Korkeakoulujen tutkimustietokantojen seurantaryhmän loppuraportti

25. maaliskuuta 1994

Tiivistelmä

Selonteko pohtii korkeakoulujen tutkimushankkeita koskevien tietokantojen perusteita ja tarvetta sekä määrittelee niiden tietosisällön siten, että eri korkeakouluissa käynnistyvät tietokannat voivat olla keskenään yhteismitallisia. Rekisterien perustaminen todetaan nykyoloissa olennaisesti ajankohtaisemmaksi kuin aiemmin. Syynä tähän on korkeakoulujen hallinnon uudet, tuloksellisuutta korostavat periaatteet. Tutkimushankkeista tullaan tarvitsemaan tietoja normaalissa tiedekuntatason hallinnossakin, kun resursseja jaetaan laitosten tarpeiden ja tuloksellisuuden mukaan.

Tutkimustietokantojen tietosisältö määritellään abstraktilla tavalla käyttäen rakenteisten dokumenttien määrittelemiseksi laadittua SGML-nimistä ISO-standardia. Tämä formalismi on varsin yleinen ja joustava, ja sallii siten tietosisällön kuvaamisen hyvin tarkoin halutunrakenteisena. Tämä on etu verrattuna saatavilla oleviin tietokantaohjelmiin, myös tekstitietokantaohjelmiin nähden.

Hyvin tärkeänä on pidetty laadullisen pikemmin kuin määrällisen tiedon keräämistä. On myös pyritty siihen, että tutkimushankkeet voisivat välttää samojen asioiden päällekkäisen raportoinnin eri tahoille. Korkeakoulun keskushallinnolle, Suomen Akatemialle ja vastaaville tahoille riittäisi yksi ja sama tietokonemuotoinen raportti, joka sisältäisi tietokantaan hankkeesta menevät tiedot. Samat tiedot siistinä paperitulosteena voisivat toimia hankkeen esitteenä.

Seurantaryhmä ei ota kantaa tapaan, jolla korkeakoulut tulevat toteuttamaan tutkimushankkeiden tietokantoja, ei myöskään aikatauluihin. Tietojen muuntamista tavanomaisesta tekstitietokannasta yhteiseen muotoon ja päinvastoin on kuitenkin kokeiltu käytännössä. Tietosisällön täsmentäminen mahdollistaa koko valtakuntaakin koskevan tietokannan rakentamisen, mutta seurantaryhmällä ei ollut toimeksiantoa eikä resursseja tällaisen yksityiskohtaisempaa suunnittelua varten.

Sisältö

Luku 1

Johdanto

Raportin alkuosassa todetaan seurantaryhmän perustamisen ja toiminnan vaiheet sekä pohditaan tutkimustietokantojen tarpeellisuutta ja erinäisiä periaatteita, joiden mukaan tietoja pitäisi hahmottaa ja tulkita. Toisessa osassa esitellään lyhyesti tietosisällön pääpiirteet ja kolmannessa annetaan lyhyt johdatus SGML-formalismiin, jota käyttäen raportin loppuosassa tietosisällön yksityiskohdat määritellään. Raportin päätteeksi on erinäisiä ehdotuksia jatkotoimenpiteiksi, joiden avulla voidaan parantaa mahdollisuuksia toimivien ja yhteensopivien tutkimushanketietokantojen perustamiseksi valtakunnan eri korkeakouluissa.

1.1 Seurantaryhmä

Korkeakoulujen tutkimustietokantojen seurantaryhmä on kokoontunut toimikautensa aikana yhteensä 11 kertaa. Yksi kokouksista pidettiin VTT:n tiloissa, yksi Tilastokeskuksessa ja loput Helsingin yliopistolla. Ryhmä on pyytänyt korkeakouluja, Suomen Akatemiaa, Tilastokeskusta, Tieteellisten kirjastojen ATK-yksikköä ja TEKESiä nimeämään yhdyshenkilön ja toimittanut näille materiaalia ja aiempia versioita tietokannan sisällön kuvailusta, sekä saanut palautetta eri tahoilta. Ryhmä on pyrkinyt siihen, että eri tahot ovat voineet hyvissä ajoin ottaa kantaa laadittuihin suunnitelmiin ja tuoda omat näkemyksensä esille. Palaute on ollut pääsääntöisesti rakentavaa.

Helsingin yliopiston nimissä anottiin opetusministeriöltä määrärahaa vuodeksi 1993 hanketta tukevaan työhön, joka raha myös myönnettiin. Sen turvin palkattiin mm. fil.yo. Mikko Lounela runsaaksi puoleksi vuodeksi. Hän suoritti suurimman osan tietosisällön ATK-teknisestä koostamisesta, testasi määritysten toimimisen käytännössä sekä laati SGML-määritystä vastaavan tekstitietokantasovelluksen ja muunnokset näiden välille. Lounela kirjoitti myös lähtöversion osasta tämän raportin tekstiä. Muut osat ja raportin viimeistelyn on tehnyt puheenjohtaja.

1.2 Taustaa

Ajatus tutkimusrekisterin perustamiseksi ei ole uusi. Jo 1970-luvun puolivälissä toimi Heikki Kallion johtama Tutkimusrekisterikomitea, joka julkaisi mietinnön,[1] jossa oli verraten yksityiskohtaiset suunnitelmat valtakunnallisen rekisterin ja sitä ylläpitävän organisaation perustamiseksi. Mietintö ei kuitenkaan johtanut käytännön toimenpiteisiin valtakunnallisten rekisterien toteuttamiseksi.

Runsasta kymmentä vuotta myöhemmin käynnistyi Tieteellisen informoinnin neuvoston (TINFO) tuella keväällä 1986 hanke, jonka tehtävänä oli selvittää Suomen tutkimusrekisteritilanne sekä laatia ehdotus tutkimusrekisterien tietosisällöstä ja tiedon esitysmuodosta. Sen tuloksena oli Nyrhisen ja Ylisen raportti,[2] jossa olevat ohjeet toistetaan opetusministeriön 10.3.1989 päivätyissä ohjeissa tutkimusrekisterien laatimiseksi.[3] Nämä synnyttivät tietyn määrän kiinnostusta, mutta eivät sinänsä johtaneet korkeakoulukohtaisten tai valtakunnallisten tutkimusrekisterien syntyyn.

Opetusministeriö rahoitti seuraavaksi ns. TUTKA-projektin, jossa pääosin Oulun yliopistossa laadittiin relaatiotietokantasovellus em. ohjeiden mukaisesti. Raportti[4] lähetettiin lausuntokierrokselle korkeakouluihin ja muihin asianosaisiin laitoksiin. TUTKA-hankkeen palautetta voidaan ehkä hahmottaa eri tavoin, mutta selvää on, että yleistä hyväksymistä hanke ei saanut, eikä sen pohjalta ole syntynyt rekistereitä.

Itse asiassa osa TUTKA-hankkeen kritiikistä oli ankaraa. Suomen Akatemia suhtautui hankkeeseen torjuvasti, samoin Helsingin yliopiston matemaattisluonnontieteellinen osasto. Tilastokeskus katsoi, ettei rekisteri toteutuessaankaan kiinnostaisi sitä. Monet palautteen antajista näyttivät suhtautuvan hyväntahtoisen myönteisesti hankkeeseen, kunhan joku muu sen toteuttaisi. Eräät korkeakoulut olivat vastahakoisia hankkimaan yhtä sovellusta varten kallista tietokantaohjelmistoa, jota muuten ei käytetty.

Korkeakoulujen tutkimusrekisteri on siten toteutunut varsin hitaasti. Tämä on jokseenkin erikoista, jos sitä verrataan siihen, että Valtion teknillinen tutkimuskeskus perusti oman tutkimusrekisterinsä jo 1975. Rekisteri oli aluksi eräkäyttöinen, mutta vuodesta 1984 lähtien tiedot ovat olleet saatavilla Valtion tietokonekeskuksen MINTTU-tietopalveluna valtakunnallisestikin. Tämä rekisteri on toiminut ennen kaikkea osana VTT:n sisäistä hallintorutiinia. Sen pohjalta myös tulostetaan vuosittaiset luettelot käynnissä olleista tutkimuksista sekä uutuusluettelot vuoden aikana. Sama tieto tarjoillaan tietokantapalveluna nykyään MINTTU-palvelun lisäksi Otaniemen tietokantaverkossa ajantasaisena.

[1] *Tutkimusrekisterikomitean mietintö*, Komiteamietintö 1976:45, Helsinki

[2] Kirsi Nyrhinen ja Ann-Britt Ylinen, *Tutkimustoiminnan rekisteröinti — ehdotus tutkimusrekisterin tietosisällöksi*, OTA-kirjasto A10, Teknillisen korkeakoulun kirjasto, Otaniemi, 1987.

[3] Mukaan lukien painovirhe, jossa tutkimuksen vaiheen koodiksi on tullut TV (p.o. VA), alkup. raportin s. 41, ohjeen s. 8.

[4] *Korkeakoulujen tutkimustietokanta TUTKA: Loppuraportti*, Opetusministeriön työryhmien muistioita, 1991:28, Helsinki 1991.

Tutkimustietokantahankkeesta aiemmin esitettyyn kritiikkiin on perehdytty ja se on pyritty ottamaan huomioon. Yhteydenpitoa eri korkeakoulujen, Suomen Akatemian ja muiden tahojen kanssa on pyritty myös hoitamaan tiiviimmin ja ennen kuin ryhmän suositukset ovat saaneet lopullisen muotonsa.

1.3 Rekisterien tarpeellisuudesta

Tärkeänä tekijänä tutkimusrekisterihankkeiden hitaalle edistymiselle korkeakouluissa lienee ollut se, että niitä ei ole koettu kovinkaan hyödyllisiksi itse tutkijayhteisön kannalta. Aiempia hankkeita perusteltiin sillä, että tutkijat löytäisivät sitä kautta toisia alan tutkijoita ja uusia yhteyksiä. Suomen pienuuden ja tutkimustoiminnan luonteen vuoksi tämä peruste oli suuresti liioiteltu. Arvostelijat ovat aivan oikein katsoneet, että tutkija, joka löytää kollegansa vasta rekisterin avulla, on perin heikko (tai aloitteleva) tutkija.

Aiemmin tutkija sai rahoitusta tutkimukselleen esim. Suomen Akatemiasta ja oli tilivelvollinen vain rahoittajalleen. Laitos ja erityisesti tiedekunta oli varsin vähän tietoinen tai kiinnostunut tutkimushankkeista. Olosuhteet ovat muuttuneet ratkaisevasti, kun korkeakoulujen hallinto on uudistunut ja samalla on siirrytty *tuloksellisuutta korostavaan ja palkitsevaan toimintamenobudjetointiin* . Nykytilanteessa tiedekunta ja sen voimavaroista päättävä dekaani arvioi laitosten kaikkea toimintaa: sekä opetuksen tuloksellisuutta että tutkimustoiminnan tasoa ja laajuutta.

Uudessa tilanteessa on jokseenkin ilmeistä, että tiedekuntataso tarvitsee tietoja laitoksilla suoritettavasta tutkimuksesta, ja että laitoksilla on intressi tarjota tällaista. Dekaani vertaa kutakin laitosta toisaalta tiedekuntansa muihin laitoksiin ja toisaalta saman alan muihin laitoksiin muissa korkeakouluissa. *Hyvän tutkimuksen erottaminen keskinkertaisesta ei voi onnistua pelkkiä määriä mittaamalla*, eikä tutkimushankkeiden rekisterillä kannata pyrkiä tällaiseen. Tavoitteeksi tulisi ottaa yleiskuvan ja avaintietojen antaminen kustakin hankkeesta siten, että kyseisen alan tai lähialan tuntija saa alustavan kuvan tutkimuksesta, sekä voi helposti hankkia lisää tietoja ja tarkistaa sekä arvioida niitä.

Aiemmissa vaiheissa tutkimusrekisterihanketta perusteltiin myös muun yhteiskunnan ja elinkeinoelämän tarpeilla. Kun korkeakoulut ovat viime vuosina avautuneet muuhun yhteiskuntaan, nämä ovat oikeansuuntaisia perusteita, mutta niiden merkitys vaihtelee tieteenaloittain. Tällaiset yhteydet saadaan helposti toteutetuksi, jos rekisteritiedot ovat olemassa, mutta ne eivät ole olleet riittäviä käynnistämään rekistereitä.

1.4 Periaatteita ja tavoitteita

Ensimmäiseksi tavoitteeksi on asetettu se, että tutkimushankkeista kerättäisiin tietoja *tutkimusyhteisön omilla ehdoilla* ja sen omiin tarpeisiin, ei ensisijaisesti ulkopuolisten tahojen palvelemiseksi. Päätavoitteeksi ei aseteta myöskään valtakunnallisten tai korkeakoulukohtaisten summatietojen keräämistä. Käytännössä tämä merkitsee tietojen laadullisen puolen korostamista määrällisen kustannuksella. Hanketta koskevat tiedot hahmotetaan pikemminkin *tutkimushankkeen yleiskäyttöiseksi esitteeksi* kuin tavanomaiseksi rekisteriksi. Esite voidaan ajatella yhden tai kahden sivun mittaiseksi yleiskatsauksellisekssi selosteeksi erotukseksi rekistereistä, joita tyypillisesti vain hallintohenkilöt käyttävät ja joista ajettavat raportit ovat yleensä summa- ja kappalemäärätietoa.

Toisena periaatteena voidaan mainita pyrkimys siihen, että tutkimushankkeiden tarvitsisi antaa tiedot käynnistymisestään ja edistymisestään *vain kertaalleen* ja, että nämä tiedot riittäisivät eri tahoille, jotka nykyään edellyttävät hankkeilta vuosittaisia tietoja. Olemassa olevien tarvitsijoiden lisäksi samoilla tiedoilla voitaisiin helposti tyydyttää muita tarvitsijoita ilman, että siitä koituu tutkijayhteisölle hallinnollista vaivaa. Tätä tietojen monikäyttöisyyttä ei kuitenkaan voida saada aivan samalla vaivalla kuin, mitä yhden tavanomaisen raportin laatiminen vie. Nykytilanteessa päällekkäinen raportointi on kuitenkin siksi yleistä, että suoranaisiin säästöihinkin voidaan päästä. Rahoittajatahoista ainakin Suomen Akatemia on alustavasti kelpuuttanut tässä raportissa esitettävän tietosisällön mukaiset raportit tutkimushankkeiden vuosittaisiksi edistymisraporteiksi. Rahoittajatahot luonnollisesti edellyttävät omien hallintorutiiniensa mukaisia minimivaatimuksia tiedoille, joista joidenkin täyttäminen on muutoin vapaaehtoista tai korkeakoulujen omien käytäntöjen varassa.

