

# Ohjelma-arviointi

## metodologisia ja arviointitiedon tiedonmuodostukseen liittyviä näkökohtia

*Petri Virtanen*

### ABSTRACT

Policy-orientation will be increased in the Finnish public sector in the near future. For instance, the on-going Central Government Reform Programme has introduced several proposals regarding this. In this article, the motives of policy evaluation, the construction of evaluation body of knowledge, the implementation of policy evaluation as a research process and the dissemination of evaluation results are scrutinized. It is argued, that the causal cycle from empirical evidence to conclusions and then to recommendations is not always made visible completely. Instead, there seems to be a kind of "black box" between these entities. Therefore, it is the key challenge to evaluators to map out the linkages between these elements of programme evaluation. The article suggests that the key elements in the programme evaluation are a) focus on utilization aspects, b) three-dimensional space with regard to time, c) linking evaluation into programme monitoring, d) the validity and reliability of the methods and data used, and finally e) situational interpretation of the findings.

**Key words:** Programme evaluation, methodology, evaluation motives, New Public Management, reporting, dissemination of evaluation results

### 1. JOHDANTO: OHJELMA-ARVIONNIN KÄSITTEESTÄ JA ARTIKKELIN LÄHESTYMISTAVASTA

Arviointi, arviointitutkimus, seuranta, politiikka, ohjelma ja projekti ovat hankalia käsitteitä. Ne ovat hankalia sen vuoksi, että ne ovat indeksikaalisia. Ne saavat eri merkityksiä erilaisissa konteksteissa. Mainitut käsitteet liittyvät joka tapauksessa, kulloisestakin määritelmästä huolimatta, toisiinsa. Esimerkiksi Jeffrey L. Pressman ja Aaron Wildavsky sanovat kuuluisissa teoksessaan *Implementation* (1984), että implementaatio (politiikan, ohjelman tai projektin) ja arviointi ovat "saman kolikon vastakkaiset puolet".

Pressmanin ja Wildavskyn (mt., xvii) mukaan implementaatio, arviointi, policy-analyysi ja suunnittelu ovat poliittis-hallinnollisessa kontekstissa kysymyksiä, joita ei voi tarkastella analyttisesti toisistaan irrallisesti. Heidän mielestään arviointi tuottaa poliitikkojen, ohjelmien ja projektien toteutukseen oppimista: "arvioimme oppiaksemme, opimme implementoidaksemme". Ohjelmien ja projektien toteuttamiseen tavallisesti liitettävät arviointikysymykset ovat milloin, missä, kenelle, mitä ja miksi (Browne & Wildavsky 1984, 206-207), jotka mukailten sopivat perustaksi arviointikysymyksiksi kaikkiin arviointitutkimuksen ja -toiminnan paradigmatteihin.

Evert Vedungin teosta *Public Policy and Program Evaluation* (1997) pidetään Frank Fischerin *Evaluating Public Policy* (1995) teoksen ohella ohjelma-arviointin klassikkona. Molempien teosten määrittelyissä ohjelma-arviointilla tarkoitetaan ennen muuta julkisin varoin toteutettujen julkisten interventioiden arvon todentamista. Tämä väljä määritelmä on tämänkin artikkelin ohjelma-arviointia koskevan käsitteellisen ymmärryksen perusta. Käytännössä tämä määrittely mahdollistaa ensinnäkin sen, että paikoitellen kun puhun ohjelma-arviointista voi tähän rajauk-

seen sisällyttää myös projektiarvioinnin. Toisaalta tietyt havaintoni voi kytkeä myös laajempaan, koko arviointitutkimusta koskevaan viitekehykseen.

