle Kaila Born: 19440413 Tyrvää

Kuva 16. XML-tiedosto esitetty XSL-tyylitiedostoa käyttäen

7 Sukututkimuspäätelyn mallintaminen

GENTECH-tietomalliprojekti [Gdm00] syntyi GENTECH:in Lexicon-nimisen projektin jatkeena. Sen tarkoituksena on sukututkimustiedon määrittelyminen, jotta tiedon vaihtaminen eri sukututkijoiden kesken olisi helpompaa. Pyrittäisiin määrittelemään tietoa jonkin aikaa ryhmä totesi, että sukututkimustiedon määrittelyminen ottamatta huomioon eri käyttöympäristöjä on vaikeata, koska eri käyttäjät käsittävät samoilla termeillä toisistaan poikkeavia asioita. Projektiryhmä päätyi tulokseen, että sukututkimustiedon määrittelyminen loogisen tietomallin yhteydessä on lopputuloksen kannalta edullisempi tapa edetä.

Projektiryhmä korostaa, että kysymys on vain sukututkimustiedon määrittelemisestä. Ryhmä ei suunnittele sen varaan toteutettavaa ohjelmistoa. Tätä mallia suunniteltaessa ei ole huomioitu nykyisten ohjelmistojen tai laitteistojen asettamia rajoituksia. Ryhmä pitää toivottavana, että tämän mallin pohjalta kehitettäisiin ohjelmistoja, vaikka tämä ei ollutkaan projektin tavoite. Mallin alaotsikon mukaan tavoitteena oli aikaansaada *kattava tietomalli sukututkimusta ja sen tulosten analysointia varten*.

Projektiryhmä kehitti loogisen tietomallin, eikä fyysistä mallia. Looginen malli kuvaa erilaisten sukututkimustietojen keskinäiset suhteet. Kun suunnittelija ryhtyy tätä mallia hyödyntämään, voi suunnittelija luodessaan tämän pohjalta fyysisen mallin muokata sitä käyttäen tunnettuja tarkkoja muunnoksia. Tyypillisiä syitä muunnosten tekemiseen on ohjelmiston monimutkaisuuden vähentäminen tai sen tehokkuuden parantaminen.

7.1 Tietomallin ”luonteenpiirteet”

Projektiryhmän tarkoituksena oli luoda tietomalli, joka täyttää seuraavat neljä peruseriaa.

1. Mallin tarkoituksena on tukea sukututkimuksen tutkimusprosessia.
2. Kutakin sukututkimustietoelementtiä varten on yksi ja vain yksi paikka, ja kutakin sukututkimustiedon osaa varten mallissa on olemassa paikka.
3. Jokaisen tutkijan ei tarvitse tuottaa kaikkea tietoa, jota perusteellinen tutkimusprosessin noudattaminen tuottaa.
4. Varsinaiset tietomalliin pohjautuvat ohjelmistojärjestelmät tulisi tehdä niin, että ne opettavat, kannustavat, muistuttavat ja neuvovat käyttäjää seuraamaan tutkimusprosessia. Näin käyttäjän on mahdollista luoda korkealuokkaista sukututkimusta, jonka välittäminen on mahdollista myös muille.

Tietomalli kuvaa käsitystämme (mahdollisuuksien puitteissa) kaikesta sukututkimuksen tiedosta ja pyrkii olemaan tässä täysin kattava. Se on tarkoitettu helpottamaan sukututkimustiedon käsitteiden ymmärtämistä sukututkimusyhteisössä. Vaikka malli on luotu systeeminsuunnittelumenetelmiä käyttäen, sitä ei luotu minkään erityisen ohjelmiston käyttöä silmälläpitäen. Se on laajennettavissa kattamaan kaikenlaisia sukututkimustietoa lisäämällä taulukoihin rivejä sellaista tietoa varten, joita ei huomioitu alkuperäisessä mallissa. Siinä on mahdollisimman vähän ”kovakoodattuna”. Näin malliin voidaan liittää tietoa, jota ei ole osattu mallin luonnin yhteydessä huomioida, jos tieto on mallin tuntemaa tyyppiä. Sen ei pidä millään tavalla vaatia, että sukututkija pakottaa tietoa sopimattomiin kenttiin vain siksi, että mallin suunnittelija ei osannut tätä huomioida. Sitä tulee laajentaa lisäämällä tietotyyppejä, joita projektiryhmä ei ole osannut huomioida. Se saattaa edesauttaa uusien sukututkimusohjelmistojen luontia, mutta malli on riippumaton mistään toteutuksesta, eikä rajoita tulevia ohjelman kehittäjiä ohjelmointikielen tai käytettävän laitteiston valinnassa. Käytetty relaatiomalli on rakenne, jonka kautta tiedon ymmärtäminen on mahdollista. Siinä on mahdollista tallentaa sukututkimusjohtopäätösten takana olevia perusteluja ja näihin liittyviä lähdetietoja sukututkimustiedon yhteydessä. Mikäli sukututkimukselliset johtopäätökset myöhemmin todetaan pätemättömiksi, malli sallii tutkijan korjata päätelmät tekemällä korjaava päätelmä eikä pelkästään purkamalla alkuperäistä johtopäätöstä. Se ei sinänsä pakota tutkijaa tekemään korjausta, jos hän mieluummin purkaa päätelmän. Se soveltuu sekä ammatikseen sukututkimusta tekeväille että vasta-alkajalle. Vasta-alkaja voi tallettaa

suoraan johtopäätöksiä ilman todisteita, kuten nykyisissä sukututkimusohjelmissa yleensä tehdään. Se estää muiden sukututkijoiden tiedon yhdistämisen tutkijan oman tiedon kanssa umpimähkään. Siinä on mahdollista liittää elektronista tietoa tutkijan omiin tietoihin kuten myös tuoda tietoa perinteisistä lähteistä. Tiedon tuomisen yhteydessä malli liittää aina tietoon sen lähteet niin, että kaikesta tiedosta on saatavana tieto siitä mistä tieto on saatu. Siinä voidaan esittää tiedon kulku todistuskappaleista johtopäätöksiin analyysien ja muunnosten kautta sekä alustavien johtopäätösten käyttö edistyneempien johtopäätösten tekemiseksi.

7.2 Mallin suhde GEDCOM-standardiin

GEDCOM on sukututkimustiedon siirtostandardi eri sovellusten välille. Tämä standardi on kehittynyt vuosien varrella ja siinä on tehty määrättyjä oletuksia sukututkimustietojen välisistä suhteista.

GENTECH-mallilla puolestaan kuvataan sukututkimusprosessia ja sen tärkein tehtävä on kuvata ymmärtämäämme sukututkimuksen tutkimus- ja analysointiprosessia.

Tätä mallia ei näin ollen voi verrata GEDCOM:iin koska nämä kuvaavat kahta erilaista rakennetta. GENTECH-tietomallia voi toisaalta verrata ”GEDCOM (Future Direction)”-ehdotuksen mukana olevaan tietomalliin vaikka mallien tekemisessä käytetyt menetelmät ovat varsin erilaiset. Kummallakin mallilla on monta yhteistä ominaisuutta, mutta ne myös poikkeavat toisistaan monella tavalla.

7.3 Sukututkimusprosessin eteneminen

Onnistunut sukututkimuksen tietomalli edellyttää, että sekä mallin tekijät että sen käyttäjät ovat samaa mieltä siitä, mistä sukututkimustiedon keräämisessä ja analysoinnissa on kysymys. Mallin suunnittelussa on katsottu, että sukututkimusprosessi etenee määrättyllä tavalla. Tutkimuksen suunnittelu aloitetaan etsimällä sukutiedoista epäjohdonmukaisuuksia, heikkoja lähdetietoja ym, valitaan mihin ryhdytään seuraavaksi sekä jakamalla tutkimus osakokonaisuuksiin. Seuraavaksi etsitään lähteitä arkistoista. Lähteiden etsimisessä käytetään arkistoja joista valitaan lähteet. Näistä etsitään viittaukset etsittyihin tapahtumiin jonka jälkeen tutkimussuunnitelmaa tarkennetaan ja tarvittaessa tilataan lähteet. Lähteistä etsitään dokumentteja käyttäen tarpeen mukaan

soveltuvia hakuehtoja, jonka jälkeen kirjataan löydetty tekstit halutuista osista. Löydetty dokumentti analysoidaan. Tarvittaessa siitä hankitaan kopio, tehdyt havainnot kirjataan ja kaikki tapaukseen liittyvä etsitään. Työn tilannetiedot päivitetään. Lainatut sitaatit tallennetaan. Lainatut kuvat tallennetaan esimerkiksi skannamalla tai valokuvaamalla. Tekstit käännetään nykykielille. Tekstistä poimitaan sopivia paloja joihin lisätään tarvittaessa huomautuksia. Löydettyjen tietojen perusteella luodaan todistuslausuntoja, joissa kirjataan todistelut kullekin johtopäätökselle, merkitään mitkä viitteet kuvaavat samaa henkilöä tai tapahtumaa. Analysoidaan todistelut ja tehdään johtopäätökset missä analysoidaan tiedot vertaamalla niitä aikaisempiin sukututkimustietoihin, luomalla uusia johtopäätöksiä ja yhdistämällä uudet päätelmät vastaaviin päätelmiin, jotka tukevat niitä. Lopuksi lisätään muiden tutkijoiden sukututkimustietoja

Edellä oleva on GENTECH-mallin tekijöiden näkemys sukututkimusprosessin vaiheista. Nämä askeleet poikkeavat Suomessa ja Euroopassa käytetystä, mutta pääpiirteittäin lienee sukututkimus USA:ssa samantapaista kuin täällä Suomessa. Suomen Sukututkimusseuran toiminnanjohtajan Leif Metherin mukaan ero johtuu siitä, että lähteet ovat Euroopassa erilaisia kuin Yhdysvalloissa, joissa lähteiden avulla usein päästään vain 100-150 vuotta taaksepäin. Suomessa lähteet riittävät aina 1500-luvulle saakka (joissakin Euroopan maissa 1300-luvulle tai jopa sitäkin kauemmaksi).