Seurantaryhmä on tulkinnut olemassa olevien ja perustettavien rekisterien kuvaavan *ensisijaisesti tutkimushankkeita, eikä siis välttämättä kaikkea tutkimustoimintaa*, jota korkeakouluissa tapahtuu. Tätä linjaa voidaan pitää selkeämpänä ja tutkimusyhteisön toiminnan kannalta käyttökelpoisempana kuin sitä, että kohteena ja jaotteluperusteena olisi tutkimus yleensä. Omaksuttu linja ei mitenkään sulje pois pienempien tai yksinomaan virkatyönä suoritettujen tutkimusten sisällyttämistä tietokantoihin niissä korkeakouluissa, missä näin halutaan. Toisaalta tällaisia tutkimuksia on perinteisesti arvioitu ja raportoitu niiden tuloksena syntyneiden julkaisujen perusteella, joten korkeakoulukohtaiset julkaisurekisterit saattavat hyvinkin riittää tällaisen tutkimuksen kuvailemiseen.

Tutkimustietokannan tietosisältö on pääosiltaan yhdenmukainen opetusministeriön aiemman suosituksen, Oulun yliopiston järjestelmän, Euroopan yhteisön ohjeiden sekä jo toiminnassa olevien VTT:n ja Teknillisen korkeakoulun tutkimustietokantojen kanssa. Tarvetta varsinaisiin muutoksiin on ollut vain vähän. Tietosisällön muotoa on lähinnä tarkennettu ja täsmennetty eräin kohdin. Näkyvin muutos on ehkä se, että tiedot on järjestetty edellä esitettyjen periaatteiden perusteella esitteelle sopivampaan järjestykseen.

Tutkimushanketta kuvaavan tietueen ajatellaan olevan *kumulatiivinen* siten, että vuosien kuluessa sen tiedot karttuvat, eivätkä uudet tiedot välttämättä korvaa aiempia. Näin esim. tutkimukseen osallistuneen henkilökunnan muutokset ilmaistaan siten, että poistuneelle henkilölle kirjataan loppupäivä ja uusi henkilö lisätään luetteloon. Vastaavasti uuden vuoden rahoitustiedot varustetaan vuosiluvulla, eikä niiden tarvitse syrjäyttää entisiä. Samoin uudet julkaisut tulevat entisten jatkoksi.

Monet tiedot ovat kiinteitä, eikä niihin ole tarvis (eikä ehkä lupakaan) kajota kesken hanketta (esim. tutkimussopimuksen numero, vastuullisen johtajan nimi, jne.). Muutama tieto (kuten tiivistelmä ja edistyminen) on ilmoitustaulun luontoinen, projekti saa ilmoittaa siinä hyödylliseksi katsomaansa tietoa ja uudistaa sitä tarpeen mukaan.

1.5 Menetelmät

Korkeakouluihin perustettavien tutkimushankkeiden tietokantojen yhteensopivuutta ei ole tässä vaiheessa mahdollista ratkaista valitsemalla yhtä ohjelmistoa ja tietokannan määrittelyä, joita kaikki korkeakoulut käyttäisivät. Sen sijaan on valittu menettely, jossa määritellään tietty *vaihtoformaatti* ja sille tarkka tietosisältö. Yksittäisen korkeakoulun tutkimustietokannan voidaan katsoa olevan yhteismitallinen ja yhteensopiva valtakunnallisesti, jos sen tiedot voidaan *ohjelmallisesti ja mekaanisesti muuntaa* tämän vaihtoformaatin muotoon siten, että tuloksessa tiedot ovat oikeanmuotoisina oikeissa kentissään, ja että kaikki tarvittavat tiedot ovat mukana. Parhaassa tapauksessa muunnokset voidaan suorittaa molempiin suuntiin, eli myös vaihtoformaatista paikallisen tietokannan muotoon. Jos edestakainen siirto onnistuu tietojen häviämättä tai vääristymättä, on yhteensopivuus kääntäen yksikäsitteinen.

Vaihtoformaatin rakenne määritellään “Standard Generalized Markup Language” eli SGML-nimisen standardin (ISO 8879) avulla. Tämän formalismin avulla on helppo määritellä tutkimushankkeen kuvailu sellaiseksi kuin on tarpeen. SGML ei aseta rajoituksia kenttien pituudelle, toistumiselle tai sille, että osakenttinä on itsessään monimutkaisen rakenteen mukaisia osia. Tässä suhteessa SGML on olennaisesti yleisempi kuin relaatiotietokannat, jotka yleensä käyttävät määrämittaisia kenttiä, tai vastaavasti tekstitietokannat, joilla on yleensä hierarkkista rakennetta koskevia rajoituksia.

Varsin pienillä kuluilla ja helposti saatavilla yksinkertaisilla välineillä[5] on voitu kokeeksi laatia TRIP-tekstitietokannan määritys ja pienimittainen koesisältö, sekä konversio-ohjelmat kumpaankin suuntaan. Samalla on laadittu muunnosohjelmat, joilla SGML-tietueista tuotetaan siistiksi muotoiltu paperidokumentti, joka voi toimia kunkin tutkimushankkeen vuosikertomuksena ja esitteenä. Muunnosohjelmat ovat vapaasti korkeakoulujen käytettävissä.

Jos ryhmän suositukset toteutuvat, on valtakunnallisen rekisterin aikaansaaminen lähinnä suoraviivainen toteutusasia, joka edellyttää tiettyjä laite- ja ylläpitoresursseja ja vastaavia ministeriötason päätöksiä. Yhteinen rekisteri olisi käytännön edellytyksenä paitsi tietojen tarjoamiseksi korkeakoulujen ja tiedekuntien johdon tarpeisiin, myös tietojen yhdistelemiseksi esim. tieteenalakohtaisiksi hakemistoiksi.

Tässä raportissa käytetty SGML-standardi saavuttaa nopeasti laajempaa suosiota ja tulee nopeasti yhä laajempien piirien tietoisuuteen ja käyttöön. Se ei tätä kir-

[5] Internet-verkosta saatavalla SGMLS-nimisellä jäsentimellä ja SGMLSASP-muuntimella, Macintoshilla toimivalla Author/Editor-nimisellä SGML-editorilla (joka oli maksullinen) ja ilmaiseksi saatavalla TeX-ohjelmalla.

joitettaessa kuitenkaan ole maassamme kovin laajalti tunnettu. Sen vuoksi kolmas luku johdattelee lyhyesti SGML-standardin pääpiirteisiin, ja sen avulla on helpompi perehtyä yksityiskohtaisempaan tietosisältöön, josta raportin loppuosa koostuu.

Luku 2

Tietosisällöstä

Tässä luvussa käsitellään toisaalta hankkeiden eli tallennettavien yksiköiden määrittelyä tutkimustietokantojen kannalta ja toisaalta käydään pääpiirteittäin läpi tietosisältö ilman muodollista SGML-formalismia.

2.1 Tutkimushankkeiden rajauksesta

Suosituksen mukaisiin tietokantoihin voidaan korkeakoulujen ja rahoittajaorganisaatioiden harkinnan ja päätösten mukaan kerätä tietoja tutkimushankkeista. Samojen tahojen on myös laadittava yksityiskohtaisia kriteerejä, joilla ratkaistaan, milloin yhteen kuuluvat eri päätöksillä rahoitetut osat raportoidaan yhtenä kokonaisuutena ja milloin erillisinä hankkeina, jotka viittaavat toisiinsa. Rahoittajat, kuten Suomen Akatemia ja TEKES laatinevat omia ohjeitaan, ja korkeakoulujen hallinnot puolestaan antanevat ohjeita, jotka ovat sopusoinnussa näiden kanssa ja ottavat lisäksi huomioon korkeakoulun omat tarpeet.

Voidaan olettaa, että tutkimustietokantoihin halutaan ainakin seuraavanlaisia kriteerejä täyttäviä hankkeita:

- Korkeakoulun ulkopuolelta (Suomen Akatemialta, TEKESiltä, säätiöiltä, yrityksiltä tms.) rahoitusta saaneita tutkimushankkeita tai tilaustutkimuksia.
- Korkeakoulusta laitostasoa ylempää myönnettyjä, ns. erityisvoimavaroja saaneita hankkeita.
- Laitostasolla tehdyllä nimenomaisella päätöksellä ns. perusvoimavaroja saaneita hankkeita.
- Laajempia tai pidempiaikaisia, riittävän laajoiksi katsottuja virkatyönä tehtyjä tutkimushankkeita, jotka halutaan tietokantaan.

Tämä vastaa käytännön termeillä sitä, että *yksikköinä ovat tutkimushankkeet, eikä tutkimus sinänsä*. Rekisterien käyttäminen myös pienimuotoisen tutkimuksen kuvaamiseen on kuitenkin teknisesti mahdollista ja korkeakoulujen päätäntävallassa, kuten aiemmin todettiin. Opetusministeriön aiemmassa suosituksessa on pidetty sisällyttämisen rajana kahdentoista henkilötyökuukauden kokonaispanosta. Seurantaryhmän kannalta yllä kuvailluilla tai vastaavilla tavoilla voimavaroja saaneet hankkeet ovat ensisijaisia raportoinnin yksiköitä. Kukin korkeakoulu voisi itse päättää tarpeestaan sisällyttää virkatyönä tehtyä tutkimusta tutkimusrekisteriinsä, vaikkakaan selvää perustetta niiden vaatimiselle ei yleisesti olekaan. Henkilötyömäärän arviointikin voi olla hankalaa silloin, kun opettajien ajankäyttöä ei kirjata tai raportoida.

2.2 Yksi vai useampi hanke?

Yhtenä yksikkönä tai tietueena ilmoitetaan yleensä yhdellä rahoituspäätöksellä käynnistetty hanke, jolla on tietty alkamis- ja päättymispäivämäärä. Jos hankkeen rinnalle syntyy liitännäis- tai tytärprojekteja, joilla on oma alkunsa ja loppunsa sekä rahoituspäätöksensä, ne ilmoitetaan erillisinä yksikköinä, mutta emo- ja tytärprojektit laitetaan viittaamaan asianmukaisesti toisiinsa.

Voidaan myös katsoa, että hanketta koskevat täydentävät rahoituspäätökset eivät sinänsä johda erillisen tietueen perustamiseen, vaikka niihin liittyisi myöntävän viranomaisen eri päätösnumero tai sopimuskoodi. Ratkaisu yksittäistapauksissa jätetään myöhempien ohjeiden ja tutkimuksen vastuullisen johtajan terveen harkinnan varaan.

Toisaalta yksi hanke voi käynnistyä useamman osapuolen yhteisellä sopimuksella, jolla kukin taho panostaa rahaa tai muita resursseja hankkeeseen. Tällaisella hankkeella olisi siis yksi nimi, yksi alkamis- ja yksi päättymispäivä, sekä yksi vastuullinen johtajakin. Tällainen voidaan mielellään käsitellä yhtenä yksikkönä, ja antaa sille tarvittaessa useampikin sopimuskoodi, jos eri osapuolilla on rahoituspäätöksilleen omat tunnukset. Edellytyksenä yhdeksi hankkeeksi tulkitsemiselle on se, että osapuolten kesken on sopimus, jolla koko hankkeen käynnistäminen määritellään.

Suurissa, valtakunnallisissa tai kansainvälisissä projekteissa voi samalla päätöksellä käynnistyä osahankkeita useammassa korkeakoulussa. Tällöin kukin osa raportoidaan oman korkeakoulun tietokantaan, mutta hankkeen tutkimussopimuksen koodina käytetään yhteisiä tunnuksia, joiden avulla suuremman hankkeen osat voidaan löytää ja tunnistaa.

Tutkimushankkeiden kuvauksessa on varattu monipuoliset mahdollisuudet viitata läheisesti liittyviin mm. saman tai eri korkeakoulun puitteissa toimiviin muihin hankkeisiin. Tätä mekanismia tulee käyttää yhteen kuuluvien, mutta eri aikana alkavien tai päättyvien, tai muutoin erillisenä raportoitavien osien kohdalla.

2.3 Tietosisällön tiivistelmä

Tässä jaksossa esitellään seurantaryhmän täsmentämää tutkimusrekisterin tietosisältöä lyhyesti ilman konkreettista formalismia. Yksityiskohtaisempi selostus tietoalkioista ja niiden rakenteesta on tämän raportin jälkiosassa. Huomattakoon myös, että kukin korkeakoulu voi omaan tietokantaansa sisällyttää muitakin tietoja kuin mitä tässä raportissa mainitaan.

Yleisiä tietoja. Nämä tiedot koskevat koko tietuetta ja ne syntyvät enimmäkseen automaattisesti ilman, että hankkeiden itse täytyy niistä huolehtia. Tutkijat eivät yleensä myöskään itse tarvitse näitä tietoja: tietueen numero, tietokannan omistaja, tutkimustietueen kieli, tietueen täyttöaste, tietueen luontipäivä, tietueen viimeinen päivityspäivämäärä. (Vrt. kohta 4.2 s. 20.)

Tutkimuksen nimi suomeksi (tai ruotsiksi) ja sama englanniksi. (Vrt. kohta 4.3 s. 21.)

Tutkimussopimuksen tiedot. Hankkeet pyritään tunnistamaan ensisijaisesti rahoittajan tms. antaman ja muutenkin olemassa olevan numeron mukaan mikäli mahdollista. Sekä rahoittajaosapuolelle että sopimuksen numerolle on varattu erillinen osakenttä. Tarvittaessa voidaan viitata useampaankin sopimukseen. Huomaa kuitenkin edellä kuvaillut periaatteet hankkeiden raportoimisesta erillisenä tai yhdessä. (Vrt. kohta 4.3 s. 21.)