Vedung osoittaa edellä mainitussa kirjassaan havainnollisesti sen, miten moniulotteista ohjelma-arviointi on ja miksi ohjelma-arvioinnin alaa on vaikeata määritellä. Kyseessä on kuitenkin perusolemukseltaan hallinnon systeemisen mekanismin tarkastelu, jossa systeemin perusosia ovat panokset, konversio (toimintatapa) ja tuotokset. Vedung kirjoittaa vakuuttavasti myös ohjelma-ajatteluun liittyvistä arviointitoiminnan vaikeuksista, kuten esimerkiksi siitä, miten nimenomaan jotkut tietyt vaikutukset ovat seurausta jostain tietyistä interventioista (siis konversio Vedungin termiä käytettäessä).

Frank Fischer puolestaan tekee hienosyisempiä määrittelyjä ohjelma-arvioinnin lähestymistavasta ja merkityksestä. Hänelle empiirinen ohjelma-arviointi liittyy ilmiöön, jota hän kutsuu ohjelman verifiointiksi. Fischerille ohjelmat identifioivat ja lausuvat julki yhteiskunnallisten politiikkojen (policies, ei siis politics!) päämääriä ja tavoitteita, (mt., 28). Tullakseen hallinoiduksi yhteiskunnallinen politiikka tulee mieltää, jäsentää ja toteuttaa (implementoida) ohjelmana.

Ohjelma-arvioinnissa näin ollen ollaan fischeriläisittäin ajatellen kiinnostuneita yksinomaan tavoitteisiin sidotusta (goal-bound) arviointinäkökulmasta. Kyse on siitä, miten ohjelmalle ja projektille etukäteen tavoitteeksi asetetut asiaintilat toteutuvat tai jäävät toteutumatta. Tästä ei pitkä matka siihen, että ohjelma-ajattelu kytketään kokeellista tai ainakin kvasikokeellista asetelmaa ihannoivaan arviointikäytäntöön. Fischer sanoo (mt., 40-42), että tällaiseen ajatteluun liittyy lukuisia ongelmia, joista vähäisin ei ole objektiivisuuden harha (eli ollaanko niin ulkopuolisia ettei enää kerta kaikkiaan ymmärretä arvioitavan kohteen, ohjelman tai projektin, toimintalogiikka?).

Ohjelma-arviointi on siis varsin moniulotteinen kokonaisuus. Tarkastelen tässä artikkelissa arviointitutkimuksen tekemistä edellä lausutun määritelmän mukaisesti ohjelma-arvioinnin näkökulmasta. Olen kytkenyt tarkasteluuni mukaan erilaisia teemoja, kuten arviointitoiminnan motiivit, arviointitiedon rakentuminen, arviointiprosessin toteuttaminen käytännössä ja arviointitiedon käyttö ja olen täysin tietoinen, että aiheen käsitteilyni on kaukana tyhjentyvästä. On varsin helppo lisätä listaani kaikenlaista, mikä arviointitutkimuk-

seen jollain tavoin kuuluu.

Uskon, että ohjelma-arvioinnin metodologiset ja tiedonmuodostukseen liittyvät näkökohdat aukeavat parhaiten tarkasteltaessa arviointitietoon liittyviä motiiveja, erilaisten arviointiasetelmien koeltavuutta (ymmärtämällä erot havaintojen, johtopäätösten ja toimenpide-ehdotusten välillä) ja konkreettisten arviointien tekemiseen liittyvien kysymysten (miten esimerkiksi edistetään arviointitiedon hyödynnettävyyttä) kautta. Artikkelissa haetaan siis vastausta kysymykseen: Mitä ohjelma-arviointi on, millaisen prosessin kautta arviointien toimenpide-ehdotukset syntyvät ja millaisena käytännön prosessina arviointihanke toteutetaan.