Sukututkija, löydettyään etsimänsä dokumentin analysoi sen sisällön, kirjaa tietonsa jonkinlaiseen lokiin ja kirjoittaa otteen muistiin ja tekee tästä yksinkertaisen päätelmän, kuten että Jaakko Matinpoika syntyi ennen vuotta 1762. Perusteellinen tutkimus edellyttää sukututkijalta näitä pieniä askeleita ja näiden yhdistämistä tarkoituksenmukaisesti hyvin perusteltuihin johtopäätöksiin. Kokenut tutkija voi käsitellä tätä usealla eri tavalla. Vasta-alkajalta voi jäädä nämä viittaukset kirjaamatta ja hän joutuu ehkä tästä syystä uusimaan tutkimuksensa. Sukututkimuksen tietomallin täytyy kuitenkin tukea tätä kattavaa tutkimusprosessia.

Tämä prosessi, jossa pienistä todisteista tehdään pieniä johtopäätöksiä, on tämän tietomallin ydin ja ehkä merkittävin asia, jonka malli antaa sukututkimustiedon ymmärtämiseksi.

7.4 Sukututkimustiedon yhtenäisteoria

On tärkeää ymmärtää, että sukututkimuksen perustiedot ovat pohjimmiltaan varsin samanlaisia. Vaikka sukututkijat käsittelevät hyvin suurta joukkoa erilaisia lähdetietoja on tämän tietomallin eräs keskeinen ajatus se, että *kaikki* sukututkimustieto voidaan jakaa joukkoon lyhyitä muodollisia sukututkimuslauseita. Nämä lauseet esitetään tässä seuraavia lyhenteitä käyttäen.

P1 esittää henkilöä numero 1. P2 esittää henkilöä numero 2. Käytännössä sekä henkilö 1 ja henkilö 2 ovat osoittimia näitä henkilöitä kuvaaviin tietueisiin eikä esim henkilön nimiin. Tietomallia kuvaavassa kappaleessa tullaan esittelemään PERSONA-olio, joka sisältää osoittimen varsinaiseen nimeen.

Termi ”paikka” viittaa joko itse paikan nimeen tai jos paikankäsittelyohjelmisto on olemassa tarkoittaa ”paikka” viittausta sitä kuvaavaan olioon.

”Päivämääräväli” sallii tarkkojen päiväysten sekä vähemmän tarkkojen päiväysten käytön kuten ”noin 1734”. Myös päivämäärävalit kuten ”välillä 1812-1815” ovat sallittuja.

”Tapahtumat” ovat henkilön elämän aikana tapahtuneita asioita, kuten henkilön syntymä, kuolema, ristiäiset, avioliiton solmiminen jne.

Termillä ”suhde” kuvataan kahden henkilön välisiä sukulaisuustietoja. Voimme esimerkiksi sanoa, että P1 on P2:n isä.

”Ominaisuus” kuvaa henkilöä. Se voi olla ammatti, oppiarvo tai muu tieto kuten esimerkiksi asuinpaikka. Ominaisuus kuvaa henkilöä ja se on yleensä liitettävissä henkilöön suhteellisen pitkän ajanjakson aikana.

Mallin lähtökohtana on, että *mikä tahansa* sukututkimustieto voidaan esittää näillä rajoitetuilla lausetyypeillä. Tämä ei sulje pois sitä tosiasiaa, että jotkut tiedot, kuten esimerkiksi maa-alueiden rajat ovat helpommin ymmärrettävissä, jos ne esitetään esimerkiksi kartan muodossa.

Lexicon-projektiryhmä päätyi siihen, että sukututkimustiedot voidaan esittää seuraavilla lausetyypeillä:

1. Tyyppi 1: lauseet sukulaisuuksista

Sukulaisuuslause on seuraavaa muotoa:

P1 / Sukulaisuussuhde / P2 / Päivämääräväli / Paikka

Esimerkki sukulaisuussuhdelauseesta on:

1234 (Mikko Simonpoika) / isä henkilölle / 2143 (Anna Mikontytär) / 1.12.1727 /
Rakkilassa

2. Tyyppi 2: lauseet tapahtumista

Tapahtumalause on seuraavaa muotoa:

P1 / Tapahtuma / Päivämääräväli / Paikka

Esimerkki sukulaisuussuhteesta on:

1234 (Mikko Simonpoika) / vihittiin / 4.12.1726 / Marttilassa

3. Tyyppi 3: lauseet ominaisuuksista

Ominaisuudet lause on seuraavaa muotoa:

P1 / Ominaisuus / Päivämääräväli / Paikka

Esimerkki ominaisuudet lauseesta on:

1234(Mikko Simonpoika) / oli isäntänä / 1722 / Rakkilassa

4. Tyyppi 4: järjestystä osoittavat lauseet

Neljännellä lausetyypillä esitetään tietojen keskinäistä järjestystä ja tällä lausetyypillä on seuraavat yleiset muodot:

Tapahtuma 1 / Järjestys / Tapahtuma 2

tai

Ominaisuus 1 / Järjestys / Ominaisuus 2

tai

Nimi 1 / Järjestys / Nimi 2

Esimerkkinä järjestyslauseesta on:

Oli ratsumiehenä / oli ennen / talonpoikana

5. Tyyppi 5: Asema ryhmässä

Tällä lausetyypillä esitetään minkä tahansa sukututkimustiedon asemaa / järjestysnumeroa ryhmässä. Lause on muotoa:

X / Asema ryhmässä / Ryhmä

tai

X / Asema ryhmässä / Ryhmä / Päivämääräväli / Paikka

Tyypit 1 – 3 ovat muodoltaan melko samantapaisia. Kaikilla on yhteisenä piirteenä päivämääräväli- ja paikkatieto. Toinen tapa tarkastella tapahtumia ja ominaisuuksia on liittää niihin arvo. Syntymä on useimmissa tapauksissa pelkkä syntymä, mutta siihen saatetaan liittää arvoksi ”keisarinleikkauksella”. Toisaalta ammatti-ominaisuus on arvoton ilman arvoa kuten ratsumies, talonpoika tai pappi. Arvon käsite osoittautuu hyödylliseksi myöhemmin, kun käsittelemme varsinaista tietomallia.

Helposti tulee mieleen, onko tyyppin 2 esimerkissä tapahtunut virhe. Eikö avioliiton kuvauksen pitäisi olla tyyppiä 1. Siis:

P1(Mikko Simonpoika) / vihittiin / P2(Liisa Heikintytär) / 4.12.1726 / Marttilassa

Kummatkin esitystavat ovat oikeita. Mikäli tutkija ei tiedä kenen kanssa Mikko Simonpoika vihittiin hän voi käyttää tyyppiä 2.

Projektiryhmä muodosti lopulta yllälueteltuja tyypejä käyttäen menettelytavan, jonka he nimittivät *superlausetyypiksi*.

Superlausetyyppi kytkee toisiinsa kaksi perustyyppiä kuvan 17 esittämällä tavoilla.

Kaaviosta puuttuu tyyppin 1 lause. Tämä johtuu siitä, että lopullisessa muodossaan tyyppin

| | | <i>Subjekti 2</i> | | | |
|-------------------|------------|-------------------|-----------|-------|------------|
| | | Henkilö | Tapahtuma | Ryhmä | Ominaisuus |
| <i>Subjekti 1</i> | Henkilö | | ✓ | ✓ | ✓ |
| | Tapahtuma | | ✓ | ✓ | |
| | Ryhmä | ✓ | | ✓ | |
| | Ominaisuus | | | ✓ | |

Kuva 17. Superlausetyypin yleisimmät yhdisteet

1 lause korvattiin voimakkaammalla, mutta vaikeammin ymmärrettävällä tavalla. Sukulaiset esitetään yhdistämällä heidät ryhmän kanssa.