Tutkimuksen tyyppi. Luokituksia, joiden avulla tilaustutkimukset voidaan erottaa muista tutkimuksista. Soveltavan, perustutkimuksen ja kehitystyön osuudet voidaan tarvittaessa ilmaista. (Vrt. kohta 4.3 s. 23.)

Tutkimuksen vastuuhenkilö. Voidaan antaa riittävällä tarkkuudella: etunimi, sukunimi, asema kyseisellä laitoksella. (Vrt. kohta 4.4 s. 24.)

Organisaatio. Riittävät tiedot korkeakoulusta ja laitoksesta kuten osoite, puhelin, telefax, sähköposti jne. Myös englanninkieliset versiot on mahdollista ja toivottavaakin antaa yksikön nimistä. Suorittavasta laitoksesta yleensä mekaanisesti riippuvat tiedot, kuten tiedekunta ja KOTA-koulutusalakoodi ovat tässä yhteydessä. Näitä tutkijoiden ei yleensä tarvitse itse täyttää. (Vrt. kohta 4.4 s. 25.)

Tiivistelmä. Erikseen on mahdollista antaa tutkimushankkeen pysyväinen kuvailu ja **edistyminen** eli kertomusvuoden lopun tilanne. Tiivistelmä voidaan antaa erikseen myös englanninkielisenä. (Vrt. kohta 4.5 s. 26.)

Avainsanat ja asiasanat. Tutkimusta kuvaavat kontrolloidut asiasanat valitaan Yleinen suomalainen asiasanasto -teoksesta, tarkistamattomat tai asiasanaston ulkopuolelta tulevat sijoitetaan avainsanoiksi. Vastaavat englanninkieliset kontrolloimattomat avainsanat ja UNESCOn SPINES-thesauruksen mukaiset kontrolloidut asiasanat annetaan myös erikseen. (Vrt. kohta 4.5 s. 26.)

Erillinen kenttä on varattu mahdollisille tutkimuksessa käytetyille erityisille **laitteille**. Joillakin tieteenaloilla voidaan tutkimuksia hyödyllisesti kuvailla **menetelmän** avulla. Nämä kaksi tietoa koskevat yleensä vain eräitä tieteenaloja. (Vrt. kohta 4.5 s. 26.)

Luokitukset. Pakollisiksi on valittu opetusministeriön ns. KOTA-oppiaineluokitus ja Euroopan yhteisön nk. CERCS-luokitus. Valinnaisesti voidaan antaa UDK-luokitukset ja Tilastokeskuksen toimialaluokitukset, sekä mikä tahansa muukin tieteenalalla yleisesti käytetty luokitus, kunhan se identifioidaan. (Vrt. kohta 4.6 s. 28.)

Aikataulutiedot. Hankkeen alkamis- ja loppumispäivämäärät. Lisäksi valmiissa tietokannassa on se päivämäärä, jota tietueen tiedot koskevat, mutta tämä tieto voidaan täyttää tietokannan ylläpitäjän toimesta. Tieto tutkimuksen vaiheesta (käynnissä, päättynyt, ...) annetaan myös tässä. (Vrt. kohta 4.7 s. 29.)

Rahatiedot. Tiedot annetaan yleensä *vuosittain* ja erikseen voidaan antaa kuluneet ns. perusvoimavarat ja erityisvoimavarat (jotka tulevat omasta korkeakoulusta) ja erikseen kustakin rahoituslähteestä tulleet varat. Tiedot ovat normaalisti markkoja, mutta voivat olla henkilötyökuukausiakin, milloin jokin firma tai esim. Akatemia kustantaa muista varoista hankkeeseen tutkijan. Koko hankkeen laajuutta arvioidaan yhteenlasketuilla henkilötyökuukausilla, joka on pakollinen tieto. Rahoista on myös mahdollista antaa lisäksi vapaamuotoinen selostus. Rahatietojen pitäisi mieluiten tulla suoraan korkeakoulun kirjanpidon tietojärjestelmistä. Kukin korkeakoulu ja mahdollisesti rahoittaja sanelee omat toivomuksensa tutkimushankkeilleen siten, että tarvitsija saa haluamansa tiedot rahojen käytöstä. Ainoastaan **henkilötyökuukausien summaa** kertomusvuodelta esitetään yleisesti pakolliseksi tiedoksi. (Vrt. kohta 4.8 s. 30.)

Tutkijat. Tutkimukseen osallistuvat ja osallistuneet tutkijat annetaan tarvittavalla tarkkuudella: Sukunimi, etunimet, oppiarvo, mukanaolon alku- ja mahdollinen loppupäivä, sekä asema tutkimuksessa. Kansalaisuutta, sukupuolta ja oppiarvon suoritusvuotta varten on kentät, jos tietoja tarvitaan. (Vrt. kohta 4.9 s. 31.)

Julkaisut. Painetut julkaisut (ei siis käsikirjoituksia) ilmoitetaan riittävällä tarkkuudella, joka voi vaihdella esim. korkeakouluittain. Itse kirjoitetuista tai toimitetuista kirjoista, kokoomateoksissa ja aikakausjulkaisuissa ilmestyneistä artikkeleista annetaan riittävät bibliografiset tiedot, esim. kirjasta vähintään tekijän nimi, kirjan nimi, kustantaja ja julkaisuvuosi, mutta mielellään lisäksi sivumäärä, ISBN-numero jne.[1] Patentit ja julkaistut tietokoneohjelmat tai muut ATK-tallenteet ilmoitetaan antamalla vastaavia tietoja. (Vrt. kohta 4.10 s. 33.)

Opinnäytteet. Opinnäytteet muodostavat olennaisen osan hankkeiden jatkokoulutustehtävän tuloksista ja niistä ilmoitetaan tekijän suku- ja etunimet, laitos, jossa suoritettu, opinnäytteen nimi ja hyväksymisvuosi. (Vrt. kohta 4.11 s. 38.)

Viitteet muihin tutkimuksiin. Tämä on tärkeä kohta, jolla varmistetaan sukulaisprojektien yhteenkuuluvuus sekä ilmoitetaan erilaiset yhteistyötahot. Viitteet annetaan ensi sijassa samanlaisilla sopimuskoodeilla kuin, millä hankkeet itsekin tietueen alkuosassa tunnistetaan. Tarvittaessa voidaan lisätä vapaamuotoista tekstiä yhteistyökertomuksena. (Vrt. kohta 4.12 s. 38.)

[1] Voidaan olettaa, että kullakin korkeakoululla on julkaisurekisteri, johon julkaisutietoja muutenkin kerätään. Tiedot tutkimusrekisteriin olisi yleensä helppo saada julkaisurekisteristä, jos siinä on varattu mahdollisuus antaa tutkimushankkeen identifioiva lisätieto kullekin julkaisulle. Tällöin riittää, että tutkijat antavat julkaisutietonsa vain kertaalleen.

Luku 3

SGML

SGML on kansainvälinen standardi (ISO 8879), jonka avulla voidaan kuvailla rakenteisia dokumentteja. Sen avulla voidaan kuvata esim. kirjan, raportin tai muun dokumentin muodollinen rakenne, siis dokumentin jakautuminen osiinsa. Kirja esimerkiksi jakautuu alkuosan tietoihin, lukuihin, ja lopussa oleviin tietoihin. Luvuilla on otsikko ja tekstiosa jne.

On tärkeää huomata, että SGML:n avulla laadittu dokumentin rakennemääritys DTD (eli Document Type Definition) ei ota kantaa dokumentin ulkoasuun, vaan ainoastaan rakenteeseen. SGML:llä voidaan kuvata myös tietueita ja tietokantoja, vaikka ensimmäiset sovellukset ovat olleet painettavien kirjojen ja sanakirjojen rakenteen kuvauksen alueella.

SGML-formalismista ja sen käytöstä on saatavilla erilaista kirjallisuutta. Sille, joka ryhtyy laatimaan dokumenttien tyyppimäärityksiä tai konversio-ohjelmia, käyttökelpoisimpana voidaan pitää teosta: Charles F. Goldfarb, *The SGML Handbook*, Oxford University Press, 1990. Maallikolle on vaikeampi löytää sopivaa kirjallisuutta, mutta seuraavasta voi olla apua: Eric van Herwijnen, *Practical SGML*, Kluwer Academic Publishers, 1990. Lyhyempiä yleisesittelyitä SGMLS:stä on ollut mm. BYTE-lehdessä (June 1992).

3.1 SGML verrattuna tavanomaisiin tietokantoihin

SGML eroaa tavanomaisista, esim. relaatiotietokannoista siten, että tietojen pituudelle ei ole kiinteitä rajoja ja sikäli, että tietoalkioiden järjestys ja hierarkkisuus on hyvin vapaasti määriteltävissä. Perinteiseen tietokantatyyliin määritellyn tietuemallin, esim. opetusministeriön aiemman tutkimustietokantoja koskevan ohjeen pukeminen SGML:n DTD:ksi onkin yleensä helppoa.

Toisaalta SGML eroaa tavanomaisista tekstitietokannoista kuten TRIPistä siinä, että tekstitietokannoissa on usein melko puutteelliset mahdollisuudet esittää hierarkkisia rakenteita, joissa toistuu itsessään rakenteen omaavia osatietoja (kuten tutkimuksen julkaisutiedot ja rahoitustiedot).[1]

Näiden hyvien ominaisuuksien vastapainoksi on todettava, että SGML on määrittelykieli, eikä itsessään tietokantaohjelma. Se, että tiedot ovat SGML:n muodossa, ei nykyisellään vielä merkitse sitä, että niitä voidaan suoraan hakea jollakin valmiilla ohjelmatuotteella. Lähikuukausien aikana tämän odotetaan tulevan yleisesti mahdolliseksi ns. oliotietokantojen tullessa markkinoille. (Näiden odotetaan tulevaisuudessa korvaavan laajalti nykyisiä relaatiotietokantoja.) Jo nyt on kuitenkin yleisesti ja helposti saatavilla joukko ohjelmia ja työkaluja SGML-muotoisen tekstitiedon käsittelemiseksi: jäsentimiä koodien oikeellisuuden tarkistamiseksi[2] ja tiedon muuntamiseksi toisiin muotoihin[3], editoreita[4] annetun DTD:n muotoisten dokumenttien laatimiseksi ja päivittämiseksi sekä erilaisia tekstinkäsittely- ja julkaisuohjelmia SGML-muotoisen tekstin siistiä tulostamista[5] varten.

3.2 Lyhyt johdatus SGML:ään

Rakenteen ilmaiseminen SGML:n puitteissa perustuu tekstiin lisättävään koodittukseen (tagging), joka on selväkielistä eli näkyvää. Koodit osoittavat tekstialkioiden alkamista ja loppumista, esim. seuraavassa on henkilöä tarkoittava `hlo`-niminen alkio, joka koostuu `sukunimi` ja `etunimet` -nimisistä alkioista:

```
<hlo><sukunimi>Virtanen</sukunimi>
     <etunimet>Artturi I.</etunimet></hlo>
```

Elementti alkaa kulmasuluissa olevalla alkion nimellä ja päättyy vastaavasti kulmasuluissa olevaan alkion nimeen, jonka edessä on kauttaviiva. Dokumentin tyyppimäärityksen eli DTD:n kielellä yllä oleva rakenne vastaa seuraavaa:

[1] TRIP-tekstitietokannoissa näyttää olevan rajoitus, jonka takia vain yhtä itsessään rakennetta sisältävää kenttää voidaan käyttää toistuvana kenttänä.

[2] Aiemmin mainittu SGMLS on ns. Amsterdamin jäsentimen julkinen (public domain) versio, joka toimii mm. Unix-järjestelmissä. Lähes sama ohjelma on myös vapaasti saatavissa MSDOS-käyttöjärjestelmän alaisuudessa toimiville mikrotietokoneille.

[3] Aiemmin mainittu SGMLSASP on saatavissa ilmaisena samoille laitteille kuin SGMLS:kin. Tämä ohjelma ottaa jäsentimen purkaman, täydentämän ja käsittelemän tiedon määrämuodossa ja korvaa siinä olevat tietoalkioiden alku- ja loppumerkinnät halutulla tavalla.

[4] Author/Editor, josta siitäkin mainittiin aiemmin, on saatavissa Macintoshille, MS Windows -ympäristöön ja Unix-koneissa käytettyihin MOTIF- ja Open Look -käyttöliittymiin. Sen avulla on helppo tuottaa varmistetusti oikeanlaista DTD:n mukaista tietoa. Ohjelma näyttää ruudulla tietueen kenttineen rakenteisena. Rakennetta voi selventää ja korostaa käyttämällä sopivia kirjasinlajeja eri tyyppisille kentille. Author/Editor edellyttää ja toisaalta tuottaa täysillä kenttien alku- ja loppumerkeillä varustettua tietoa.

[5] Tavanomaiset Unix-järjestelmissä toimivat Free Software Foundationin groff-ohjelmat, Donald Knuthin TeX ja LaTeX ovat vapaasti saatavissa, ja mitä tahansa niistä voidaan käyttää tuottamaan siistejä tulostuksia tutkimustietokantojen sisältämistä tiedoista. Monet näistä toimivat myös muissa ympäristöissä. Kaupallisista ohjelmista mainittakoon WordPerfectin INTELLITAG-ohjelma, joka muuntaa SGML-dokumentin WordPerfect-tiedostoksi ja lisää siihen määrittelyn mukaisen muotoilun.

```
<!ELEMENT   hlo                - o (sukunimi,etunimet)>
<!ELEMENT (sukunimi|etunimet) o o (#PCDATA)>
```

Alkiot, jotka eivät jakaudu edelleen osiinsa (kuten `sukunimi` ja `etunimi`) merkitään `#PCDATA` -tyyppiseksi (Parsed Character Data). Alkio `hlo` puolestaan määrityksen mukaan jakautuu osa-alkioihin `sukunimi` ja `etunimi`.