En ole kuormittanut tekstiä runsaalla ajan-kohtaisella keskustelulla tai edes konkreettisilla ohjelma-arvioinnin suomalaisilla esimerkeillä (edes niillä, joita olen itse tehnyt) enkä tällä ratkaisulla millään tavalla väheksy muiden tekemisiä. Pikemminkin päinvastoin. Jäsenenän tässä artikkelissa esitettävät ajatukset osaksi kolmen artikkelin sarjaa, jota olen kirjoittanut vuodesta 1998 ja jossa käsillä oleva kirjoitus on viimeinen osa. Vuonna 1998 julkaistussa artikkelissa erittelin arviointitiedon hyödyntämisen mahdollisuuksia (Virtanen 1998) ja vuonna 2000 ilmestyneessä tekstissä analysoin arviointiosaamisen tilaa suomalaisessa julkishallinnossa ja kolmannen sektorin kentässä (Virtanen 2001).

Kirjoitus- ja asioiden käsittelytapaani voi syyllä luonnehtia korostuneen subjektiiviseksi, enkä tätä yritä millään tavoin peittää tai häivyttää. Metodisessa mielessä lähestymistavalleni on löydettävissä ajatuksellinen lähtökohta hijattain poismenneen sosiologin Pierre Bourdieun oivaluksista. Bourdieu puhui keskeisissä teoksissaan refleksiivisestä tutkimusotteesta yhteiskuntatieteilijän perusominaisuutena. Bourdieun refleksiivisyys viittaa nimenomaisesti siihen, että kirjoittajan on paljastettava ajattelunsa perusta, jotta lukijat voisivat perusteellisesti punnita kirjoittajan motiiveja ja intentioita. Tässä siis lukuohje olisi jotakuinkin sellainen, että tämän kirjoituksen luettuaan pystyy paremmin arvioimaan allekirjoittaneen empiirisiä ohjelma-arviointeja.<sup>1</sup>

<sup>1</sup> Olen kiitollinen Jari Stenvallille oivaltavista ja rakentavista kommentista artikkelin viimeistelyvaiheessa.

## 2. OHJELMA-ARVIOINNIN AJANKOHTAISUUDESTA

Evert Vedung (2001) on hiljattain sanonut, että arviointitoiminta on yksi merkittävimmistä julkishallinnon toimintaa rakenteistavista megatrendeistä. Hänen mukaansa elämme parhaillaan ilmiötä, jota hän kutsuu lennokkaasti termillä "audit explosion". Voitaisiinko kohta puhua ilmiöstä, jonka voisi Vedungiin viitata nimitä "programme evaluation explosion"?

Vedungin ilmauksen uudelleen muotoilun jälkeen on viimeistään paikallaan esittää kysymys, miksi ohjelma-ajattelu ja ohjelma-arviointi sitten ovat ajankohtaisia kysymyksiä Suomessa 2000-luvun alussa?

Viime vuosien aikana tapahtunut paljon selaista, joka antaisi olettaa ohjelma-ajattelun vahvistuvan poliittis-hallinnollisessa suunnittelussa ja päätöksenteossa. Esimerkiksi syksystä 2000 käynnissä olleen keskushallinnon uudistamishankkeen osana syntyneessä kohtalaisen tuoreessa raportissa (Kekkonen 2001) puhutaan voimaperäisesti ohjelmajohtamisen puolesta. Kyseisessä raportissa esitetään ajatus uudesta valtioneuvoston työtavasta, joka käytännössä tarkoittaisi valtioneuvoston kanslian vahvistamista, ohjelmien rakentamisesta ministeriöiden "putkihallinnollisten" koneistojen välimaastoon ja uudentyypistä koordinaatiota (ohjelmista vastaavat ministerit, ohjelmajohtajat ja ohjelmien perustuminen olemassa olevaan budjetin kehysmenettelyyn). Kyse on isosta ajatuksellisesta ja periaatteellisesta muutoksesta, jos edes osa Sirpa Kekkonen esittämistä ajatuksista toteutuu.