7.5 Tietomallille asetetut vaatimukset

Projektiryhmän päätarkoituksena oli luoda tietomalli, joka kuvaa sukututkimustiedon sisäiset suhteet. Lisäksi määrätyt genealogiseen tutkimukseen liittyvät vaatimukset vaikuttavat eri tietoelementtien määrittelyyn.

Tietomallilta odotetaan, että siinä tulee voida kytkeä kaikki oletukset suoraan lähteistä ja sitaateista koostuvaan todistemateriaaliin. Täytyy voida olettaa, että kukin tiedonjyvän kohdistuu eri henkilöön, vaikka nämä olisivat samannimisiä, kunnes toisin on todistettu. Mallin tulee sallia vasta-alkajan sekä kokeneen tutkijan tarpeisiin tehdyn järjestelmän toteutus. Käytännössä tämä tarkoittaa sitä, että ohjelmisto lisää puuttuvat yhdistävät linkit automaattisesti esimerkiksi huomautuksilla ”Ei käytössä”.

Nimien tallennuksesta vaaditaan mm, että nimet tulee voida tallettaa täsmällisesti kuten ne esiintyvät lähteessä. Nimien tulee olla tutkijan käytettävissä yhtenä kokonaisuutena tietona. Kukin erilainen nimi esitetään erillisenä Persona-ID-kenttänä PERSONA-oliossa. Todellinen henkilö koostuu useasta personaoliosta, jotka kuvaavat henkilön nimeä hänen elämänsä aikana.

Paikannimille asetettavia vaatimuksia ovat, että ne tulee voida tallettaa kokonaan sellaisena, kun ne esiintyvät lähteessä. Kukin paikannimi tai sen variaatio on erillinen Place-ID. Lisäksi paikannimessä tulee sallia tarkennuksia kuten ”lähellä”.

Päiväyksiin on liitetty mm vaatimus, että päiväys tulee voida tallettaa tarkasti sellaisena kun se esiintyy lähteessä. Noudatettava kalenteri on määriteltävä (esim Gregorian-Ruotsi). Epätarkat päiväykset (noin, arvioitu) ovat sallittuja. Päivämäärävalit ja suhteelliset päiväykset ovat sallittuja. Päivämäärätieto tulisi tallettaa myös siten, että sitä on mahdollista käyttää myös lajittelukenttänä (esim muodossa 19440413). Mikäli varsinainen järjestelmään talletettu päiväys ei ole tässä muodossa tulee tämä lajittelukelpoinen päiväys myös tallettaa sillä tarkkuudella joka on tiedossa. Päiväykset on voitava tutkijan valinnan mukaan esittää joko kuten ne on tallennettu tai yhtenäisen muunnetun muodon mukaisesti (kuten 13.4.1944 tai 1944-04-13).

Tietomallin tulee sallia kaiken todiste- ja päättelytiedon jäljittämisen siihen tutkijaan, joka sen talletti. Mallin tulee mahdollistaa kaiken ulkoa tuodun tiedon jäljittämisen muihin tutkijoihin ja mallin tulee tukea erilaisia järjestelmiä alkaen yhtä sukua tutkivasta yhden käyttäjän järjestelmästä usean tutkijan järjestelmään jotka tutkivat useita sukuja.

7.6 GENTECHin sukututkimuksen tietomalli

Looginen tietomalli on mahdollista jakaa osiin monella eri tavalla. Tärkein osa on päättelytietojen suhteita kuvaava osa kuten henkilön syntymätieto ja todistetieto, joka johti varsinaisen tiedon perustana olevaan päätelmään. Todistetietojen talletus ja käsittely on merkittävin ero tämän tietomallin ja useimpien tämän hetken sukututkimus-ohjelmistojen käytännön tietomallien välillä.

Hallinnollinen osamalli ohitetaan tässä luettelolla sen olioista. Osamallissa kuvataan kyseessä olevan tutkijan ja hänen projektinsa välisiä suhteita. Malli liittyy todiste-osamalliin SEARCH-olion kautta. SURETY-SCHEME-PART kuvaa SURETY-SCHEME olion osaa. SURETY-SCHEME kuvaa tutkijan arvioimaa projektin todenpitävyyttä. RESEARCHER kuvaa sukututkijan. RESEARCHER-PROJECT liittää sukututkijan kyseiseen projektiin. PROJECT olio kuvaa käsiteltävää projektia. Kuhunkin projektiin liittyy todenpitävyyttä kuvaava SURETY-SCHEME. RESEARCH-OBJECTIVE-ACTIVITY:llä liitetään projektiin loput hallinnollisesta osamallista. RESEARCH-OBJECTIVE olioon tutkija voi kirjata tutkimuksensa tavoitteet. ACTIVITY oliolla liitetään RESEARCH-OBJECTIVE joko hallinnolliseen tehtävään tai etsintään (SEARCH). ADMINISTRATIVE-TASK kuvaa tutkimukseen liittyvää hallinnollista toimenpidettä. SEARCH kuvaa suunniteltua tai jo toteutettua etsintää. Etsintä voi tapahtua jossain nimeltä mainitussa arkistossa, käyttäen lähteenä (SOURCE) esimerkiksi kirkonkirjaa ja koostuu jonkun henkilön nimen, paikan tai muun tiedon tutkimisesta. SEARCH on tämän hallinnollisen ketjun päätepiste, joka yhdistää hallinnollisen osamallin todisteet-osamalliin.

Hallinnolliseen malliin lasketaan vielä lähteisiin liittyvät oliot SOURCE-GROUP jolla yhdistetään yhteenkuuluvia lähteitä ja SOURCE-GROUP-SOURCE liittää SOURCE-GROUP olion SOURCE olioon todiste-osamallissa.

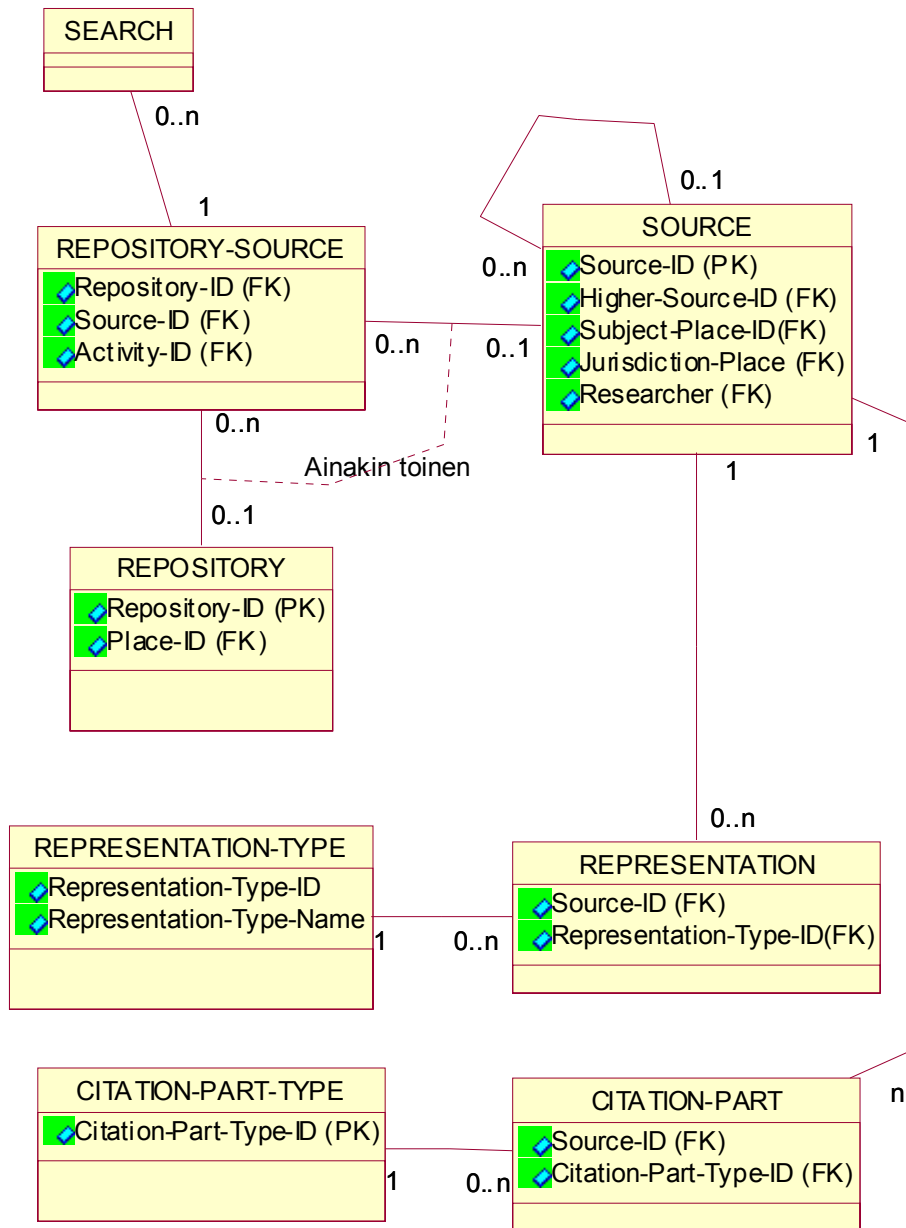
Hallinnollisen osamallin hyödyllisyyttä yksityisen sukututkijan tarpeisiin on epäilty mutta nähty sen mahdollisuudet sellaisen ammattimaisen sukututkijan käytössä, joka tutkii useita sukuja samalta alueelta [Pri98].