Esimerkin kooditus oli täydellistä, eli jokaisen alkion alkaminen ja loppuminen oli nimenomaisesti merkittynä. SGML antaa mahdollisuuden jättää pois sellaisia alku- ja loppukoodeja, jotka voidaan yksiselitteisesti päätellä asiayhteydestä. Esimerkki voitaisiin kirjoittaa lyhyemmin monellakin eri tavalla, esim.:

```
<hlo>Virtanen<etunimet>Artturi I.</hlo>
```

Koodien poisjättämisen sallimista tai kieltämistä merkitään elementin määrittelyssä elementin nimen perässä olevalla merkinnällä (`- o`) tai (`o o`). Ensimmäinen merkki vastaa alkukoodin pakollisuutta (`-`) tai valinnaisuutta (`o`), ja jälkimmäinen vastaavasti loppukoodin poisjättöä eli minimointia.

Alkiot voivat sisältää alkioita, jotka voivat puolestaan sisältää alkioita, ilman rajoitusta rakenteen syvyydelle. Samannimisiä alkioita, kuten esimerkin `hlo` voi esiintyä rakenteisen dokumentin eri osissa, kuten juuri tämä alkio tutkimustietueessa vastuuhenkilönä, hankkeen tutkijoina ja julkaisujen tekijöinä.

Joitakin tietoalkioita voi olla toistettuna useampiakin kappaleita, esim.

```
<!ELEMENT tutkijat - o (tutkija)+>
```

Tässä `tutkijat` alkio koostuu yhdestä tai useammasta `tutkija` -tyyppisestä alkiosta. Jos plus-merkin sijasta olisi käytetty tähteä (`*`), olisi sallittu nolla, yksi tai useampia alkioita. Kysymysmerkillä puolestaan merkitään valinnaista alkiota, joita saa olla yksi tai nolla kappaletta, esim.:

```
<!ELEMENT pvm o o (pp?,kk?,vv)>
```

Tässä päivämäärässä `pvm` saa olla päivä `pp` tai se saa puuttua, samoin kuukausi, mutta vuosi `vv` on pakollinen.

SGML sallii muunkinlaisia lyhennysmerkintöjä. On esimerkiksi helppo määritellä, että aiemman esimerkin sukunimen ja etunimen välissä oleva pilkku (`,`) merkitsee `sukunimi`-alkion sisällä sen loppumista ja `etunimi`-alkion alkua. Esimerkki yksinkertaistuisi täten edelleen. Emme kuitenkaan käytä tätä ns. SHORTREF-mekanismia tässä raportissa pitääksemme esityksen selkeänä (tinkien siis tietueiden koodituksen lyhyydestä). Lyhennyskäytäntöjen käyttö on tietenkin mahdollista, eikä ole ristiriidassa tämän suosituksen kanssa, koska sekä optimoinnit että lyhennysmerkinnät voidaan ohjelmallisesti muuntaa täydelliseksi koodituksesi.

Tiedot, jotka kuuluvat kuvattavan tietueen näkyvään osaan, esitetään yleensä alkioina. Jotkut tiedot eivät ole luonteeltaan osatietoja, vaan kuvaavat koko tietueen ylläpitoon liittyviä ominaisuuksia tms. Näitä varten SGML:ssä on ns. attribuuttimekanismi. Attribuuteille voidaan antaa oletusarvo ja tarkemmin rajattu arvoalue.

Tässä suosituksessa on käytetty attribuutteja säästäen lähinnä, koska niiden formalismi on jonkin verran monimutkaisempi kuin alkioiden.

Tietoalkioiden sisällössä voidaan käyttää lyhenteitä eli olioita (entity), jotka toimivat kuten makrot. Lyhenne edustaa sille määriteltyä aukikirjoitettua tekstiä, jollaiseksi se korvataan käsittelyn kuluessa. Oliot merkitään edeltävällä et-merkillä ja lopussa olevalla puolipisteellä, esim:

```
<orgnimi>&HY;</orgnimi>
```

Tässä käytetään seuraavalla tavalla määriteltyä SGML-oliota:

```
<!entity HY "Helsingin yliopisto">
```

Lyhenteet avataan kulloisenkin DTD-määritelmän mukaisesti, ja siten sama lyhenne voidaan yhtä tarkoitusta varten avata suomenkieliseksi termiksi, ja toista varten vaikkapa sen englanninkieliseksi vastineeksi. SGML-olioita käytetään tutkimustietokannan määrittelyssä jossakin määrin tällaista käyttöä ajatellen, jotta puhtaan SGML:n puitteissa saadaan jonkin verran yleisempi ja monikäyttöisempi tietosisältö.

Alkioiden ja olioiden nimeämiseen liittyy tiettyjä rajoituksia. Nimet saavat olla korkeintaan kahdeksan merkkiä pitkiä ja niiden sallittu merkistö koostuu englantilaisista aakkosista, numeroista ja muutamista välimerkeistä. Nämä rajoitukset häiritsevät suomen kielen osalta sen tavoitteen toteutumista että alkioiden nimeämisen tulisi olla mahdollisimman havainnollista.

3.3 Alkioiden pakollisuus

Dokumentin tyyppimääritys (DTD) ottaa kantaa tietoalkioiden vapaaehtoisuuteen tai pakollisuuteen. Tämä kannanotto on tekninen siten että tietue, josta pakollinen kenttä puuttuu tulee hylätyksi, eli ohjelmat eivät ota sellaista käsiteltäväksi, vaan antavat virheilmoituksen tms. Tällaisen teknisen pakollisuuden lisäksi on erikseen ohjeita siitä, mitä tietoja tietoalkioihin pitäisi sisällyttää. Nämä ohjeet ovat tutkimustietokannan määrittelyssä varsinaisia DTD-määritelmiä edeltävissä sanallisissa selityksissä. Näissä pakollisuus pyritään nimenomaisesti mainitsemaan.

On luonnollista, että laajempi tietoalkio, esim. artikkelijulkaisu voi olla valinnainen, mutta jos sellainen alkio on sisällytetty, sen tietyt osat (esim. julkaisuvuosi) voivat olla pakollisia kullekin annetulle artikkelijulkaisulle.

Pakollisuus tämän suosituksen puitteissa tarkoittaa useimmiten sitä, että koko korkeakoulun koottuja tietoja yhteiseen muotoon siirrettäessä kyseisen tietoalkion täytyy olla mukana. Korkeakoulun sisällä, kun tietoja kerätään laitoksilta tiedekuntiin tai keskushallintoon, tilanne voi olla toinen niiden tietojen osalta, jotka on tavalla tai toisella pääteltävissä tai ohjelmallisesti lisättävissä. Korkeakoulujenhan ei ole edellytetty käyttävän suosituksen määrityksiä sisäisesti, vaan ne voivat toteuttaa keruun esim. jollakin tekstitietokantaohjelmalla. Tällainen muu ohjelma ja korkeakoulun oma, hieman poikkeava SGML:n mukainen DTD ja paikalliset ohjeet tiedon keräämiseksi ovat kunkin korkeakoulun sisäisiä asioita.

Tämän suosituksen teknisesti pakollisiksi luokittelemien tietojen tulisi muodostaa ehdoton minimivalikoima olennaisimmista tiedoista jotka edellytetään aina annettaviksi kussakin tutkimushankkeessa.

Tekninen pakollisuus käy ilmi SGML-määrittelystä seuraavasti:

- Jos alkion nimeä ei seuraa mikään merkeistä (*, +, ?), elementti on pakollinen ja sitä pitää olla tasan yksi kappale.
- Jos elementin nimeä seuraa plus-merkki (+), elementti on pakollinen ja sitä pitää olla vähintään yksi kappale, mutta saa olla useampiakin.
- Jos elementin nimeä seuraa kysymysmerkki (?), elementti on vapaaehtoinen ja sitä saa olla korkeintaan yksi kappale.
- Jos elementin nimeä seuraa asteriski (*), elementti on vapaaehtoinen ja sitä saa olla nolla, yksi tai useampia kappaleita.

Muistettakoon, että silloin, kun tutkimushankkeet antavat tietoja korkeakoulunsa tietokantaan, voivat monet tässä tietokannan sisällön määrittelyssä olevat pakollisiksi luokitellut tiedot olla vielä valinnaisia, jos niiden arvot lisätään esim. automaattisesti ennen, kuin tietokantaa muunnetaan lopulliseen siirtomuotoon.

Luku 4

Tietosisältö

Tässä ehdotettu vaihtoformaatti ei aseta rajoituksia kenttien pituuksille, eikä alkioiden lukumäärille erilaisissa luetteloissa. Se pyrkii muutenkin olemaan salliva siten, että eri korkeakouluissa laaditut tietueet voisivat sopia siihen. Lukuisten tietoalkioiden salliminen ei sinänsä merkitse niiden vaatimista, vaan sitä, että tällaisille hyödyllisinä pidetyille tiedoille on paikkansa, jos niitä on kerätty. Toivomme, että vain harvoissa tapauksissa on tarpeen jättää pois tietoalkioita yhteisestä muodosta, jos ne on katsottu tarpeelliseksi paikallisesti. Tällöin kaksisuuntainen muunnos on mahdollinen ilman tietojen häviämistä tai huononemista.

Tietosisällön määrittely sisältää kaksi puolta: (1) täsmällisen SGML-standardin mukaisen rakenteisen tiedon määrittelyn (ns. DTD:n) tutkimusta kuvaaville tietueille sekä (2) selityksiä jotka valottavat ja täsmentävät eri tietojen sisältöä, tulkintaa ja muotoa. Alaviitteissä viitataan toistuvasti kahteen keskeiseen asiakirjaan:

- "EY:n suositukseksi" kutsutaan EY:n komission suositusta 91/337/CEE tutkimustietokantojen sisällöksi ("Commission recommendations of 6 May 1991 concerning the harmonization within the Community of research and technological development databases", Official Journal of the European Communities, L 189, Volume 34, 13 July 1991). EY:n suosituksesta käytetään jatkossa usein lyhennettä "CERIF".

- "OPM:n suositukseksi" kutsutaan opetusministeriön 10.3.1989 päivättyä aloitetta korkeakouluille "Ohjeet tutkimusrekisterien laatimiseksi".

Osat etenevät vuorottelevassa tahdissa: ensin selitys ja sen jälkeen tarkka SGML-määritelmä. [1] Selitykset ovat normaalia tekstiä, kun taas määritelmät ovat vaakaviivojen välissä kirjoituskonekirjasimilla esitettynä. Kutakin selitystä vastaavan SGML-tietoelementin nimi on aina hakasulkeissa `[ ]` siinä kohdassa josta siihen selityksessä viitataan. Jos elementin nimi on täsmälleen sama kuin tekstissä oleva viittaus ei erillistä viitettä anneta.

[1] Tähän raporttiin jaksoina sisältyvä tutkimustietokannan DTD-määrittely on eristetty mekaanisesti erilliseksi tiedostoksi, jota on käytetty testiajoissa ja validisoinneissa.

Lukija voi aluksi jättää teknisen SGML-määrityksen pintapuolisen silmäilyn varaan, ja vasta myöhemmillä lukukerroilla pyrkiä selvittämään rakenteen teknistä puolta itselleen.

4.1 Ylimmän tason rakenne

Ylimmällä tasolla tietokannan määrittely ilmaisee jokseenkin triviaalin asian, että **tu**tkimus**h**anke**ti**etokanta `tuhti` koostuu yhdestä tai useammasta tietueesta `tutkimus`. Toisto ilmaistaan elementin jäljessä olevalla plus-merkillä.

Toinen, vähemmän triviaali määritelmän osa sanoo, että tietokannassa saa olla huomautuksia eli `komm`-elementtejä vapaasti missä tahansa. Ilmeisen tarkoituksensa lisäksi tämä antaa mahdollisuuden sijoittaa sellaisia tietoja tietokantaan, joille ei muuten ole varattu omaa paikkaa. Tieto saadaan itsensä selittäväksi esim. lisäämällä siihen sopiva vakioteksti alkuun tunnisteeksi.

```
<!ELEMENT tuhti    - - (tutkimus)+ +(komm)>
```

Yhtä tutkimushanketta koskevat tiedot on pyritty järjestämään siten, että ne paperille tulostettuna sopivasti muotoiltuna muodostaisivat käyttökelpoisen kuvailun hankkeesta. Tässä suhteessa ehdotus poikkeaa opetusministeriön aiemmasta suosituksesta, joissa kaikki pakolliset kentät luetellaan ennen valinnaisia. Ratkaisu heijastaa myös irrottautumista tavanomaisesta kiinteänmittaisiin kenttiin perustuvasta tietokantamallista.

Kunkin tutkimushankkeen tiedot koostuvat seuraavista alkioista, joista osa jakautuu edelleen osiin.

- Tutkimuksen nimi suomeksi tai ruotsiksi `[otsake]`.
- Tutkimuksen nimi englanniksi `[title]`.
- Sopimus- tai projektikoodit `[tutksop]`.
- Tutkimuksen tyyppi - onko tilattu, kehitystyön osuus, soveltavan tutkimuksen osuus `[tyyppi]`.
- Vastuuhenkilö `[vasthenk]`.
- Tutkimusta tekevän organisaation tiedot - laitos, osoite, ym. `[tutkorg]`.
- Tutkimuksen sisältöä koskevat tiedot - tiivistelmät, avainsanat, luokitukset `[sisalto]`.
- Tutkimuksen aikataulutiedot `[aikat]`.
- Rahoitustiedot `[rahoitus]`.

- Tutkimuksessa työskentelevät tutkijat [`tutkijat`].
- Tutkimukseen liittyvät julkaisut [`julk`].
- Opinnäytetyöt [`opinnayt`].
- Osa- ja jatkotutkimukset sekä yhteistyötahot [`yhtstyo`].
- Mahdollisia muita tietoja [`muuta`].