Käynnissä olevan keskushallintohankkeen – johon sivumennen sanoen liittyy kovasti odotuksia myös julkisuudessa, esimerkiksi lehtitietojen perusteella on luvassa "hallintovallankumous" (Helsingin Sanomat 29.12.2001) – ohella toinen merkittävä ohjelma-ajattelua suomalaisessa hallinnossa edistänyt seikka liittyy Suomen jäsenyyteen Euroopan unionissa.

Euroopan unionin varoin toteutetut rakennerahastot ovat nimensä mukaisesti ohjelmia ja nimenomaan taloudellisessakin mielessä isoja yhteiskuntapoliittisia ohjelmia. Rakennerahasto-ohjelmien toimeenpano on edellyttänyt suomalaiselta hallintokulttuurista jos ei täydellistä uudelleen suuntautumista niin ainakin merkittävää uutta hallintoajattelua. Toki on sanottava,

että ohjelma-ajattelua on suomalaisessa julkisessa hallinnossa Euroopan unionin ohella edistänyt myös muiden OECD-maiden hyvät esimerkit hallinnollisen järjestelmän kääntämisestä perinteisestä byrokraattisesta "laatikkoajattelusta" uudentyypiseen ohjelmaperusteiseen hallintoon (esim. Bouckaert et al. 2000; Pollitt & Bouckaert 2000). Rakennerahastoihin liittyvä arviointivelvoite on merkinnyt mittavaa ohjelma-arviointien lisääntymistä suomalaisessa hallinnossa ja päätöksenteossa.

## 3. ARVIOINNIN MOTIIVIT

Ohjelma-arvioinnin olemusta on syytä lähteä pohtimaan erittelemällä siihen liittyviä motiiveja. Tässäkään yhteydessä ei ole mitään syytä jättää mainitsematta Michael Q. Pattonin (1997) usein siteerattua luokitusta arviointitiedon käyttötaivoista.

Pattonin mukaan arviointitoiminnan motiivit liittyvät yhtäältä tiedolliseen näkökohtaan (joskus näkee puhuttavan akateemisesta intressistä), toisaalta kehittämistoimintaan ja yleensä oppimiseen ja kolmanneksi tilintekovastuuseen, joista viimeksi mainittu motiivi on hyvin lähellä myös valvonnan ja seurannan lähtökohtia.

Tiedolliset näkökohdat ovat sinällään olenainen ulottuvuus arviointitoiminnan motiivien kokonaisuudessa. Yksi tärkeä syy tähän liittyy ohjelma-arvioinnin osalta metodikehitykseen, joka Suomessa on – ottaen huomioon ohjelma-arvioinnin ajankohtaisuus (vrt. pohdinta edellisessä jaksossa) – on vielä tätä artikkelia kirjoitettaessa vähäisesti pohdittu kysymys.

Näen tiedolliset motiivit juuri tätä taustaa vasten, en niinkään substantiaalisten akateemisten opinnäytteiden tuottamisena. Tiedolliset näkökohdat varmistava menetelmäkehitys varmistaisi Suomessa toteutettujen ohjelma-arviointien metodisen korkean tason. Tällä hetkellä tilanne ei ole siinäkään mielessä otollinen, koska Suomessa ei tällä hetkellä ole arviointitutkimukseen pelkästään painottuneita akateemisia professoritason oppituojeja.

Kehittämiseen ja oppimiseen liittyvät motiivit ovat aina tavallaan haastavia ja pulmallisia. Olen itse huomannut, että tämän motiivin ääneen sanominen yleensä vapauttaa arvioinnin kohteet valvonnan pelon ikeestä. Suomalaisessa mielenlaadussa elää voimakkaasti pelko tulla nolatuksi

ja rangaistuksi, eikä suhtautuminen arviointitoimintaan tee tässä suhteessa poikkeusta. Olen itse ollut mukana muutamissa ohjelmaston arviointihankkeissa, joissa on menneet varsinaisen arviointihankkeen käynnistämiseen kohtuullisen paljon aikaa juuri tämän pelon hälventämiseen. Ohjelmasta vastaava projektijohtaja on alun perin luullut joutuvansa tiille virheistään. Sitten olen ollut myös mukana sellaisissa ohjelma-arvioinneissa, joissa alusta alkaen on puhuttu auki arviointitoiminnan oppimiseen ja toiminnan parantamiseen tähtäävä tavoite. Esimerkkinä tästä käy vaikkapa Paavo Lipposen II hallituksen aluekeskusohjelman arviointihanke, jonka tiimin johtajana olen toiminut kesästä 2001.