Todisteet-osamalli sisältää lähteitä kuvaavat tiedot, eli sukututkimuksen todistemateriaalin.

SEARCH hallinnollisessa mallissa kytkeytyy kuvan 18 esittämään todisteet-osamalliin aputaulun REPOSITORY-SOURCE avulla, joka kytkee yhteen sekä REPOSITORY- että SOURCE- taulun. REPOSITORY sisältää arkistoa kuvaavaa tietoa, kuten arkistolaitoksen nimi ja osoite. Arkistoksi luokitellaan kansallisarkiston lisäksi myös mm yksityishenkilöt, joilta sukututkimustieto on saatu. SOURCE kuvaa tutkijan käyttämää lähdettä.

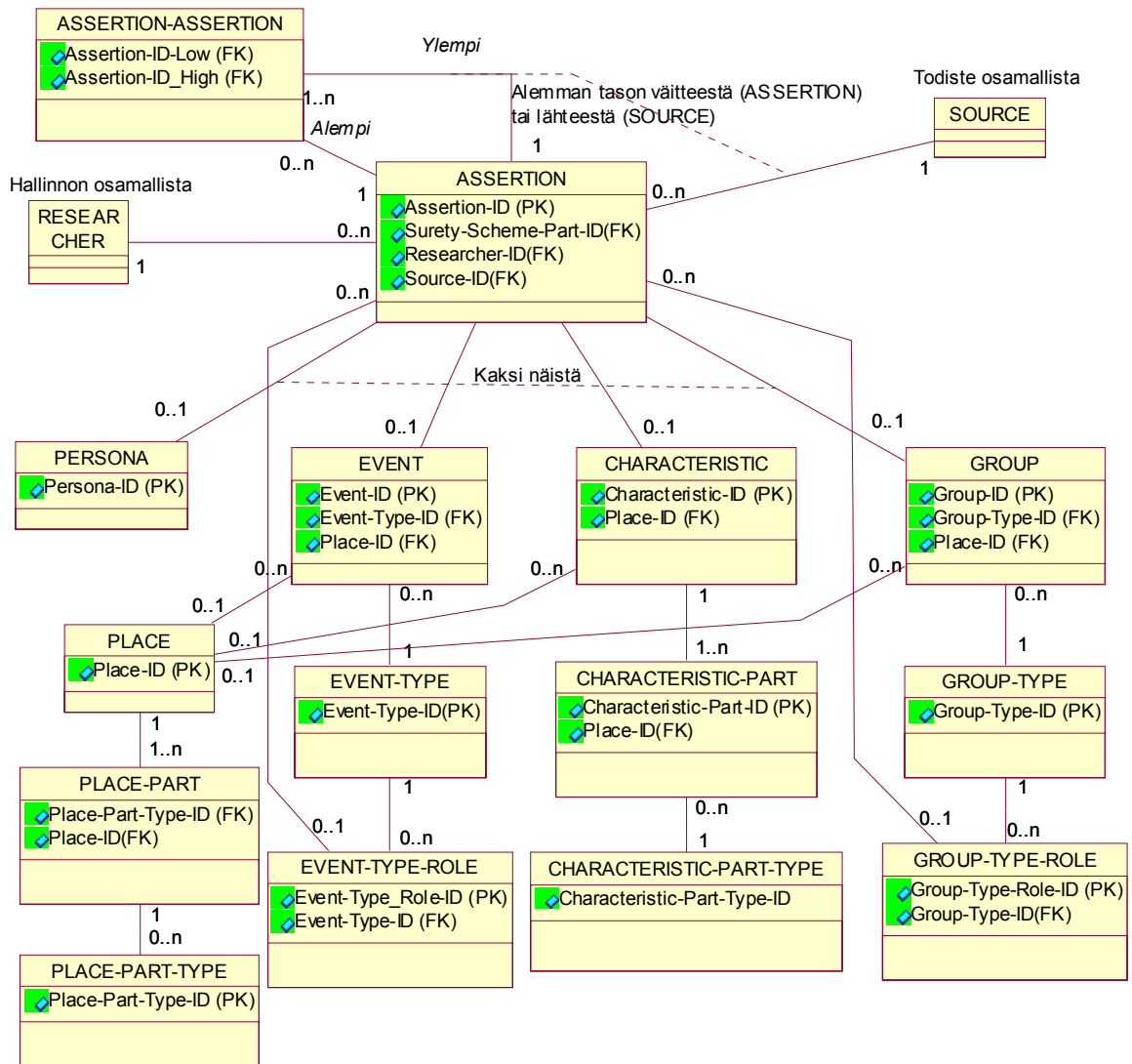
REPOSITORY sisältää yleensä useita lähteitä. Toisaalta yksi lähde voi löytyä useasta arkistosta. On myös mahdollista, että tutkija tietää määrätyn lähteen olemassaolosta, mutta ei tiedä mistä sen löytäisi. Tästä johtuu, että lähde löytyy 0..n arkistosta eikä 1..n arkistosta kuten voisi muuten ajatella. Merkintä ”ainakin toinen” osoittaa, että kunkin REPOSITORY-SOURCE-olion tulee olla kytkettynä lähteeseen tai arkistoon tai kumpaankin. Yleisessä etsintä-tapauksessa tämä kytkeytyy kumpaankin. Lähde voi sisältää muita lähteitä. Tämä näkyy kaaviossa siitä, että lähde viittaa itseensä.

Esimerkkinä voitaisiin ajatella tuomiokirjoja, jotka sisältävät useiden oikeustapausten pöytäkirjoja. Lähteet sisältävät sekä ensisijaisia että toissijaisia lähteitä. Esimerkkinä vaikkapa toissijaisena lähteenä ns. mustat kirjat, jotka puolestaan ovat alkuperäislähteiden eli historiankirjojen toissijainen lähde. SOURCE-taulu ei sisällä mitään viittauksen tarvitsemaa tietoa kuten otsikko tai tekijä. Viittaukset talletetaan CITATION-PART tauluun, joka voisi sisältää esim. Mustat kirjat. CITATION-PART-TYPE voisi tällöin olla kastettujen luettelo. Sitaatin varsinainen sisältö talletetaan REPRESENTATION tauluun. REPRESENTATION voi sisältää myös multimedia-sisältöä, esimerkiksi äänitetyn haastattelun tai valokuvan. Tämän tyyppiä kuvaava taulu REPRESENTATION-TYPE sisältää tämän tyyppin, kuten GIF tai MP3.



Kuva 18. Todisteet osamalli

Tietomallin ydin on *johtopäätökset-osamalli* (kuva 19), erityisesti väittämä-taulun



Kuva 19. Johtopäätökset osamalli

ASSERTION rooli. Kun todistuskappale on eristetty yksinkertaiseksi otteeksi lähteestä, tutkija käyttää tätä tehdäkseen väitteen (ASSERTION). Vaikka varsinainen ote sijaitsee REPRESENTATION taulussa, se on väitteen käytettävissä SOURCE tietueen kautta.

Väittäjä (ASSERTION) voidaan luoda kahdella tavalla. Ensimmäinen tapa on muuntaa lähteen (SOURCE) ote väittäjäksi (ASSERTION). Muunnos tehdään kuvassa 17 esitetyn superlausetyypin tavalla. Kukaan lähde voi johtaa yhteen tai useampaan väittämään, mutta väittäjä, joka on johdettu tällä tavalla, pohjautuu yhteen ja vain

yhteen lähteeseen. Toinen tapa on muodostaa väittämä, joka pohjautuu yhteen tai useampaan väittämään.

Väittämän käsitteen vahvuus on siinä, että kerätyt todistuskappaleet löytyvät lähteestä ja näitä on suoraan käytetty muodostamaan alemman tason johtopäätöksiä. Korkeamman tason johtopäätökset on puolestaan muodostettu alemman tason väittämistä tarpeen mukaan. Tästä seuraa, että tehdyille johtopäätöksille saadaan täydellinen jäljitysketju todisteisiin.

Tietomalli sallii myös negatiivisen väittämän, siis lauseen jossa todetaan, että jotain ei ole tapahtunut. Esim Merkintää Vapun ja Simon häistä ei ole Marttilassa vihittyjen avioparien luettelossa. Ne lienee pidetty muualla. Negatiivinen tieto voi esiintyä tietomallissa myös monessa muussa kohdassa. Tämä rakenne hallitsee kaikkia sukututkimuksen lausetyyppejä. Nämä lauseet jaetaan neljään aiheen mukaiseen tyyppiin, jotka käsitellään seuraavaksi.