Jotkut näistä tiedoista koostuvat useammista osista, toiset ovat yksinkertaisia kenttiä. Kukin selitetään jatkossa tarkemmin.

```
<!ELEMENT tutkimus - - (otsake, title, tutksop*,  tyyppi?,
                        vasthenk, tutkorg, sisalto, aikat,
                        rahoitus, tutkijat, julk?, opinnayt?,
                        yhtstyo?, muuta?)>
```

4.2 Tietuetta koskevat attribuutit

Yllä mainittujen tietoalkioiden lisäksi tutkimustietueeseen liittyy erinäisiä tietoja jotka eivät varsinaisesti kerro tutkimushankkeesta vaan pikemminkin tietueesta itsestään. Nämä ovat ns. attribuutteja joilla on attribuutin nimi ja arvo. Teknisesti näihin liittyy myös se mahdollisuus, että niille voi asettaa oletusarvoja ja mahdollisten arvojen luetteloita. Kaikki attribuutit pitäisi voida päätellä automaattisesti siten, että tietokannan loppukäyttäjän ei tarvitse kiinnittää niihin huomiota.

Tietuenumero [`id`]. Tietueen yksiselitteinen tunnistenumero [2] voi olla käytöltään pelkästään tekninen ja numero voidaan generoida automaattisesti. Mikä tahansa systemaattisesti käytetty tunnusnumero joka erottaa korkeakoulun tutkimukset toisistaan käy tähän.

Tietokannan omistaja [`tietok`]. Sen tietokannan, eli käytännössä sen korkeakoulun tunnus, jonka tutkimustietokannasta tietueen tiedot ovat peräisin.[3] Tämä arvo generoitaisiin normaalisti ohjelman avulla silloin, kun korkeakoulun tietoja siirretään yhteiseen vaihtoformaattiin. Koska tätä tietoa ei ole tarkoitettu käyttäjien luettavaksi, siinä voidaan käyttää opetusministeriön KOTA-tilastoissa käyttämiä lyhenteitä.[4]

Tutkimustietueen kieli [`kieli`]. Se kieli, jolla tutkimustietueen tiedot on annettu. [5] Käytännössä otsikko, tiivistelmä ja avainsanat annetaan tässä ilmoitetulla

[2] OPM suosituksessa `NR`, joka koostuisi usein erilaisista merkityksellisistä osista. EY:n CERIF-suosituksen kenttä 11, *Record control number* identifioi tietueen.

[3] OPM:n suositus ei tunne tätä, EY:n suosituksen kenttä 00, *Database owner*.

[4] Esim.: *KOTA: Tietoa korkeakoulujen toimintaa kuvaavasta tietokannasta 1991*, Opetusministeriö, Korkeakoulu- ja tiedeosasto, 1992, sivu 8, vasemmalla olevat kirjaintunnukset.

[5] Alkio vastaa EY:n suosituksen kenttää 13, *Original working language of the record*.

kielellä, joka voi olla suomi tai ruotsi. Oletusarvona voidaan pitää korkeakoulun ensisijaista kieltä: suomea `fi` tai ruotsia `sv`.[6] Paikallisesti voidaan olettaa, että korkeakoulun virallinen tai yleisempi kieli on oletus. Tutkimuksista, joiden kieli poikkeaa oletusarvosta, kieli annetaan eksplisiittisesti.

Tietueen täyttöaste `[taytto]`. Tämä kolmiarvoinen tieto voidaan generoida joko automaattisesti tai esimerkiksi informaatikon toimesta: [7]

- kaikki tietueen pakolliset ja valinnaiset tietoelementit on annettu: `A`,
- kaikki pakolliset tietoelementit on annettu: `B`,
- kaikkia pakollisiakaan tietoelementtejä ei ole annettu: `C`.

Tietueen luontipäivä `[luontpvm]`. Tämä voidaan generoida automaattisesti.

Tietueen viimeisen päivityksen päivämäärä `[paivpvm]`. Se päivämäärä, jolloin tietueen tietoja on viimeksi päivitetty.[8] Tieto voidaan normaalisti generoida automaattisesti. (Vrt. myöhempänä oleva `[tietopvm]` -kenttä, joka ilmaisee, mitä ajankohtaa tiedot kuvaavat.)

```
<!--     elementti attr    arvoalue                     oletusarvo
-->
<!ATTLIST tutkimus  id      ID                           #IMPLIED
                    tietok  (HY | JY | OY | JOY |
                             KUY | TY | TAY | ÅA | VY |
                             LY | EKK | TKK | TTKK |
                             LTKK | HKKK | SHH | TUKKK |
                             TAIK | SIBA | TEAK |
                             KUVA | MUU)                 "HY"
                    kieli   (fi | sv)                    "fi"
                    taytto  (A | B | C)                  "B"
                    luontpvm CDATA                       #IMPLIED
                    paivpvm  CDATA                       #IMPLIED>
```

Tietokannan omistajan oletusarvona kukin korkeakoulu pitää luonnollisesti omaa lyhennettään, samoin esim Åbo Akademi kielikoodina ruotsin kieltä `sv`.

4.3 Tunnistetiedot

Tutkimuksen nimi `[otsake]`. Tutkimuksen nimi eli otsikko[9] tutkimustietueen kielellä, siis suomeksi tai ruotsiksi — vrt. attribuutti `kieli`. Useimmiten tutkimuksella on tutkimussopimuksen perusteella virallinen nimi, ja tällöin sitä tulee käyttää.

[6] Nämä kaksikirjaimiset kielikoodit ovat EY:n CERIF-suosituksen mukaiset.

[7] EY:n suosituksen kenttä 02, *Level of completeness code*

[8] EY:n suosituksen kenttä 01, *Date of last updating of the record*. Siellä edellytetään ISO 2014:n muotoista päiväystä YYYYMMDD.

[9] OPM:n suosituksen `NI`, EY:n CERIF-suosituksen kenttä 21.2, *Original project title*.

Jos nimeä ei ole ennestään kiinnitetty, valitaan mahdollisimman informatiivinen nimi — mitä tutkitaan, ja miksi. Pakollinen tietoelementti.

```
<!ELEMENT otsake   - o (#PCDATA) >
```

Tutkimuksen nimi englanniksi `[title]`. Tutkimustietueen toinen kieli on tämän ja aikaisemman suositusten mukaan aina englanti.[10] Tietoelementti on pakollinen.

```
<!ELEMENT title    - o (#PCDATA)>
```

Tutkimussopimuksen tiedot `[tutksop]`. Tutkimussopimus, joka yksilöi hankkeen yleispätevällä tavalla ja jota voidaan pitää yhtenä tärkeimmistä tunnisteista.[11] Tietue voi sisältää tiedot useistakin tutkimussopimuksista. Kukin tieto koostuu kahdesta osasta:

- Sopimuksen toinen osapuoli `[osapuoli]`: tutkimusta rahoittava organisaatio, tilaaja, tai sellaisen puuttuessa se yksikkö, jossa tutkimus suoritetaan. Korkeakouluista ja tunnetuimmista rahoittajista käytetään sovittuja lyhenteitä ja mieluiten alla määriteltyjä SGML-olioita.[12] Mikäli saman sopimuksen toisia osapuolia on useita, annetaan useampi kenttä.
- Sopimuskoodi `[sopkoodi]`: sopimuksen numero, diaarinumero tai projektinumero. Sellaisen puuttuessa sopimuksen päivämäärä tai muu riittävän yksiselitteinen (esim. juokseva) numero kelpaa.

Jos molempia tietoja ei ole saatavissa, toinen riittää, kunhan se on oletettavasti yksiselitteinen valtakunnallisestikin.

Jos kyseessä on tilaustutkimus, on ensimmäisenä mainitun tutkimussopimuksen osapuolen oltava tutkimuksen tilaaja.

Mahdolliset muut sopimukset asetetaan järjestykseen toisen osapuolen organisaation laajuuden mukaan, alkaen suurista kansainvälisistä yhteisöistä (esim. YK, EY)

[10] OPM:n suosituksen kenttä `TI`. EY:n CERIF-suositus edellyttää tutkimuksen nimen ja tiivistelmän myös muulla EY:n kielellä kuin tutkimuksen alkuperäisellä kielellä. Tutkimuksen nimi muilla kielillä on CERIFin kentässä 22, *Project title — other language(s)*. CERIFissä annetaan osakenttänä kielikoodi, jota tässä ei tarvita, koska toinen kieli on aina sama. (Lisäksi CERIFin mukaan kaikki tietueessa käytetyt kielet paitsi alkuperäinen kieli luetellaan kentässä 14, *Other languages of the record*).

[11] OPM:n suosituksessa kenttä `SO`, *sopimus- tai projektikoodi*, jonka diaarinumeroon tai tutkimussopimuksen numeroon osapuolen lyhenne voisi sisältyä. EY:n suosituksessa lähinnä vastaava kenttä on numero 12, *Project identifier*, joka muotoillaan kansallisen tai tietokantaan liittyvän käytännön mukaan ja varustetaan maakoodilla. CERIFin mukaan mikä tahansa yksiselitteinen tunnus riittäisi.

[12] Esim. `&HY;` Helsingin yliopisto, `&OY;` Oulun yliopisto, `&TKK;` Teknillinen korkeakoulu, ..., ja `&SA.HUM;` Suomen Akatemia/Valtion humanistinen toimikunta, `&TEKES;` jne.

ja päätyen pieniin kotimaisiin (esim. oma yliopisto). Koska kaikilla tutkimuksilla ei voi olettaa olevan sopimusta, eivät nämä tiedot ole pakollisia.

Tutkimussopimuksen tunnisteiden avulla on voitava löytää samaan laajempaan hankkeeseen kuuluvat eri korkeakouluissa suoritettavat osat. Yhdistävänä voi olla yksinkertaisesti sama sopimuskoodi tai tietyn systematiikan mukaan rakennetut koodit, joissa on osahankkeille yhteinen osa.

```
<!entity SA  "Suomen Akatemia">
<!entity SKR "Suomen kulttuurirahasto">
<!entity HY  "Helsingin yliopisto">
<!entity JY  "Jyväskylän yliopisto">
<!entity OY  "Oulun yliopisto">
<!entity JOY "Joensuun yliopisto">
<!entity KUY "Kuopion yliopisto">
<!entity TY  "Turun yliopisto">
<!entity TAY "Tampereen yliopisto">
<!entity AA  "Åbo Akademi">
<!entity VY  "Vaasan yliopisto">
<!entity LY  "Lapin yliopisto">
<!entity EKK "Eläinlääketieteellinen korkeakoulu">
<!entity TKK "Teknillinen korkeakoulu">
<!entity TTKK "Tampereen teknillinen korkeakoulu">
<!entity LTKK "Lappeenrannan teknillinen korkeakoulu">
<!entity HKKK "Helsingin kauppakorkeakoulu">
<!entity SHH  "Svenska handelshögskolan">
<!entity TUKKK "Turun kauppakorkeakoulu">
<!entity TAIK "Taideteollinen korkeakoulu">
<!entity SIBA "Sibelius-Akatemia">
<!entity TEAK "Teatterikorkeakoulu">
<!entity KUVA "Kuvataideakatemia">
```

```
<!ELEMENT tutksop  - o (osapuoli*, sopkoodi?)>
<!ELEMENT (osapuoli, sopkoodi)
                   - o (#PCDATA)>
```

Tutkimuksen tyyppi `[tyyppi]`.Tyyppiä koskevat tiedot ovat siinä mielessä pakollisia että niillä on määrätty arvoalue, ja jos tietoa ei anneta, annetaan niille ennalta määrätty arvo.

- Onko tilaustutkimus? `[tilattu]`. Oletusarvona on se, että kyseessä ei ole tilaustutkimus, jolloin koko kenttä jätetään pois (tai tyhjäksi). Jos kyseessä on tilaustutkimus, arvona on SGML-vakio `&TIL;`. Silloin myös kohdassa *sopimus- tai projektikoodi* on ensimmäisen sopimuksen osapuolen oltava sopimuksen tilaaja.

- Soveltavan tutkimuksen osuus `[sovpros]`. Tässä annetaan kokonaislukuna soveltavan tutkimuksen prosenttiosuus, ellei se ole nolla. [13]

- Kehitystyön osuus `[kehpros]`. Tässä puolestaan kehitystyön osuus prosentteina, ellei se ole nolla. Loput, eli se mikä ei ole soveltavaa tutkimusta tai kehitystyötä on perustutkimusta. Oletusarvona on että koko tutkimus on perustutkimusta, ja jos tästä poiketaan joko soveltavan tutkimuksen tai kehitystyön suuntaan, on esitettävä arvio poikkeaman osuudesta. Arvio esitetään kokonaislukuna, prosenttiosuutena koko tutkimuksen työmäärästä. Näin ollen kaikkien kolmen osuuden yhteenlaskettu osuus on aina tasan sata.

```
<!ELEMENT tyyppi   - o (tilattu?, sovpros?, kehpros?)>
<!entity  TIL          "tilaustutkimus">
<!ELEMENT tilattu  - o (#PCDATA)>
<!ELEMENT sovpros  - o (#PCDATA)>
<!ELEMENT kehpros  - o (#PCDATA)>
```

4.4 Johto ja laitos

Tutkimuksen vastuuhenkilö `[vasthenk]`. Vastuuhenkilö eli vastuullinen tutkija on yleensä myös yksi tutkimussopimuksen allekirjoittajista.[14] Vastuuhenkilön nimitiedot ovat pakolliset, mutta tieto hänen asemastaan tutkimusta suorittavassa organisaatiossa (ei siis itse tutkimushankkeessa) on vapaaehtoinen.

- Sukunimi `[sukunimi]`

- Etunimet `[etunimet]`

- Asema tutkimusta suorittavassa organisaatiossa `[asema]` — esim. "paleontologian prof."

Vastuuhenkilön nimestä, kuten muistakin henkilöistä erotetaan kaksi osaa: sukunimi ja toisaalta etunimet yhtenä kenttänä.[15]

[13] OPM:n suosituksen kenttä `TT`, tutkimuksen tyyppi kertoo onko tutkimus perustutkimusta, soveltavaa tutkimusta vai kehitystyötä

[14] OPM:n suosituksen kenttä `TU`, tutkija tai vastuuhenkilö. EY:n suosituksen kenttä 31, *Project director*.