Tilintekovastuu (accountability) on tärkeä näkökohta arviointitoiminnassa. Ohjelma-arvioinnissa kyse voi tilintekovastuun osalta olla siitä, missä määrin julkisin varoin rahoitetuilla ohjelmilla saadaan aikaiseksi lisäarvoa. Tähän motiiviin liittyy paljon muitakin teemoja, muun muassa arviointitulosten käyttöä koskeva avoimuus ja raporttien saatavuus (joko sähköisessä muodossa Internetistä pdf-tiedostoina tai fyysisinä painettuina raporteina). Itseeni esimerkiksi teki suuren vaikutuksen hiljattain se – kirjoittaessani muun tutkijaryhmän kanssa vertailevaa analyysia julkisen yritysrahoituksen vaikuttavuusarviointimalleista – miten avoimesti Isossa-Britanniassa periaatteessa kaikki julkista toimintaa koskevat ja julkisin varoin toteutettujen arviointiraporttien tulokset ovat kaikkien saatavilla ja periaatteessa julkisia (ks. Virtanen ym. 2000).

Tilintekovastuun vaatimus voi saada aikaan myös yllättäviä reaktioita arviotavassa kohteessa. Eräässä institutionaalisisessa arvioinnissa – jonka kohteena oli valtioapua saava yleishyödyllinen yhteisö – kyseisen toimintaorganisaation johtaja oli suoraan sanoen hysteerinen odottaessaan omaa organisaatiotansa koskevan arviointiraportin käsiinsä. Tutkijakollegani mukaan kyseinen johtaja suoranaisesti painosti kollegani johtamaa arviointitiimiä kirjoittamaan kyseiselle organisaatiolle suotuisia johtopäätöksiä. Kun paineet ovat kovat, otetaan käyttöön uskomattomia puolustuskeinoja.

Pattonin esittämä luokittelu ei luonnollisesti ole tyhjentävä, vaan pikemminkin suuntaa antava. Pattonin ajatuksia voi täydentää toteamalla, että ohjelma-arvioinnin motiivit voi myös liittää veloitteeseen tai normiin (nykyään on Suomessakin muotia säättää arviointivelvoitteet lain tasoiseksi

veloitteeksi tai vähintäänkin hallinnollisiksi toimintatapaohjeiksi), mikä sinänsä voi merkitä edistysaskelta, mutta ei välttämättä. Esimerkiksi lääninhallitusten tehtävien muutosten yhteydessä 1990-luvun jälkipuoliskolla kuviteltiin, että lakisääteisyys tekee entisistä lääninhallitusten virkamiehistä arviointitoiminnan ammattilaisia. Näin ei suinkaan tapahtunut, vaan osaamisen kasvattaminen on vienyt viisi vuotta ja vasta nyt – vuonna 2002 – ollaan tilanteessa, jossa lääninhallitusten peruspalvelujen arviointitoiminnan alkaa muistuttaa kriittistä arviointia mukaan lukien tarkoituksenmukainen menetelmien ja raportointitapojen käyttö.