Kukin väittämä sisältää seuraavaa muotoa olevan lauseen. (Katso kuvan 17 esittämä superlausetyyppi, jossa tämä tiivis sukututkimuslause selitetään).

Subjekti1-tyyppi / Subjekti1-tunnus / Subjekti2-tyyppi / Subjekti2-tunnus / Arvo

Sallitut neljä subjektityyppiä ovat henkilö (PERSONA), tapahtuma (EVENT), ryhmä (GROUP) ja ominaisuus (CHARACTERISTIC). Tunnus osoittaa kyseistä tyyppiä edustavaan käsitteeseen. Tiedot voidaan kytkeä kuvan 17 mukaisesti. Ne kentät, jotka sisältävät ruksin (✓), esittävät todennäköisimpiä vaihtoehtoja.

Esimerkiksi kuvan 17 ylin rivi esittää subjekti1:n eli henkilön kytkettynä henkilöön ja sen suhteeseen toiseen *henkilöön* (henkilö 1 on henkilö 2:n isä). Huomioi, että tämä S1-tyyppiä oleva lause esitetään myöhemmin voimakkaammalla tavalla esittämällä suhteet ryhmä-käsitettä käyttäen. *Tapahtumaan* henkilön elämässä (henkilö oli sulhasena häissä). *Ryhmään* jossa henkilö on jäsenenä (henkilö oli korpraalina joukko-osastossa) sekä *ominaisuuteen* joka henkilöllä on (henkilö oli maanviljelijä).

Esimerkit esitettynä taulukkomuodossa ovat:

| | | | | |
|---------|----------------------------|----------------|--|----------|
| PERSONA | 1234 (Heikki Yrjänänpöika) | PERSONA | 4321 (Liisa Heikintytär) | isä |
| PERSONA | 1234 (Heikki Yrjänänpöika) | EVENT | 222 (Mikon ja Vapun häät) | sulhanen |
| PERSONA | 1234 (Heikki Yrjänänpöika) | GROUP | 111 (Turun ja Porin läänin kaksikasratsuväkirykmentti) | korpr. |
| PERSONA | 1234(Heikki Yrjänänpöika) | CHARACTERISTIC | 333 (ammatti) | mv. |

Nämä neljä aiheyyppiä käsitellään erikseen seuraavassa.

Kukin väittäjä johdetaan joko lähteen sisältämästä todistusaineistosta tai yhdestä tai useammasta alemman tason väitteestä. Väitteistä voidaan muodostaa korkeamman tason väitteitä tarpeen mukaan. Myöhemmin virheelliseksi todetut väitteet voidaan poistaa tiedoista, mutta parempi tapa on merkitä ne vääriksi, jolloin toteamukset säilyvät myöhempää tarkastelua silmälläpitäen.

Kuvassa 19 ASSERTION:n alla olevalla pisteiviivalla esitetään, että käsitteistä PERSONA, EVENT, CHARACTERISTIC ja GROUP kaksi yhdistetään ASSERTION-käsitteellä.

Mikäli toinen liitoksista, joko subjekti1 tai subjekti2, kuvaa ominaisuutta, kytketään tämä ominaisuus-käsitteeseen CHARACTERISTIC. Päinvastoin kuin ryhmä- tai tapahtumatyyppisten käsitteiden tapauksessa ei väittäjä kytketä suoraan ominaisuuden osaan CHARACTERISTIC-PART, joka sisältää viittauksen kohteena olevat arvot vaan kytkentä tehdään ominaisuus-käsitteeseen.

Mikäli väittäjä liittyy ominaisuuteen, kuten henkilön nimeen, sisältää ominaisuuskäsite yleensä useita osia. Jos kysymyksessä on ammatti-ominaisuus, sisältää ominaisuus vain yhden osan. Kukin ominaisuuteen liittyvä ominaisuusosa on aina samaa tyyppiä (CHARACTERISTIC-PART-TYPE). Jos ominaisuus sisältää useita osia, kuten nimien tapauksessa on laita, järjestetään osat haluttuun järjestykseen järjestystä osoittavilla arvoilla.

Sukututkimustiedon käsittely

Rajoituksena ominaisuudet-käsitteen käytölle on, että se voi olla subjekti1:nä vain, jos subjekti2:na on ryhmä. Tämä mahdollistaa ominaisuuksien ryhmittämisen, mutta estää muut vaihtoehdot, koska ne eivät tunnu mielekkäältä.

Ammatti-ominaisuus esitetään seuraavasti:

```
ASSERTION
  Subject1-Type =          Persona
  Subject1-ID =           Persona-ID
  Subject2-Type =          Characteristic
  Subject2-ID =           C001

CHARACTERISTIC
  Characteristic-ID =      C001

CHARACTERISTIC-PART
  Characteristic-ID =      C001
  Characteristic-Part-Name = Torppari
  Characteristic-Part-Type-ID = P001

CHARACTERISTIC-PART-TYPE
  Characteristic-Part-Type-ID = P001
  Characteristic-Part-Type-Name = Ammatti
```

Nimi-ominaisuus esitetään vastaavasti seuraavasti:

```
ASSERTION
  Subject1-Type =          Persona
  Subject1-ID =           Persona-ID
  Subject2-Type =          Characteristic
  Subject2-ID =           C002

CHARACTERISTIC
  Characteristic-ID =      C002
  Ascending-Descending-None = Ascending

CHARACTERISTIC-PART
  Tietue1:
  Characteristic-ID =      C002
  Characteristic-Part-Name = Juho
  Characteristic-Part-Type-ID = P002
  Sequence-Number =       1

  Tietue2:
  Characteristic-ID =      C002
  Characteristic-Part-Name = Simonpoika
  Characteristic-Part-Type-ID = P003
  Sequence-Number =       2

CHARACTERISTIC-PART-TYPE
  Characteristic-Part-Type-ID = P002
  Characteristic-Part-Type-Name = Etunimi

  Characteristic-Part-Type-ID = P003
```

Sukututkimustiedon käsittely

Characteristic-Part-Type-Name = Patronyyymi

Mikäli toinen väittämän osista kuvaa tapahtumaa, se kytketään tapahtumakäsitteeseen EVENT. Tapahtuma voisi olla esimerkiksi ”Mikko Simonpojan ja Liisa Heikintyttären häät”. Kunkin tapahtuma on määrättyä tapahtumatyyppiä EVENT-TYPE. Tässä tapauksessa tapahtumatyyppin nimi olisi ”Häät”. Tapahtumatyyppi voi luonnollisesti esiintyä 0..n tapahtumassa.

Subjekti1 tai subjekti2 kytkeytyy tapahtumaan. Väittämän arvo kytketään tapahtumatyyppin rooliin EVENT-TYPE-ROLE. Koska väittämässä voi olla vain yksi arvo, vaikka kumpikin liitos osoittaisi tapahtumakäsitteeseen voi väittämä sisältää 0..1 tapahtumatyyppin roolia. Tapahtumatyyppin rooli on usein tyhjä; esimerkiksi syntynyt-tapahtuma ei tarvitse roolia.

Tapahtuma voisi sisältää seuraavat arvot:

ASSERTION

| | |
|-----------------|------------|
| Subject1-Type = | Persona |
| Subject1-ID = | Persona-ID |
| Subject2-Type = | Event |
| Subject2-ID = | E001 |
| Value = | R001 |

EVENT

| | |
|-----------------|--|
| Event-ID = | E001 |
| Event-Type-ID = | T001 |
| Event-Name = | Mikko Simonpojan ja Liisa Heikintyttären häät |

EVENT-TYPE

| | |
|--------------------|------|
| Event-Type-ID = | T001 |
| Event-Type -Name = | Häät |

EVENT-TYPE-ROLE

| | |
|-------------------------|---------------|
| Event-Type-Role-ID = | R001 |
| Event-Type-Role -Name = | Morsiamen isä |

Mikäli toinen väitteen liitoksista, joko subjekti1 tai subjekti2, kuvaa ryhmää kytketään tämä ryhmään GROUP. Ryhmä kuvaa määrättyä henkilöiden tai muiden olioiden joukkoa.

Ryhmätyyppi GROUP-TYPE on käytössä 0..n ryhmässä. Esimerkiksi ”Komppania” ryhmätyyppi esiintyy kaikissa varsinaisissa komppania-ryhmissä.

Väittäjä, joka on kytketty ryhmään, kytketään myös ryhmätyypin rooliin GROUP-TYPE-ROLE väittäjän arvo-kentän välityksellä. Ryhmätyypin roolin nimi voi olla esimerkiksi ”Kersantti” tai ”Eversti”. Puolisoiden lapsikatrasta kuvaavan ryhmän eräs ryhmätyypin rooli voi olla ”3:s”.

Kutakin ryhmätyypin roolia vastaa 1..n ryhmätyyppiä. Koska monella ryhmällä on arvojärjestys käytössä, ryhmätyypin roolin järjestysnumeroa voidaan käyttää tähän tarkoitukseen.