[15] Konversio, joka poimii ensimmäisen tai vastaavasti viimeisen osan sukunimeksi toimii yleensä oikein. On huomattava kuitenkin, ettei "Sågvall Hein Anna Lena" tai "Anna Lena Sågvall Hein" -muodoista voi aina päätellä mikä osa on sukunimeä. Sen sijaan pilkulla varustetut muodot ovat aina yksiselitteisiä: "Sågvall Hein, Anna Lena". Sekä OPM:n suositus että EY:n CERIF edellyttävät sukunimeä ensin sekä pilkun käyttöä, eli mahdollistavat helpon mekaanisen konversion.

```
<!ELEMENT vasthenk - o (hlo, asema?)>
<!ELEMENT hlo      o o (sukunimi, etunimet)>
<!ELEMENT (sukunimi, etunimet, asema)
                   o o (#PCDATA)>
```

Organisaatio. Sen organisaation (korkeakoulun laitoksen tai tutkimuslaitoksen osaston tms.) tiedot, jossa tutkimusta tehdään. [16]

Näihin kuuluvat:

- Laajemman tutkimusorganisaation nimi[17]

  `[orgnimi]`, (korkeakoulu, tutkimuslaitos, ...). Tässä käytetään mielellään lyhyttä SGML-oliota (esim. `&JY;`). Pakollinen tieto.

- Laajemman tutkimusorganisaation nimi englanniksi `[orgname]`.
- Tutkimusta tekevän laitoksen[18] tai vastaavan[19] yksikön nimi `[laitos]`. Pakollinen tieto.
- Tutkimusta tekevän laitoksen tai vastaavan yksikön nimi englanniksi `[dept]`.
- Mahdollisen alayksikön — ryhmä, laboratorio tms. — nimi[20] `[alayks]`.
- Postiosoite[21] `[postios]`.
- Sijainti, ellei käy ilmi postiosoitteesta `[sijainti]`.
- Puhelinnumero[22] `[puh]`. Kotimaisista puhelinnumeroista annetaan suuntanumero puhelinnumeron edessä miinusmerkillä erotettuna. Mahdollisissa ulkomaisissa numeroissa annetaan maakoodi plus-merkin edeltämänä alussa.
- Telefax-numero[23] `[fax]`.

[16] OPM:n suositus kokoaa kaikki tunnistetiedot yhteen kenttään `YT`, johon laitetaan organisaation nimi, tutkimusta tekevä laitos tai osasto, osoite ja puhelinnumero. EY:n CERIF sisältää kolme kenttää (50, 51 ja 60) ja niiden alakenttiä. Näiden vastaavuudet seuraavissa alaviitteissä.

[17] CERIFin osakenttä 50.1, *Name of the parent organization*.

[18] Vastaa osittain EY:n CERIFin osakenttää 50.3, *research unit, intermediate levels*. CERIF määrittelee kolme tasoa: *parent organization*, *intermdiate levels* ja *most specific*. Jos tämän suosituksen kaikki kolme tasoa on annettu, ne vastaavat tässä järjestyksessä CERIFin em. kenttiä. Jos **`alayks`** puuttuu, **`orgnimi`** vastaa *parent organization* -kenttää ja **`laitos`** *most specific* -kenttää. Laitoksen taso saattaa siten olla joko *intermediate* tai *most specific*.

[19] Esim. jos tiedekunnassa ei ole laitoksia, annetaan itse tiedekunta.

[20] EY:n CERIF-suosituksen osakenttä 50.4, *Research unit (most specific)*.

[21] EY:n CERIFin osakenttiä 51.1, *address* ja 51.2, *town*. CERIF-suosituksessa on lisäksi osakenttä 51.3, *country code*.

[22] EY:n CERIFin osakenttä 60.1.

[23] EY:n suosituksen kenttä 60.2, CERIFissä suositellaan myös telex-numeroa kenttään 60.3.

- Sähköpostiosoite[24] Internet-muodossa `[email]`.
- Lisätiedot`[muuta]`.
- Laitoskoodi[25] `[ltskoodi]`. Vastuualuekoodi tai muu korkeakoulun oman hallinnon käyttämä lyhyt tunnus, joka erottaa laitokset tai tulosyksiköt toisistaan. Valtakunnallisella tasalla valinnainen, mutta todennäköisesti kussakin korkeakoulussa vaadittu tieto.
- Tiedekunta tai korkeakoulun osasto[26] `[tdk]`.
- KOTA-koulutusalakoodi `[koulala]`. Tämä määräytyy laitoksessa tai lähinnä suuremmassa yksikössä valmistuvien opiskelijoiden perustutkinnon mukaan.[27] Pakollinen tieto.

Osoitetiedot voivat olla laitokselle, ryhmään tai vaikka vastuuhenkilölle riippuen siitä mikä on paras tapa päästä yhteyteen tutkimuksesta jotain tietävien henkilöiden kanssa.

```
<!ELEMENT tutkorg  - o (orgnimi, orgname?, laitos, dept?,
                        alayks?, postios?, sijainti?,
                        puh?, fax?, email?, muuta?,
                        ltskoodi?, tdk?, koulala)>
<!ELEMENT postios  - o (rivi+) >
<!ELEMENT (orgnimi, orgname, rivi, sijainti, email, puh,
           fax, laitos, dept, ltskoodi, tdk, koulala, alayks)
                   - o (#PCDATA)>
```

4.5 Kuvailu ja avainsanat

Tutkimuksen sisällön kuvailu `[sisalto]` koostuu seuraavista tiedoista:

- **Tiivistelmä** tutkimustietueen omalla kielellä[28]

  `[tiivis]`. Tässä kuvataan tutkimushankkeen suunnitelma ja tavoitteet, siis seikkoja, jotka yleensä eivät juuri muutu tutkimuksen kuluessa. Pakollinen tieto.

[24] CERIFin osakenttä 60.4.

[25] Koodi on luultavimmin hyödyllisimmillään korkeakoulun sisäisessä käytössä, jossa se varmaan usein määrätään pakolliseksi, koska siitä voidaan usein päätellä `laitos`, `dept`, `postios`, `tdk` ja `koulala`. Valtakunnan tasolla sen arvo on vähäisempi, se on jo työnsä tehnyt.

[26] Voidaan useimmiten päätellä mekaanisesti laitoskoodista, jos se on käytettävissä.Vastaa osittain CERIFin kenttää 50.3.

[27] Koodi luokittelee siis tutkimusta suorittavan organisaation hallinnollisesti, eikä sen tarvitse aina kuvailla itse tutkimuksen laatua. Se päätellään usein laitoskoodin perusteella automaattisesti (vrt. `kotaoppi`, joka luokittelee tutkimushanketta).

[28] OPM:n suosituksen kenttä `TV`, EY:n CERIFin kenttä 23, *Project abstract in original language.*

- Tutkimuksen **edistyminen** `[edist]`. Tutkimuksen tilanne raportointikauden lopussa. Tämä selostus on luontevimmin se, jossa rahoittajalle tai muulle hallinnolle kerrotaan kertomusvuoden tuloksista.
- **Tiivistelmä englanniksi**[29] `[abstract]`. Suositellaan täytettäväksi.
- **Avain- ja asiasanat** tutkimustietueen kielellä `[avainsan]`. Koostuu avainsanoista[30] `[av]` ja asiasanoista[31] `[as]`. Avain- ja asiasanat suositellaan annettavaksi oman tieteenalan yleisen käytännön mukaisesti. Selvän käytännön puuttuessa täsmällisemmin kuvailevat avainsanat annetaan ensiksi ja laajemmat viimeksi. Asiasanat ovat *Yleinen suomalainen asiasanasto*, Helsingin yliopiston kirjasto, 1994 -julkaisun mukaisia kontrolloituja avainsanoja. Muut eli pelkät avainsanat ovat vapaasti itse valittuja tutkimusta mahdollisimman hyvin kuvailevia avainsanoja. Pakollinen tieto.
- **Avain- ja asiasanat englanniksi**[32] `[keywords]`. Koostuu avainsanoista `[kw]` ja asiasanoista `[sp]`. Englanninkieliset kontrolloidut asiasanat tulee valita UNESCOn SPINES-thesauruksesta.[33]
- Tutkimuksessa käytetty erikoislaitteisto[34] `[laitteet]`. Tässä ei ole syytä ilmoittaa tavanomaisia laitteita, kuten mikrotietokoneita, tms.
- Tutkimuksessa käytetty **menetelmä** `[menetel]`. Tätä kenttää täytetään sikäli, kun tieteenalalla on annettavissa lyhyt ja informatiivinen menetelmätieto. Useilla aloilla tästä kentästä ei ole hyötyä. Menetelmän lisäksi voidaan antaa käytetty mittari tms. Tämäkin on kovin tieteenalakohtaista.

```
<!ELEMENT sisalto  - o (tiivis, edist?, abstract?,
                        avainsan, keywords?, laitteet?,
                        menetel?, luokitus)>
<!ELEMENT avainsan - o (av | as)+>
<!ELEMENT keywords - o (kw | sp)+>
<!ELEMENT (tiivis, edist, abstract, laitteet, av, menetel,
           as, kw, sp)
                   - o (#PCDATA)>
```

[29] OPM:n suosituksen kenttä `AB`, CERIFin kenttä 24, *Project abstract in other language(s)*.

[30] OPM:n suositus ei erottele avain- ja asiasanoja, kenttä `AS`. CERIFin kenttä 70, *Free key words*, jossa jokainen avainsana varustetaan kielikoodilla.

[31] EY:n CERIFin kenttä 71, *controlled terms*, jossa jokainen asiasana varustetaan kielikoodilla. Vain "Common European Research Theaurus" -termejä. Muita sanastoja varten kehotetaan luomaan kenttä numero 73.

[32] OPM:n suosituksen kenttä "KW".

[33] "SPINES thesaurus : a controlled and structured vocabulary for information processing in the field of science and technology for development", Unesco, Paris, 1988, 1122 s., ISBN 92-3-102257-1.

[34] EY:n CERIFin kenttä 94, *Special equipment*.

4.6 Luokitukset

Tutkimuksen **luokitukset**[35] `[luokitus]` kuvaavat tutkimushankkeen sisältöä, joka voi siis joskus poiketa tutkimusta suorittavan organisaation tavanomaisesta profiilista (vrt. KOTA-koulutusalaluokitus yllä, jonka ei ole tarkoitus kuvailla tutkimusta, vaan laitos).

- KOTA-oppiaineluokitus `[KOTO, kotaoppi]`. Ks. aiemmin mainittu KOTA-käsikirja, sivut 66–74, joilla luetellaan oppiaineiden kolminumeroiset koodit. Pakollinen tieto.
- EY:n CERCS-luokitus `[EYT, EYluok]`. Ks. aiemmin mainittu EY:n CERIF-suosituksen Annex II, suosituksen sivut 23–34. Koodit koostuvat yhdestä kirjaimesta ja kolmesta numerosta, esim. "P176" *Artificial intelligence*. Pakollinen tieto.
- UDK-luokitus `[UDKT,UDK]`.
- Tilastokeskuksen toimialaluokitus `[toimalat, toimiala]`.
- Muut luokitukset `[muuluok]`.

Luokituksista vähintään yhden KOTA-oppiaineluokituksen ja EY:n CERCS-luokituksen antaminen on pakollista. UDK- ja toimialaluokituksia saa antaa lisäksi. Jos on tarpeen, voi samasta luokitustyypistä antaa useita koodeja.

On myös mahdollista antaa luokituksia, joita ei tässä listassa ole. Niistä laitetaan luokituksen tyyppi `[ltyyppi]` ja luokituksen koodi `[luokoodi]`.

```
<!ELEMENT luokitus - o (KOTO, EYT, UDKT?, toimalat?, muuluok*)>
<!ELEMENT KOTO     - o (kotaoppi)+>
<!ELEMENT EYT      - o (EYluok)+>
<!ELEMENT UDKT     - o (UDK)+>
<!ELEMENT toimalat - o (toimiala)+>
<!ELEMENT (kotaoppi, EYluok, UDK, toimiala)
                   - o (#PCDATA)>
<!ELEMENT muuluok  - o (ltyyppi , luokoodi)>
<!ELEMENT (ltyyppi, luokoodi)
                   - o (#PCDATA)>
```

[35] OPM:n suosituksen kenttä `LU`. EY:n CERIF-suosituksen kenttä 72, *Common subject classification codes — Common European Research Classification Scheme* on varattu CERCS-luokitukselle. Mahdollista muuta (paikallisesti käytettyä) luokitusta varten CERIFissä on kenttä numero 74.

4.7 Aikataulutiedot

Tutkimushankkeet ovat projekteja, ja projektit tunnetaan siitä, että niillä on tietty alku ja rajallinen kesto. Tällöin voidaan siis antaa aikataulutiedot `[aikat]`.[36]

- **Tutkimuksen alkamispäivämäärä** `[alkupvm]`. Tieto ei muutu, katsottu pakolliseksi tiedoksi.
- **Tutkimuksen loppumispäivämäärä** `[loppupvm]`. Päivämäärä, johon saakka tutkimushankkeen rahoitus on toistaiseksi varmistettu, tai johon saakka tutkimushankkeen on arvioitu kestävän. Muuttuminen hankkeen kuluessa mahdollista.
- **Päivämäärä jota annetut tiedot koskevat** `[tietopvm]`. Yleensä päättyneen vuoden viimeinen päivä. Tietoja voidaan päivittää useamminkin. Pakollinen tieto.[37]
- **Tutkimuksen vaihe** `[vaihe]`. Jos tutkimus on käynnissä, kuten varmaan useimmat tietokantaan tulevat tutkimukset ovat, jätetään tämä täyttämättä.[38] Mahdolliset vaiheet ovat: keskeytynyt, päättynyt tai päättynyt ja loppuraportoitu. Tämä tieto on pakollinen siten, että, jos sen jättää täyttämättä, oletetaan, että tutkimus on käynnissä.
  - Keskeytynyt: `&KE;`.
  - Päättynyt, mutta vielä loppuraporttia vaille valmis: `&PA;`.
  - Kokonaan valmis, eli loppuraportoitukin: `&RA;`.

```
<!entity KE "keskeytynyt">
<!entity PA "päättynyt">
<!entity RA "päättynyt ja loppuraportti valmis">
```

```
<!ELEMENT aikat    - o (alkupvm, loppupvm?, tietopvm, vaihe?)>
<!ELEMENT (alkupvm, loppupvm, tietopvm)
                   - o (pvm)>
<!ELEMENT vaihe    - o (#PCDATA )>
```

[36] OPM:n suosituksen kenttä `AT` sisältää alkamis- ja loppumisajankohdan. CERIFin kentät 40, *Starting date of the project* ja 41, *Expected ending date of the project*.