Arvioinnin motiivit voivat liittyä toiminnan legitimaatioon tai jopa markkinointiin. Ensin mainitussa on kyse siitä, että esimerkiksi jonkin tietyn yhteiskuntapoliittisen ohjelman toteutusta koskeva strateginen päätös – operatiivisen painopisteen muutos, rahoitukseen liittyvä merkittävä määrärahan karsinta, tietyt henkilövalinnat – voidaan tehdä etukäteen ja päätökselle haetaan justifikaatio ulkoisesti teetetyt arviointihankkeen kautta. Markkinointiin liittyvät motiivit puolestaan voivat esimerkiksi tarkoittaa sellaista tilannetta, jossa jonkun tietyn yhteiskuntapoliittisen ohjelman merkitystä lanseerataan siitä tehdyn etukäteisarvioinnin (ex ante) avulla. Ohjelma ikään kuin "myydään" eri osallisille. Pertti Ahonen (2001b, 69) pukee ajatukseni hyvin sanoiksi todetessaan, että arviointi on tietyissä tapauksissa imagopoliittikkaa, jolla evaluaation sidosryhmät osoittavat eettisyyttään, kuten säästäväisyyttään, huolenpitoaan osakseen tulleista voimavaroista sekä siitä, että ne täyttävät tulivelvollisuutensa.

#### 4. ARVOSTELMIEN MUODOSTAMISEN LOGIIKAN ONGELMAKOHDAT

Oli ohjelma-arvioinnin toteuttamiseen liittyvä motiivi mikä hyvänsä, arviointitoiminnan prosessi etenee periaatteessa mukaillen seuraavanlaisia kokemukseni perusteella aina seuraavanlaisesti. Arviointi lähtee liikkeelle siitä, että evaluaattori tuottaa arviointitehtävästä, käytössä olevista tutkimusmenetelmistä ja –aineistoista riippuen joukon tosiasiaväittämiä, ikään kuin raakahavaintoja. Sen jälkeen evaluaattorin ammattitaidosta riippuen raakahavaintoja pitäisi pystyä tulkitsemaan, selittämään ja kytkemään osaksi laajempaa kontekstia. Kolmanneksi arvioinnin tekijän tulee

muodostaa arvostelmat eli onnistuneisuutta kuvastavat johtopäätökset (tehdn jossain vaiheessa kriittisen päätöksen hyvän ja huonon kriteerin välillä), ja lopulta muovata arviointikysymyksiin, -menetelmiin, -aineistoihin, näistä johdettuihin raakahavaintoihin ja arvostelmiin palautettavissa olevat toimenpidesuosituksia ja kehittämisehdotukset. Ilman kehittämisehdotuksia arviointiprosessi on hyödytöntä toteuttaa.

Arvostelmien muodostamisen logiikka ei kuitenkaan koskaan tapahdu yksinkertaisesti tai yksiselitteisesti. Tässä on yksi ohjelma-arvioinnin – ja laajemminkin ottaen arviointitutkimuksen – keskeinen haaste. Ohjelma- ja projektiarvioinnissa on monia tiedonmuodostukseen liittyviä ongelmia, niin kuin arviointitutkimuksessa yleensäkin. Nostan seuraavassa esille kaksi periaatteellista kysymystä, joihin olen törmännyt viimeisen viiden vuoden aikana erilaisissa arviointihankkeissa, joissa olen ollut mukana.

Ongelmista ensimmäinen liittyy siihen, ettei arviointihankkeissa aineistoista johdettuja tosiasioista voida suoraan johtaa arvostelmia (eli oletuksia asioidentilojen hyvyydestä tai huonoudesta). Tämä pulma ei ole uusi, vaan filosofit ovat kuluneiden vuosisatojen muodostaa kantaan siihen.

Esimerkiksi brittiläiset George Berkeley ja David Hume yrittivät jo 1700-luvulla osoittaa, etteivät aistielimet kuvasta meitä ympäröivää suoraan ja välittömästi. Esimerkiksi Berkeley kiteytti tämän ilmiön lausumaksi *esse est percipi* – olemassa oleminen on koetuksi tulemista (ks. Gerholm & Magnusson 1983). Humelle oivalus luonnontieteellisten totuuksien rajoitetusta arvosta merkitsi isoa pulmaa. Hänen mielestään viisas ihminen suhteuttaa ajattelunsa ja uskonsa todistuksiin. Tällöin Humelle avautuva näköala liittyy kysymykseen, täytyykö tiettyä tapahtumaa (syytä) ehdottomasti seurata toinen ilmiö (seuraus).