Ryhmä voisi sisältää seuraavat arvot:

```
ASSERTION
  Subject1-Type =          Persona
  Subject1-ID =           Persona-ID
  Subject2-Type =          Group
  Subject2-ID =           G001
  Value =                R002

GROUP
  Group-ID =              G001
  Group-Name =            Mikko Simonpojan ja Liisa
                          Heikintyttären lapset

GROUP-TYPE
  Event-Type-ID =         T001
  Event-Type -Name =      Häät

EVENT-TYPE-ROLE
  Group-Type-Role-ID =    R002
  Group-Type-Role -Name = Lapsi
  Sequence-Number =      3
```

Ryhmäkäsitteen käytölle asetettujen sääntöjen mukaan uusi ryhmä ei voi olla subjekti 1. Ryhmä muodostetaan liittämällä ryhmään henkilö, tapahtuma, ominaisuus tai olemassa oleva ryhmä. Väittäjä jossa molemmat subjektit ovat ryhmiä tulee tehdä niin, että olemassa oleva ryhmä on subjekti 1 ja uusi ryhmä on subjekti 2. Ryhmän kaikkien jäsenten tulee olla samaa tyyppiä: henkilö, tapahtuma, ominaisuus tai ryhmä. Vain samaa tyyppiä olevat ryhmät voidaan yhdistää uudeksi ryhmäksi. Ryhmän tulee olla samaa tyyppiä kuin sen alimman tason jäsenet. Tutkija ei voi esimerkiksi muodostaa henkilöä joukosta tapahtumia.

Huomaa jäljempänä oleva henkilöä koskeva vastaava sääntö.

Mikäli toinen väittämän liitoksista kuvaa henkilöä, se kytetään henkilö-olioon PERSONA. Henkilö-olion nimikenttä voi sisältää arvon ”Mikko Simonpoika”. Tätä nimikenttää sukututkija voi käyttää henkilöstä. Yksityiskohtaiset nimen osat, jotka pohjautuvat suoraan todistemateriaaliin, on talletettu ominaisuusosiin.

Tutkija tekee väittämän todistemateriaaliin perustuen henkilö-oliosta. Tiedonkeräyksen yhteydessä löydetty materiaali voidaan kytkeä moneen eri henkilö-olioon, joilla kaikilla voi olla sama nimi, mutta joista ei vielä tiedonkeräysvaiheessa voida sanoa, että ne edustavat samaa henkilöä.

Riittävän todistemateriaalin saatuaan tutkija voi kerätä nämä henkilöoliot yhdeksi henkilöksi. Tämä tehdään muodostamalla väittäjä ja ryhmäkäsité, jolla voidaan yhdistää useita tutkijan mielestä samaa henkilöä edustavaa henkilö-oliota. Ryhmän järjestysnumeroa ei tällöin tarvita, ellei esimerkiksi haluta sillä osoittaa henkilön nimenvaihdokset hänen elämänsä aikana.

Huomioi, että henkilö-olio voidaan luoda vain joko niin, että henkilö-olio luodaan *suoraan lähdetiedoista* tai se luodaan *väitteellä*, joka yhdistää kaksi tai useampaa väittämää. Tämä käsittää kahden tai useamman henkilö-olion yhdistämistä uudeksi henkilö-olioksi käyttäen ryhmäkäsitetä.

Nämä huomiot johtavat sääntöön, jonka mukaan olemassa oleva henkilö-olio ei voi olla subjekti2. Henkilö-olio voi olla subjekti2:na vain silloin, kun uusi henkilö-olio luodaan ryhmästä. Jos henkilö-olio luodaan lähteestä on se aina subjekti1 ja subjekti2:na on tällöin ominaisuus, tapahtuma tai ryhmä.

Tarkastellaan kuvan 17 mukaan seuraavia mahdollisia vaihtoehtoja, joissa henkilö-olio on subjekti2:na.

| <u>Subjekti1</u> | <u>Subjekti2</u> |
|------------------|------------------|
| Ryhmä | Henkilö |
| Tapahtuma | Henkilö |
| Ominaisuus | Henkilö |
| Henkilö | Henkilö |

Yleinen väittämän muoto, jossa henkilö-olio on subjekti2:na on se, jossa väittäjä luodaan ryhmästä olemassa olevia henkilö-olioita kuten

Ryhmä-173 (joka sisältää esimerkiksi henkilöoliot 1-7) on henkilö-olio 8

Väittämä ”Ryhmä on uusi henkilö” on sallittu. Tarkastellaan puolestaan seuraavia väittämiä, joissa ei ole järkeä:

Tapahtuma-3 on Henkilö-5
Ominaisuus-2 on Henkilö-13

Tarkastellaan vielä väittämän henkilö – henkilö muotoa. Tämä muoto on selvästi järkevä, mutta tämän sijaan on päädytty yhdistämään henkilöt ryhmäkäsitteen avulla.

Seuraavan esimerkin henkilö-henkilö muodot pitäisi käsitellä ryhmäkäsitteellä:

Henkilö-1 on sama kuin Henkilö-2
Henkilö-1 on isä henkilölle Henkilö-2
Henkilö-1 on syntynyt ennen Henkilö-2
Henkilö-1 on pidempi kuin Henkilö-2

Säännön tarkoituksena ei ole henkilö-olion käytön estäminen, vain uusien henkilö-olioiden luomisen rajoittaminen edellä mainittuihin kahteen tapaan.

S1-lausetyyppi yhdistää kaksi henkilöä suoraviivaisesti (P1 / on isä henkilölle / P2). Jos käytössämme on todiste, joka yhdistää isän 6:een lapseen tarvittaisiin 6 tällaista lausetta. Mikäli samanaikaisesti esiintyisi muita suhteita tulisi tästä huomattavan monimutkainen.

Yksinkertaisella esimerkillä tämän voisi esittää:

| <u>Subjektii1</u> | <u>Subjektii2</u> |
|-------------------|--------------------------|
| Henkilö1 | Ryhmäl (roolina "Isä") |
| Henkilö2 | Ryhmäl (roolina "Lapsi") |

Tähän ryhmään (perhe) on helposti lisättävissä muita lapsia sekä äiti.

Huomaa edellä oleva ryhmää koskeva vastaava sääntö.

GEDCOM-L-postituslistalla on käyty vilkasta keskustelua GENTECH-mallin puolesta ja vastaan. Yhtenä keskustelun aiheena on ollut perheenjäsenten oikea järjestys. Useimmat sukututkimusohjelmistot ja GEDCOM pystyvät järjestämään perheenjäsenet haluttuun järjestykseen. Puutteenksi ovat eräät keskustelijat nähneet sen, että näillä menetelmillä ei voida erottaa satunnaista järjestystä tutkijan asettamasta järjestyksestä. Monen muun käsitys asiasta on ollut, että jo GEDCOM:n tapa järjestää perheenjäsenet on riittävän tarkka. Toisaalta sukututkimusohjelmistoilla voidaan tallettaa tekstimuotoista tietoa eri tietueitten yhteyteen, ja sinne erikoisetkin perhesuhteet voidaan kuvata sanallisesti.

GENTECH-tietomalli ei sisällä tekstitiedolle omaa käsitettä. Mallin suunnittelijat ovat ilmeisesti nähneet tekstimuotoisen tiedon lähinnä lähdetiedoksi, josta tulisi väittämiä

käyttäen luoda tapahtumia tms. Tämä ei ole riittävää suomalaisessa sukututkimuksessa, jossa tekstitietoja käytetään melko runsaasti. Tulosteissa tekstitiedot halutaan voida sijoittaa tapahtumien joukkoon sopivaan kohtaan.

Tietomalli on todettu vaikeaksi ja monimutkaiseksi. On epäilty, voiko näin monimutkaisen mallin avulla rakentaa toimivaa sukututkimusohjelmistoa. Malli sisältää toisaalta runsaasti oivalluksia, joista tulee olemaan hyötyä. Esimerkiksi se, että kaikki mallin mukaisesti tallennettu tieto voidaan jäljittää tiedon lähteelle, on erittäin hyödyllistä. Ohjelmiston tehtävä on huolehtia siitä, että näiden tietojen syöttäminen ei koidu käyttäjälle liian raskaaksi. Tämä tapahtuu esimerkiksi istunnon alussa, jolloin käyttäjä valitsee käyttämänsä lähdetiedot. Tällöin käyttäjän täytyy erikseen huomioida vain poikkeavat tapaukset.

7.7 DeadEnds –tietomalli

LifeLines-ohjelman tekijä Thomas Wetmore osallistui aktiivisesti GEDCOM-L-postituslistan keskusteluihin GENTECH-tietomallin soveltamisesta syksyllä 2000. Thomas Wetmore on useasti esittänyt mielipiteenään, että relaatiomalli ei sovellu sukututkimusohjelmistojen tarpeisiin erityisen hyvin. Hän on ilmoittanut pitkäntähtäimen suunnitelmanaan Java-pohjaisen DeadEnds-nimisen sukututkimusohjelmiston kehittämisen.