[37] Tieto voidaan täyttää korkeakoulun tietokannassa automaattisestikin, jos muulla tavoin tiedetään oikea päivämäärä.

[38] OPM:n suosituksen kenttä `VA` voi olla: suunnitteilla "SU", tekeillä "TE", keskeytynyt "KE", päättynyt "PÄ" tai loppuraportoitu "RA".

Minimaalinen päivämäärä [pvm] käsittää pelkän vuosiluvun [vv]. Jos kuukaudella [kk] ja päivällä [pp] on merkitystä annetaan päiväyksistä ensin päivä ja kuukausi, sitten vuosi.

```
<!ELEMENT pvm      o o (pp?, kk?, vv)  -(pvm | summa | org)>
<!ELEMENT (pp, kk, vv)
                   - o (#PCDATA )>
```

4.8 Rahat ja resurssit

Tutkimuksen rahoitus [rahoitus][39] raportoidaan kausittain, joka on pääsääntöisesti kalenterivuosi. Erityisistä syistä, jos esim. korkeakoulun oman hallinnon tai laskennan takia ei toistaiseksi saada vuosittaisia tietoja, voidaan antaa muu kausi tai käyttää tutkimushankkeen kokonaisrahoituksen tietoja (jättämällä **kausi** pois).

Rahoitustieto on kumuloituvaa tietoa. Uuden rahoituskauden tietojen tullessa vanhoja tietoja ei ole tarvis poistaa, vaan ne voidaan säilyttää tietueessa. Tutkimuksen kokonaisrahoitusta koskevat tiedot voidaan siten laskea summaamalla.

Rahoitukseen kuuluvat tutkimuksen kunakin rahoituskautena käyttämät resurssit. Parhaassa tapauksessa korkeakoulun kirjanpidosta tulevina lukuina, muuten myönnettyjen lukujen perusteella tai arviona.

Rahoituskausi koostuu seuraavista tiedoista:

- Kausi, se aika jota nämä rahoitustiedot koskevat [rahkausi]. Kausi voi olla vuosi tai koostua alkamisajankohdasta [alkupvm] ja loppumisajankohdasta [loppupvm]. Mikäli kauden pituus on sama kuin koko tutkimuksen kesto, ei tätä tietoa tarvita.

- Oman laitoksen tai vastaavan perusbudjetista tulevat rahat, eli ns. perusvoimavaroista tätä hanketta varten irrotetut varat [perus].

- Oman korkeakoulun tai vastaavan organisaation myöntämät, laitoksen perusbudjetin ulkopuolelta tulevat rahat - tarkoitukseen myönnetyt lisärahat, stipendit, laiterahat tms.[erityis].

- Kustakin vieraasta eli oman korkeakoulun ulkopuolisesta rahoituksesta rahan myöntäjä [antaja], käytetty summa [maara] ja annetun tuen laatu [yksikko]. Yksikkö ilmoitetaan yleensä markkoina, mutta henkilötyökuukausina, esim. milloin Suomen Akatemia tai ulkopuolinen firma on sijoittanut

[39] OPM:n kenttä **RA**, johon voidaan laittaa rahoituslähde ja -määrä vuosittain. EY:n CERIFin kentät 80 ja 81. Kenttä 80, *Financing source* sisältää rahoittavan organisaation nimen ja koodin. Kenttä 81, *Amount* tutkimuksen kokonaisrahoituksen yleisesti *...total amount of the financial support for the entire project...* (81.1) ja vuosittain *...total annual amount of the financial support...* (81.2). Lisäksi kenttään 81 kuuluvat osakentät 81.3 - 81.6, *Average amount per year*, *Annual personnel cost*, *Annual working cost* ja *Total Equipment cost*.

tutkimushenkilöstöä hankkeeseen siten, että palkkarahat eivät kulje vastuullisen johtajan tai korkeakoulun kautta.

- Tutkimukseen kuluneet henkilötyökuukaudet [40] `[henktkk]`. Pakollinen tieto.

```
<!ELEMENT rahoitus - o (rahkausi)+>
<!ELEMENT rahkausi - o (kausi?, menot?, henktkk, rahkerto?)>
<!ELEMENT kausi    - o (vuosi | (alkupvm, loppupvm))>
<!ELEMENT menot    - o (omat?,vieraat?)>
<!ELEMENT omat     - o (perus?, erityis?)>
<!ELEMENT vieraat  - o (antaja, maara, yksikko)+>
<!ELEMENT (antaja, perus, erityis, henktkk, vuosi, maara, yksikko)
                   - o (#PCDATA)>
```

Lisäksi on varattu tilaa vapaalle rahoituskertomukselle `[rahkerto]`. Tässä on mahdollisuus selittää mistä rahaa on saatu ja mihin sitä on käytetty.

Rahoituskertomus on vapaata tekstiä, johon sisältyviä merkityksellisiä kohtia voidaan erityisesti merkitä näkyville. Tällaisia ovat organisaatioiden nimet ja yhteystiedot `[org]`, päivämäärät `[pvm]` ja rahasummat `[summa]`. Tällainen merkitseminen on vapaaehtoista, mutta saattaisi parantaa haun mahdollisuuksia.

```
<!ELEMENT rahkerto - o (#PCDATA) +(pvm | summa | org)>
<!ELEMENT summa    - - (#PCDATA) -(pvm | summa | org)>
<!ELEMENT org      - o (orgnimi, postios?, sijainti?,
                        puh?, fax?, email?, muuta?)>
```

4.9 Tutkijat

Tutkijoiden yhteydessä ei ilmoiteta apuhenkilökuntaa. Kustakin tutkimuksessa työskentelevästä tutkijasta annetaan seuraavat tiedot `[tutkijat]`:[41]

- Sukunimi. Pakollinen tieto.
- Etunimet. Pakollinen tieto.
- Oppiarvo `[oppiarvo]`.

[40] OPM:n kenttä `HT`, henkilötyön määrä. CERIFin kenttä 82, *Number of researchers* sisältää tiedot täysitoimisten tutkijoiden määrästä yhteensä ja työskentelyjaksosta kuukausina, joten projektin vuosittaiset henkilötyökuukaudet saadaan kertolaskulla.

[41] OPM:n kenttä `MU`, muiden kuin vastaavan tutkijan nimet, CERIFin kenttä 32, *Principal researchers*. Molemmissa halutaan vain nimitiedot.

- Oppiarvon (tai korkeimman tutkinnon) suoritusvuosi `[vuosi]`. Täytetään sikäli, kun rahoittaja tietoa vaatii (esim. Akatemia).
- Sukupuoli `[sukup]`. Arvona SGML-lyhenne `&NAINEN;` nainen tai `&MIES;` mies. Täytetään, jos rahoittaja tietoa vaatii (esim. Akatemia).
- Kansalaisuus, ellei Suomen `[kansal]`. Maan nimi, täytetään, jos rahoittaja tietoa vaatii (esim. Akatemia).
- Tutkimuksessa mukanaolon aloituspäivämäärä `[alkupvm]`. Pakollinen tieto.
- Tutkimuksessa mukanaolon lopetuspäivämäärä `[loppupvm]`.
- Asema tutkimuksessa `[asema]`. Jos hankkeella on useampia rahoittajia, ja ainakin yksi näistä edellyttää erittelyä, merkitään kunkin tutkijan vakanssin perään rahoittajan lyhenne suluissa. (Esim. Akatemia edellyttänee tätä).
- Ulkomailla suoritetun tutkimuksen maa `[maa]`, jos rahoittaja tietoa edellyttää (esim. Akatemia).

Tutkimuksen vastuuhenkilö mainitaan tässä uudestaan, jos hän on tutkijana mukana tutkimuksessa. Lisäksi annetaan tutkimuksessa mukanaolon aloituspäivämäärä ja lopetuspäivämäärä. Jos lopetuspäivämäärää ei ole annettu, merkitsee se sitä että tutkija on aktiivisesti mukana tutkimuksessa. Tämän lisäksi saa vielä antaa oppiarvon ja aseman tutkimuksessa. Oppiarvo ja asema eivät ole pakollisia tietoja.

Jos sama tutkija työskentelee useita eri jaksoja tutkimuksen parissa, annetaan kustakin työskentelyjaksosta erikseen alkupäivämäärä ja loppupäivämäärä sekä tutkijan asema tutkimuksessa mikäli se on muuttunut.

```
<!entity NAINEN "nainen">
<!entity MIES   "mies">

<!ELEMENT tutkijat - o (tutkija)+>
<!ELEMENT tutkija  o o (hlo, oppiarvo?, vuosi?,
                        sukup?, kansal?, tyojakso+)>
<!ELEMENT tyojakso - o (alkupvm, loppupvm?, asema?, maa?)>
<!ELEMENT (oppiarvo, sukup, kansal, maa)
                   o o (#PCDATA)>
```

4.10 Julkaisut

Julkaisuina [julk][42] ilmoitetaan valmiit, painetut (tai muutoin julkisuuteen saatetut) kirjat, artikkelit ja vastaavat. Painossa olevia tai suunnitelman asteella olevia hengentuotteita ei saa luetella tässä. Niistä voi mainita tiivistelmässä, jos sen katsoo aiheelliseksi. Tutkimushankkeen tuloksena syntyneet julkaisut on jaettu kuuteen eri typpiin, joista annetaan asianmukaiset tiedot.

```
<!ELEMENT julk     - o (kirja | kirjart | lehtiart |
                         patentti | atktall | muujulk)+>
```

4.10.1 Kirja

Kirja [kirja] eli ns. monografia kuvataan seuraavilla tiedoilla:

- Kirjoittajat tai toimittajat [kirj, toim]. Kenttä on rakenteeltaan sama kuin [hlo], eli sisältää osakenttinään sukunimen ja etunimet. Pakollinen tieto.
- Kirjan nimi [otsikko]. Pakollinen tieto.
- Painos [painos]. Muuttamattomia uusintapainoksia ei saa ilmoittaa uusina julkaisuina.
- Kustannuspaikka [kustpaik]. Jos pitkä luettelo eri kaupunkeja, kuten tunnetuilla kustantajilla, jätetään täyttämättä.
- Kustantaja [kust]. Pakollinen tieto.
- Julkaisuvuosi [vuosi]. Pakollinen tieto.
- Kirjan sivumäärä [sivuja].
- Sarja, johon kirja kuuluu [sarja]. Teoksen numero sarjassa sisällytetään tähän kenttään. Tieto annetaan, jos katsotaan hyödylliseksi.
- Volyymin numero [vol], jos on kyse useampiosaisesta kirjasta.
- ISBN väliviivoineen [ISBN].

[42] OPM:n kenttä JU: julkaisut, julkaisusuunnitelmat ja julkaistut raportit tai artikkelit yms. CERIFin kenttä 93, *Interim results* sisältää julkaisujen lukumäärän ja lueteltuina patentit ja rakennetut prototyypit.

- Julkaisun luokitus ja mahdollinen ulkomaisuus [julkluok]. Jos kyseessä on laitossarjassa julkaistu teos, laitetaan kenttään SGML-olio &LTSJULK;. Jos kyseessä on Korkeakoulun sarjassa julkaistu teos, laitetaan kenttään SGML-olio &KORKJULK;. Jos kyseessä on muualla julkaistu teos, laitetaan kenttään SGML-olio &MONOGR;. Jos kyseessä on asianomaisen toimittama teos, laitetaan kenttään SGML-olio &TOIMJULK;. Jos kyseessä on oppikirja, laitetaan kenttään SGML-olio &OPPIK;.

  Jos julkaisu on ulkomainen, lisätään SGML-olio &ULKOM;.

- Mahdolliset huomautukset [huom].

```
<!entity MONOGR "monografia">
<!entity KORKJULK "korkeakoulun sarjassa">
<!entity LTSJULK "laitoksen sarjassa">
<!entity TOIMIT "toimitettu teos">
<!entity OPPIK "oppikirja">
<!entity ULKOM "ulkomailla julkaistu">
```

```
<!ELEMENT kirja    - o ((kirj | toim), otsikko, painos?,
                        kustpaik?, kust, vuosi, sivuja?,
                        sarja?, vol?, ISBN?, julkluok?, huom?)>
<!ELEMENT (kirj, toim)
                   - o (hlo)+>
<!ELEMENT (otsikko, painos, kustpaik, kust, sivuja,
          sarja, vol, ISBN, julkluok)
                   - o (#PCDATA)>
```

4.10.2 Kokoomateoksen osa

Artikkeli kirjassa (esim. kongressijulkaisussa) eli kokoomateoksen luku tai osa [kirjart]. Pelkkiä tiivistelmiä (abstrakteja) ei saa sisällyttää näihin, vaan ne voidaan ilmoittaa muina julkaisuina. Kokoomateoksen osa kuvataan seuraavasti:

- Kirjoittajat [kirj]. Pakollinen tieto.
- Artikkelin nimi [otsikko]. Pakollinen tieto.
- Kirjan toimittajat [toim].
- Kirjan nimi [kirnimi]. Pakollinen tieto.