Mikä on siis Humen kokemuksen merkitys? Mikä selittää yksittäisten havaintojen ja ilmiöiden välistä suhdetta? Kyseessä on itse asiassa musta laatikko yksittäisestä empiirisestä havainnosta muihin empiirisiin havaintoihin. Arviointitutkimuksen näkökulmasta Humen ajatuksen – sananmukaisen kokemuksen – sovellus voisi olla pohdinta siitä, miten päästään empiirisistä havainnoista toimenpidesuosituksiin. Miten tämä tapahtuu? Pelkään pahoin, ettei yksiselitteistä vastausta voi antaa, vaan kyse on arvioinnin teki-

jän ammattitaidosta (tämän ulottuvuuksista esim. Virtanen 2001).

Kyse on siitä, että arvioinnin tekijä on kokeuksensa myötä kehittänyt arvioivaa käytäntöä, johon perustuen hän pystyy tekemään tarkoituksenmukaisella tarkkuudella toimenpidesuosituksia, tässä tapauksessa ohjelman tai projektin uudelleen suuntaamiseksi tai parantamiseksi. Pertti Ahosta (2001a, 114-115) värikkäitä muotoiluja mukaillen voisi sanoa, että ”ahnaat käsityöläisevaluaattorit”, ”urheat evaluaattorisoturit” ja ”viisaat filosofievaluaattorit” toimivat arviointitutkimustoimeksiannoissaan omien kompetenssiensa puitteissa eikä todellakaan ole yksiselitteistä, miten arviointikysymyksistä päädytään johtopäätöksiin ja siitä sitten käytäntöjen muuttamista koskeviin suosituksiin.

Vakavammin ja vähemmän lennokkaammin puhuen kyse on arviointiosaamisesta, mihin jo edellä viittasin. On tärkeätä tehdä ero arviointiosaamisesta (arviointitutkimuksen tekemisen kohdalla) välineellisen ja teknisen hyvän välillä (jaottelu palautuu muun muassa von Wrightin 2001, 49-52, tekemään jaotteluun), jota avaan seuraavassa hieman.

Välineellinen hyvä (ammattitaito) liittyy pääasiassa erilaisiin työkaluihin, instrumentteihin ja välineisiin, kuten arviointitutkimuksessa erilaisten tutkimusmenetelmien hallintaan ja esimerkiksi moninaisten tietokoneohjelmien saatavuuteen. Tekninen hyvä liittyy taas arviointitutkimuksen tekijän kykyyn ja taitoon. Taustalla on siis oletus, että asiaa, jossa joku on hyvä, voidaan kutsua taidoksi laajassa merkityksessä, ja tällöin ollaan lähellä sitä mitä kreikan kielessä tarkoitetaan sanalla *tekhné*. Tekniseen hyvään liitetään tavallisesti sellaisia määreitä kuin taitavuus, kykenevyys tai aikaansaaminen. Olen melko varma, että tällaisia taitoja arvostetaan arviointitutkimuksen tekijöiden piirissä. Oma lukunsa on se, missä tällaisia taitoja voi opiskella, harjoittaa ja missä määrin nämä taidot ovat luontaisia.

Arviointitutkimuksissa raportoitavien johtopäätösten ja toimenpidesuosituksien syntyminen ei siis ole yksiselitteistä. Tiedostavatko kaikki arviointitutkimuksia tekevät tämän tosiseikan? Ajatellaanko niin, ettei arviointitutkimuksen kirjoittamisprosessi olisi määriteltävissä, vaan se on olemassa sui generis, ikään kuin redusoimattomana kirjoittamisen ja tulkinnan muotona.