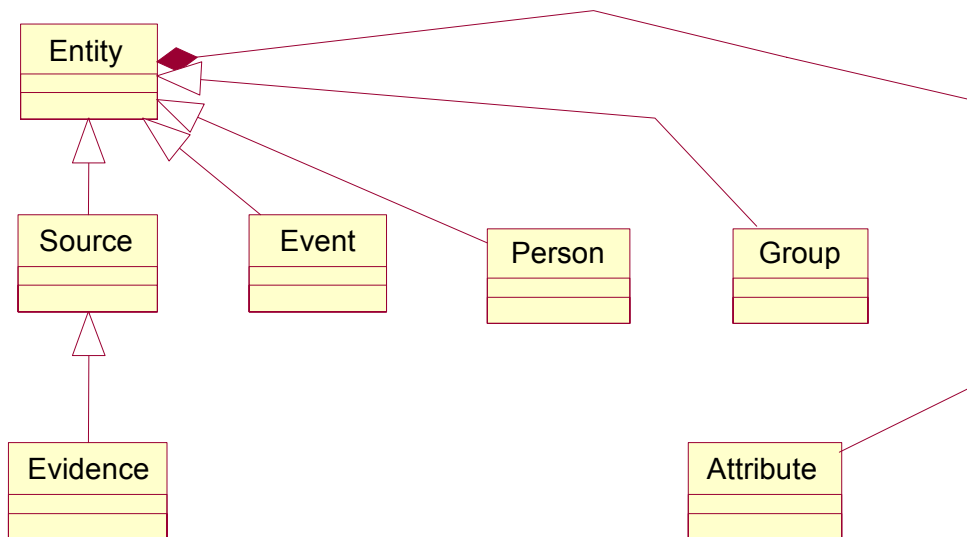
Thomas Wetmore kehitti DeadEnds-tietomallin [Pri99] syksyllä 2000 GEDCOM-L-listalla käytyjen keskustelujen yhteydessä. Hän on antanut mallille versionumeroksi 0.0001, mikä osoittaa, että kysymyksessä on lähinnä aiheesta paperille laitettuja ajatuksia. Ajatukset ovat sukututkimusohjelmistojen merkittävän vaikuttajan ja ”toisinajattelijan” ajatuksia GENTECH-mallin ajatusten soveltamiseksi tässä tapauksessa tähän hänen DeadEnds-ohjelmistoonsa.

Kuva 20 esittämä malli koostuu kahdesta alimman tason luokasta Entity (olio) ja Attribute (attribuutti). Kaikki muut rakenteelliset luokat periytyvät näistä. Malli sisältää myös muutaman jakamattoman tietoelementtityypin. Esimerkiksi nimi, päiväys ja paikka ovat tällaisia.

Kaikki erilliset luokat periytyvät *Olio-luokasta*. Oliot ovat soveltamismielessä tietokannan tietueita. Olio määritellään siten, että sillä on yksi tunnus, 0..1 lähdettä ja 1..n attribuuttia. Tunnus on jakamaton tietoelementti. Tunnus on tietokannassa yksiselitteinen. Oliolla voi olla lähde joka kuvaa sen lähdettä tai muuta oikeutusta. Oliolla ei ole yliluokkaa.

Attribuutti on perusluokka kaikille luokille, jotka kuvaavat Olio-luokan ominaisuuksia. Attribuutilla on yksi nimiö, 0..1 arvoa ja 0..n attribuuttia. Nimiö vastansee GEDCOM:in nimiötä. Nimiö ja arvo ovat jakamattomia tietoelementtejä.

Lähde (source) on olion aliluokka, joka edustaa sukututkimuksen lähdetietoja. Lähteen aliluokka todiste edustaa alkuperäistä sukututkimustietoa. Lähteitä on useita erilaisia. Näiden ominaisuuksia kuvaavat erilaiset attribuutit. Lähde määritellään siten, että lähde on olio, sillä on 0..1 lähdettä ja 1..n attribuuttia.



Kuva 20. DeadEnds-tietomallin luokkakaavio

Koska lähde voi sisältää toisen lähteen voivat lähteet linkittyä rakenteeseen, joka sallii laajempien lähteiden sisältää yksilöityjä lähteitä. Lähde voisi kuvata arkistoa, jonka sisältämä lähde on kirja. Yksi lähteen attribuutti määrittelee sen tyyppin, muut attribuutit riippuvat sen tyyppistä.

Todiste edustaa sukututkimuksen alkuperäisaineistoa. Tämä sisältää sukututkimuksen raaka-ainetta ja johtopäätösten perustan. Esimerkkeinä todisteista ovat kirkonkirjojen

otteet, haastattelujen tuloksena saadut muistiinpanot jne. Todiste määritellään siten, että todiste on lähde (näin ollen todiste on olio). Sillä on lähde ja 0..n attribuuttia

Todisteen kuvaus-attribuutti kuvaa ulkoisen todistemateriaalin sisältämättä tarkemmin sen sisältöä. Sen tarkennus-attribuutti sisältää materiaalista sen osan, jonka tutkija on arvioinut tärkeäksi. Ote-attribuutti sisältää materiaalin puhtaaksikirjoitettuna mahdollisimman täydellisenä ja tiedostoviittaus sisältää tietokannan ulkopuolisen sähköisesti talletetun tiedon osoitteen.

Tapahtumaluokka edustaa tapahtumia henkilön elämän aikana. Todisteista löydetty tapahtumat tulee pääsääntöisesti tallettaa tapahtumina. Tapahtuma määritellään siten, että tapahtuma on olio, sillä on 1 lähde, 1..n attribuuttia, (0..n roolia) ja 0 tai 2..n tapahtumaa. Tapahtuman yleisimmät attribuuttityypit ovat päiväys, paikka ja roolit.

Mikäli tapahtumalla ei ole alitapahtumia sisältää tapahtuma kaiken tiedon mitä siitä tunnetaan. Kahden tai useamman tapahtuman sisältävä tapahtuma edustaa näiden alitapahtumien johdosta tehtyä päätelmää.

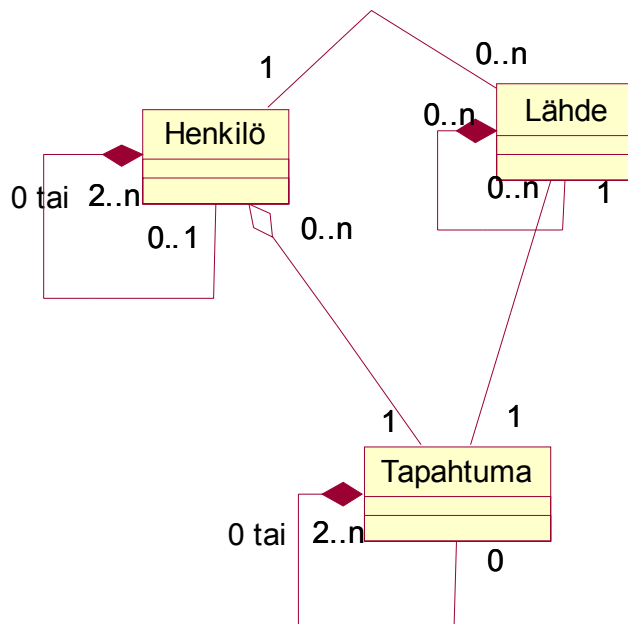
Yleisiä tapahtuman attribuutteja ovat mm tyyppi, päiväys, paikka ja rooli. Rooli on erityinen attribuutti jossa attribuutin nimiö edustaa roolia tapahtumassa ja attribuutin arvo viittaa henkilöön joka on tässä roolissa.

Henkilöolio edustaa henkilöitä. Kukin henkilöolio sisältää tietoa, jonka tutkija uskoo liittyvän yhteen todelliseen henkilöön. Henkilö määritellään siten, että se in olio. Sillä on 1 lähde, 1..n attribuuttia ja 0 tai 2..n henkilöä. Henkilön attribuutit sisältävät henkilöön liittyvää tunnettua tietoa. Attribuutteja relaatio, rooli ja vitaali käytetään henkilöoliossa määrittelemään näiden suhteet muihin henkilöihin ja tapahtumiin. Nämä attribuutit yhdessä tapahtumaolion rooli-attribuutin kanssa ovat mallin tarjoamat menetelmän suhteiden esittämiseen.

Henkilöillä on joukko yleisiä attribuutteja kuten nimi, sukupuoli, sukulaisuussuhde ja rooli, joka henkilöllä on tapahtumassa. Henkilöllä voi lisäksi olla vitaali attribuutti, joka sisältää tapahtuman joka esitetään attribuuttina. Vitaali on siksi *kevytrakenteinen tapahtuma*. Muita henkilön attribuutteja ovat ammatti, oppiarvo, sotilasarvo, maa, asuinpaikka jne.