- Painos [painos].
- Kustannuspaikka [kustpaik]
- Kustantaja [kust] Pakollinen tieto.
- Vuosi [vuosi]. Pakollinen tieto.
- Sarja, johon kirja kuuluu [sarja]. Teoksen numero sarjassa sisällytetään tähän kenttään. Tieto annetaan, jos katsotaan hyödylliseksi.
- Volyymi tai numero sarjassa [vol].
- ISBN [ISBN].
- Sivut joilla artikkeli on [sivut]. Pakollinen tieto.
- Julkaisun luokitus ja mahdollinen ulkomaisuus [julkluok]. Koska kyseessä on kirjan tai kongressijulkaisun osana julkaistu artikkeli, laitetaan kenttään SGML-olio &KIRART;.

  Jos julkaisu on ulkomainen, lisätään SGML-olio &ULKOM;
- Mahdolliset huomautukset [huom].

```
<!entity KIRART "artikkeli kirjassa">

<!ELEMENT kirjart  - o (ulkom?, kirj, otsikko, toim?, kirnimi,
                        painos?, kustpaik?, kust, vuosi,
                        sarja?, vol?, ISBN?, sivut,
                        julkluok?, huom?)>
<!ELEMENT (kirnimi, sivut)
                   - o (#PCDATA)>
```

4.10.3 Artikkeli

Artikkeli tieteellisessä aikakausjulkaisussa [lehtiart]. Tässäkään ei saa ilmoittaa pelkkää abstraktia. Artikkeli kuvataan seuraavasti:

- Kirjoittajat [kirj]. Pakollinen tieto.
- Otsikko [otsikko]. Pakollinen tieto.
- Lehti [lehti]. Pakollinen tieto.
- ISSN [ISSN].

- Vuosi [vuosi]. Pakollinen tieto.
- Volyymi tai vuosikerta [vol]. Pakollinen tieto.
- Numero [nro]. Silloin, kun aikakausjulkaisun sivunumerointi on volyymikohtainen, voidaan numerotieto jättää pois. Näin ei tietenkään voi tehdä, jos volyymin kussakin numerossa sivunumerot alkavat 1:stä. Terve harkinta on paikallaan tässäkin.
- Sijainti lehdessä [sivut]. Pakollinen tieto.
- Julkaisun luokitus ja mahdollinen ulkomaisuus [julkluok]. Jos kyseessä referoitu lehti laitetaan kenttään &REFART;, jos muussa, siis referoimattomassa lehdessä, laitetaan SGML-olio &MUUART;.

  Jos julkaisu on ulkomainen, lisätään SGML-olio &ULKOM;
- Mahdolliset huomautukset [huom].

```
<!entity REFART "artikkeli referoidussa lehdessä">
<!entity LTSART "artikkeli laitossarjassa">
<!entity MUUART "artikkeli muussa, referoimattomassa lehdessä">

<!ELEMENT lehtiart - o (kirj, otsikko, lehti,
                        ISSN? ,vuosi,vol?, no?, sivut,
                        julkluok?, huom?)>
<!ELEMENT (lehti, nro, ISSN)
                   - o (#PCDATA)>
```

4.10.4 Patentti

Patentille olennaisia tietoja ovat:

- Keksijät [keksijat]. Pakollinen tieto.
- Patentin nimi [patnimi]. Pakollinen tieto.
- Patentin haltija [haltija].
- Patentin julkaisijamaa [patmaa]. Pakollinen tieto.
- Myöntämisvuosi [vuosi]. Pakollinen tieto.
- Patentin julkaisunumero [nro]. Pakollinen tieto.
- Patentin sisällön kuvaus lyhyesti [patsis].
- Mahdolliset huomautukset [huom].

```
<!ELEMENT patentti - o (keksijat, patnimi, haltija?,patmaa,
                        vuosi, nro, patsis?, huom?)>
<!ELEMENT haltija  - o (hlo | #PCDATA)>
<!ELEMENT (patsis, patmaa, patnimi)
                   - o (#PCDATA)>
<!ELEMENT keksijat - o (hlo)+>
```

4.10.5 Ohjelmat, tallenteet ym.

Tietokonetallenne `[atktall]`, useimmiten julkaistu (siis myytävänä oleva tai muuten yleisön saataville saatettu) tietokoneohjelma, tietokanta tai muu aineisto:

- Tekijät `[tekijat]`. Pakollinen tieto.
- Nimi `[otsikko]`. Pakollinen tieto.
- Julkaisuvuosi `[vuosi]`. Pakollinen tieto.
- Mahdolliset huomautukset `[huom]`.

```
<!ELEMENT atktall  - o (tekijat, otsikko, vuosi, huom?)>
<!ELEMENT tekijat  - o (hlo)+>
```

4.10.6 Muut julkaisut

Luokkaan muut julkaisut `[muujulk]` voidaan laittaa sellainen materiaali joka ei sovi yllä oleviin kaavoihin, esim. videonauhoitteet. Muun julkaisun rakenne on määrittelemätön, se on vain juoksevaa tekstiä.

```
<!ELEMENT muujulk  - o (#PCDATA)>
```

4.11 Opinnäytteet

Tutkimuksen yhteydessä suoritetut opinnäytteet `[opinnayt]`, joista kerrotaan:[43]

- Tekijä `[kirj]`. Opinnäytteen tekijän sukunimi ja etunimet kuten muissakin `[hlo]`-tiedoissa. Pakollinen tieto.
- Laitos, jonka piirissä olevassa oppiaineessa opinnäyte on hyväksytty `[laitos]`. Pakollinen tieto.
- Opinnäytteen laatu - pro gradu, väitöskirja tms.`[opinn]`. Pakollinen tieto.
- Opinnäytteen nimi `[opinnimi]`. Pakollinen tieto.
- Hyväksymisvuosi `[vuosi]`. Pakollinen tieto.
- Muita tietoja `[muuta]`. Väitöskirjojen osalta rahoittajataho (esim. Akatemia) saattaa vaatia tekijän syntymävuoden suluissa, esim. `(1956)`. Edelleen, tällaisen opinnäytteen pääasiallinen rahoittaja, jos se ei ole tutkimushankkeen päärahoittaja saatetaan toivoa merkittäväksi tähän kohtaan, esim. “TEKES 70 %”.

```
<!ELEMENT opinnayt - o (nayte)+>
<!ELEMENT nayte    - o (kirj, laitos, opinn, opinnimi, vuosi, muuta?)>
<!ELEMENT (opinn, opinnimi)
                   - o (#PCDATA)>
```

4.12 Yhteydet

Viitteet muihin tutkimuksiin `[yhtstyo]`.[44] Jos tutkimuksella on emoprojekteja `[emoproj]`, tytärprojekteja `[tytarpr]`, ja/tai rinnakkaisprojekteja `[sisarprj]`, annetaan niiden tiedot kuten oman tutkimuksen tutkimussopimuksen `[tutksop]` tiedot, tutkimussopimuksen osapuoli ja sopimusnumero `[sopkoodi]`. Jos jompikumpi näistä on tuntematon, voi antaa vain toisen tiedon.

Muista yhteyksistä `[yhtkerto]` voi kertoa vapaasti tekstimuodossa. Tänne voi erikseen merkitä sopimuskoodeja `[sopkoodi]` ja organisaatioita `[org]` ja niiden yhteystietoja.

---

[43] OPM:n kenttä `ON`, ilmoitetaan diplomityöt, lisensiaattityöt, väitöskirjat, pro gradut ym. joista kerrotaan tekijä, oppilaitos, osasto, opinnäyte, opinnäytteen nimi ja päivämäärä.

[44] OPM:n kentät `YH`, tutkimukseen osallistuvat muut laitokset ja mahdollisesti myös niiden yhteyshenkilöt, ja `OT`, osa- ja jatkotutkimukset. EY:n CERIF-suosituksen kenttä 92, *Related to cooperation* jossa luetellaan tutkimusprojektin kannalta tärkeät yhteydet muihin tutkijoihin tai tutkimuslaitoksiin, kuuluu kategoriaan *Additional information*.

```
<!ELEMENT yhtstyo  - o (emoproj*, tytarpr*, sisarpr*,
                        yhttahot?, yhtkerto?)>
<!ELEMENT yhttahot - o (org, yhttapa?)+>
<!ELEMENT (yhttapa, yhtkerto)
                   - o (#PCDATA)+>
<!ELEMENT (emoproj, tytarpr, sisarpr)
                   - o (tutksop)>
```

4.13 Mahdolliset muut tiedot

Jos on jotakin sellaista relevanttia tietoa `[muuta]`,[45] joka ei sopinut aikaisempiin kohtiin mutta olisi kuitenkin tutkimuksen kannalta olennaista, voi sen lisätä tähän kohtaan.[46]

Tietokannassa on kolmentyyppisiä huomautuksia `[muuta, huom]` ja `[komm]`. Näistä muuta-elementit ovat niitä tietoja varten jotka katsotaan jossakin aiheellisiksi, mutta joille ei ole omaa elementtiä tässä formaatissa, huom-elementit on tarkoitettu lähinnä jossakin muussa tietoelementissä olevan tiedon kommentoimiseen, ja komm-elementit on varattu jonkinlaisiksi toim.huomeiksi, joita voi laittaa minne tahansa tietueen sisällä.

```
<!ELEMENT (muuta | huom) - o (#PCDATA)>
<!ELEMENT komm     - - (#PCDATA)>
```

[45]CERIFin kenttä 95, *Other relevant information*.

[46]OPM:n suosituksessa on kaksi kenttää, joita ei tässä tietosisällössä ole: `KK`, tutkimuksen kokonaiskustannukset (jotka kuitenkin ovat tämän tietosisällön perusteella laskettavissa) ja `XX`, tutkimuksen sisällön kuvailu muilla halutuilla kielillä. EY:n CERIFissä esiintyy kenttä 91, *Related to research programme*, joka sisältää tutkimusohjelman nimen ja akronyymin jos tutkimus on osana jotain EY:n tutkimusohjelmaa.

Luku 5

Ehdotukset jatkotoimenpiteiksi

Seuraavassa esitetään erinäisiä mahdollisia toimenpiteitä, jotka edesauttaisivat seurantaryhmän työn kohteena olleiden tutkimustietokantojen toteutumista ja sitä, että tietokannat olisivat mahdollisimman yhteismitallisia ja hyödyllisiä eri tarvitsijoille.

5.1 Olosuhteet ja edellytykset

Olosuhteet ovat melko otolliset korkeakoulujen tutkimushankkeita kuvaavien tietokantojen käynnistymisen kannalta. Toisaalta voidaan tunnistaa useita jarruttavia tekijöitä:

- Korkeakoulut ja niiden tieteellinen hallinto, erityisesti tiedekunta- ja laitostasolla ei kaikkialla ole vielä mieltänyt tutkimusrekistereitä omakseen, vaan pitää sellaisia vain valtakunnallisen keskitetyn hallinnon asiana. Nykytilanteessa korkeakoulujen olosuhteet ja pelisäännöt ovat nopeasti muuttuneet. Kaikkialla ei ole ehditty kokea tarvetta laadullisesti riittävän tiedon saantiin, vaan ollaan tyydytty kappale- ja summatietoihin.

- SGML on uusi ja verraten vähän tunnettu paitsi hallintoportaan, myös ATK-keskusten piirissä.

- Korkeakoulujen yhteistyölle tämäntapaisissa hankkeissa ei ole vakiintuneita muotoja, vaikka tietoverkkojen (FUNETin ym.) ansiosta tekniset mahdollisuudet ovat erinomaiset.

Tietokantojen perustaminen ei välttämättä vaadi kalliita laitteita, eikä paljon maksullisia ohjelmiakaan. SGML-editori saattaisi olla paras työväline hankkeiden tietojen syöttöä varten, ja tällaiset editorit ovat maksullisia ohjelmatuotteita.

5.2 Toimenpiteet

Seurantaryhmä esittää, että *opetusministeriö kehottaisi korkeakouluja toteuttamaan tutkimushankkeita koskevat tietokannat* tässä raportissa lueteltujen periaatteiden mukaisesti sekä *koordinoisi ja tukisi tietokantojen perustamista*. Ryhmä pitää tärkeänä, että perustettavien tietokantojen tiedot ovat julkisia ja tulevat yleisesti saataville.

Seurantaryhmä katsoo, että tutkimushankkeiden osalta korkeakoulujen ja laitosten tuloksellisuuden arvioinnissa tulisi käyttää *ensisijaisesti laadullisia kriteerejä*, joita käsillä olevassa raportissa ja sen suosituksissa on korostettu. Periaatteen tärkeyden vuoksi ja tutkimustietokantojen perustamisen nopeuttamiseksi olisi toivottavaa, että mm. *rehtorien neuvostoa informoitaisiin* seurantaryhmän tuloksista. Ryhmä näkee tietojen keruun ensisijaisesti korkeakoulujen tieteelliseen tuloksellisuuteen perustuvan johtamisen ja päätöksenteon edellytyksenä pikemmin kuin tavanomaisena hallinnollisena ATK-hankkeena.

Korkeakoulujen ja opetusministeriön tulisi *järjestää yhteinen seminaari*, jolla helpotettaisiin tietokantojen perustamista eri korkeakouluissa ja luotaisiin alustava yhteyksien verkosto.

Seurantaryhmä esittää, että opetusministeriö *järjestäisi korkeakoulukohtaisten tutkimusrekisterien koordinoinnin*. Sen lisäksi, että ministeriö vastaa koordinoinnista, se voisi sopia sopivan tahon, esimerkiksi jonkin korkeakoulun kanssa yhteyksien ylläpitämisestä ja eri korkeakoulujen kokemuksien, ehdotuksien sekä tulkintojen keräämisestä. Koordinoinnin tulisi edistää keskustelua korkeakoulujen kesken muutos- ja parannusehdotuksista.

Kun pääosa korkeakoulukohtaisista tutkimustietokannoista on toiminnassa, tulisi opetusministeriön ryhtyä toimenpiteisiin *valtakunnallisen tutkimustietokannan ATK-tekniseksi toteuttamiseksi*, koska huomattava osa korkeakoulujen itse tarvitsemasta tiedosta koskee vertailua toisten korkeakoulujen vastaavien oppiaineiden tuloksiin. Muutkin tutkimusrekisterien hyödyt saavutettaneen parhaiten rekisterillä, johon on koottu eri korkeakoulujen tutkimushankkeiden tiedot.