Ryhmä on olion aliluokka, joka edustaa joukkoa eri henkilöitä ja joiden olemassaolo on todettu johonkin todistusaineeseen perustuen. Ryhmä määritellään siten, että se on olio. Sillä on 1 lähde, 1..n henkilöä ja 0..n attribuuttia.

Suhteiden esittäminen voi olla tietomallin monimutkaisin kohta. DeadEnds malli tarjoaa kolme tapaa esittää suhteita. Ensimmäisellä tavalla (relaatio) henkilöt ovat pareittain suhteessa toisiinsa. Kuvan 21 esittämällä toisella tavalla (rooli) liitetään henkilöt ja tapahtumat toisiinsa (vertaa tyyppi S1 kohdassa 7.4). Kolmas tapa esittää suhteita on vitaali-attribuuttia käyttävä, jolla suhteen henkilö on kuvattu attribuutissa.



Kuva 21. DeadEnds tietomallin olioiden välisiä suhteita

Esityksessä ei mainita ryhmän käyttöä henkilöiden suhteen esittämisessä, mutta toisaalta tämä on todettu ryhmän määrittelyn yhteydessä joten tämän GENTECH-mallissa omaksuttu lähestymistapa voidaan katsoa olevan käytössä.

7.8 Tekstitiedon esittäminen

Erillistä tekstitietoa kuvaavaa attribuuttia ei GENTECH- tai DeadEnds-mallissa ole mainittu. Tekstitiedon tallettaminen on Suomen sukututkijoiden keskuudessa hyvin

yleistä, mutta ilmeisesti vähemmän tärkeätä näiden mallien suunnittelijoiden ajatusmaailmassa. Tekstitiedolla voidaan esittää pitkiäkin kertomuksia henkilön elämästä ilman että tekstin joka kohdasta saadaan muodostettua erillisiä tapahtumia tai ominaisuuksia. Tekstitiedon lisääminen toteutettavaan ohjelmistoon ei tuota ongelmaa, koska tekstitieto on rinnastettavissa aika pitkälle ominaisuuksiin / attribuutteihin.

8 Yhteenveto

Sukututkimustiedon käsittely tietokoneella voidaan katsoa alkaneeksi 1960-luvulla, jolloin mormonikirkko ryhtyi järjestelemään laajoista mikrofilmatuista kokoelmistaan saatuja nimiä uskonnollisia tarkoituksia varten. Henkilökohtaisten tietokoneiden yleistymisen myötä 1980- ja erityisesti 1990-luvulla yleistyi myös tietokoneiden käyttö sukututkijoiden keskuudessa. Sukututkijat havaitsivat varhain tietokoneista saatavan hyödyn laajojen tietojensa järjestämiseksi ja julkaisemiseksi.

Tietokoneiden ja käyttöjärjestelmien yhdenmukaistuessa 1980-luvulta alkaen on myös sukututkimuskäyttöön suunniteltujen ohjelmistojen käytettävyys parantunut. Mormonien kehittämä GEDCOM-standardi on yhdenmukaistanut sukututkimusohjelmistojen käyttämää tietoa ja mahdollistanut tiedonsiirron ohjelmien välillä.

Sukututkimustiedon mallintaminen on vielä puutteellisella tasolla. GEDCOM-standardiin liittyvä yksinkertainen malli on tavalla tai toisella huomioitu useimmissa sukututkimusohjelmistoissa. GENTECH-työryhmän esittämä sukututkimuksen päättelyä kuvaava malli on herättänyt kiinnostusta ja keskustelua, mutta siihen pohjautuvia sovelluksia ei vielä ole olemassa.

On odotettavissa, että myös sukututkimusohjelmistot ryhtyvät hyödyntämään XML-tekniikan luomia mahdollisuuksia laajemmin. GEDCOM-siirtoformaattia vastaavan XML-formaatin käyttöönotto on odotettavissa. Toistaiseksi ei yhteistä XML-määritystä sukututkimustiedon käyttöön ole kehitetty josta syystä sen käyttö on ollut vähäistä. Tapahtuuko sukututkimustiedon XML-määritysten käyttöönotto sukututkimusohjelmistojen kehittäjien yhteisen työn tuloksena vai ryhtyykö jokin (yhdysvaltalainen) markkinajohtaja tai käyttäjä (kuten mormonit) sitä ensiksi soveltamaan jota sitten muut seuraavat, jää nähtäväksi. Sukututkimustiedon XML-talletusformaatti tuo mukanaan

paremman tuen myös muille kuin länsieurooppalaisille aakkosille. Mm. tämä on eräs GEDCOM-standardin puute, jolle odotetaan parannusta XML-muotoisen tiedon käytön yhteydessä.

Lähdeluettelo

- Ber01 Sukukirja, Suomen aatellittomia sukuja, Suomen Muinaismuistoyhdistyksen puolesta toimittanut Axel Bergholm, Helsingissä Kustannusosakeyhtiö Otava 1901.
- Dis00 Föreningen för datorhjälp i släktforskningen (DIS), <http://www.dis.se>
- Dom02 Document Object Model (DOM). <http://www.w3.org/DOM/>
- Eas01 Dick Eastman Online niminen "internet-lehti".
<http://www.beliefnet.ancestry.com/library/view/columns/eastman/eastman.asp>
- Fam00 FamilySearch Personal Ancestral File. <http://www.familysearch.org/>
- Ged96 The GEDCOM standard, release 5.5. Prepared by the Family History Department, The Church of Jesus Christ of Latter-day Saints, 2 January 1996. <ftp://gedcom.org/pub/genealogy/gedcom/>.
- Ged98 The GEDCOM Standard (Future Direction), Prepared by the Family History Department, The Church of Jesus Christ of Latter-day Saints, 24 April 1998.
<ftp://gedcom.org/pub/genealogy/gedcom/>.
- Gdm00 Gentech, Irving, Texas, Genealogical Data Model Phase 1, May 29,2000. <http://www.gentech.org>
- Gen98 Genus Sukututkimusohjelmisto, Citius Software, Kajaani <http://www.citiussoftware.com>
- Juu00 Juuret ohjelmisto. Consegur Oy, Kajaani, <http://www.juuret.fi>
- Kai00 Kankarin Kailat nro 6 sukuluettelo (Espoo 2000), Kaarle Kaila ISBN 952-91-2598-4.
- Kai01 Kaarle Kaila, KK-Software, www.kk-software.fi
- Kay99 Michael Kay, GedML <http://users.iclway.co.uk/mhkay/gedml/>
- Luh00 Seppo Luhtasaari, Sukujutut ohjelmisto, Seinäjoki, <http://www.sukujutut.fi>
- Lut59 Georg Luther, Om personnumrering i antavlor, Suomen Sukututkimusseuran Vuosikirja 37 (1957-1959) sivut 42-58.
- Per00 Pervasive Software, Btrieve. Tietokantaohjelmisto, <http://www.pervasive.com>
- Pri00 Rafal Prinke, Genealogy software I like, <http://hum.amu.edu.pl/~rafalp/g-soft.htm>
- Pri01 Pedigree in the Machine: The Past and Future of Genealogical Computing. Pre-publication of article to be published in the History and Computing journal in 2001.
- Pri96 Rafal Prinke, Quo Vadis, GEDCOM? The Computer Genealogist Vol V, No.6 (July/August 1996)
- Pri98 Rafal Prinke, What Is Your Data Model? The Computer Genealogist Vol VII, No.5 (September/October 1998) sivut 3-12

Sukututkimustiedon käsittely

- Pri99 Rafal Prinke, Varieties of Genealogical Experience, 11 September 2000,
<http://main.amu.edu.pl/~rafalp/GEDCOM/voge.html>
- Roo02 RootsWeb.com. The oldest and largest FREE genealogy site, supported by Ancestry.com,
<http://www.rootsweb.com>. Verkkosivun tietojen mukaan vanhin ja suurin ilmainen
sukututkijoiden verkkosivusto. Osa MyFamily.com, Inc yhteisöä.
- Sax02 Official website for SAX. <http://www.saxproject.org>.
- Sip97 Seppo Sippu, On the meaning of FAMilies in GEDCOM, The Computer Genealogist Vol VI, No
2, March/April 1997 sivut 9-15
- Sss00 Suomen Sukututkimusseura, <http://www.genealogia.fi>
- Ste82 John Steed, Brother's Keeper, official homepage, <http://www.bkwin.com>
- Sur71 Sursillin suku, Genealogia Sursilliana, Weilin+Göös Tapiola 1971.
- Wet91 Thomas T. Wetmore IV, Lifelines, second generation genealogy software,
<http://lifelines.sourceforge.net/>
- XML98 Extensible Markup Language (XML). <http://www.w3.org/XML>
- Xsc02 XML Schema <http://www.w3.org/XML/Schema>
- Xsl99 XSL Transformations (XSLT), Version 1.0, W3C Recommendation 16 November 1999
<http://www.w3.org/TR/xslt>