Tässä on ohjelma-arvioinnin – ja ylipäänsä arviointitutkimuksen – keskeinen haaste ja kehiti-

tämialue. Esimerkiksi arvoteoria voi tarjota paljon yhteiskuntatieteelliselle arviointitutkimukselle. Georg Henrik von Wrightin jo 1950-luvulla lanseeraama deonttisen logiikan eli normatiivisten käsitteiden formaalinen teoria voi antaa selkänöjää tämättyypisissä ongelmissa.

Perustan käsitykseni siihen, että arvokäsitteet tulkitaan varsin usein normatiivisiksi käsitteiksi. Wrightin ajatukset auttavat ymmärtämään tämän virhetulkinnan laajuutta – taipumusta sekoittaa arvokäsitteet normatiivisiin käsitteisiin. Wrightin (2001, 248) mielestä esimerkiksi painotettaessa tieteen arvovapaata luonnetta perustelut liittyvät ennen muuta siihen, että tiede voi kertoa meille, miten asiat ovat, mutta ei sitä, miten asioiden pitäisi olla.

Edellä sanottu pätee Wrightin mielestä yleensä tieteellisessä ajattelussa. Kun taas otetaan huomioon arviointitutkimuksen – ja sen sisällä ohjelma-arvioinnin – erityisluonne yhteiskuntapolitiikan muutosta muovaavana tekijänä havaitaan erittäin pulmallinen näkökohta. Arviointitutkimuksessa tämä Wrightin jaottelu ei enää toimikaan. Arviointitutkimukselta odotetaan tavallisesti konkreettisia toimenpidesuosituksia siitä, miten asioiden pitäisi olla. Hume'n hengessä puhuvien filosofien puhuessa tosiasioiden ja normien, olemisen ja pitämisen, välisestä suhteesta he todella puhuvat arviointitutkimuksen "mustasta laatikosta".

Kokonaan oma lukunsa on, missä määrin arvot tai normit ovat perustavampia. Tästä on länsimaisessa filosofian historiassa taitettu sanalista peistä vuosituhansia. Arvoetiikalla, siinä muodossa kuin ymmärrämme sen von Wrightin (2001) tekstien kautta, voidaan ajatella näkemystä, jonka mukaan arvo on perustava ja normit tavalla tai toisella "esiin nostettavia" asioidentiloja tutkailemalla arvoja tai suoraan niihin nojautumalla.

Paljastin edellä, että ohjelma- ja projektiarviointiin liittyy kaksi periaatteellista ongelmaa, joita tarkastelen tässä yksityiskohtaisemmin. Se, että tosiasioista ei suoraan voida johtaa toimenpidesuosituksia on pulma, jonka tiedostaminen on tarpeellista. Pelkkä tiedostaminen ei vielä johda pitkälle. Meidän tulee ottaa tarkasteluun toinen, ehkä vieläkin periaatteellisempi kysymys. Arvioinnit ovat riippuvaisia henkilöistä ja heidän preferensseistään.

Kun edellä viittasin tosiasian ja tulkinnan väliin eroon, tarkoitin tällä lähinnä prosessuu-

aalisen ymmärryksen tarvetta arviointitutkijalla. Seuraavaksi kysymme, mikä on teko, mikä on tekemiseen johtava tapa ja miten näitä lottuvuoksia arvioidaan. Teko on tahdonvaraisesti tai tuotettu muutos maailmassa. Käytännössä tämä voisi tämän artikkelin teeman osalta viitata jonkin yhteiskuntapolitiittisen ohjelman seurauksena aikaan saatu muutos. Konkreettisesti tämä voisi tarkoittaa jonkin aluepolitiittisen ohjelman kohdalla aluerakenteen muutosta jollain sosioekonomisella mittarilla tarkasteluna. Tekemisen tapa puolestaan on jokin teko tai toiminta, jonka avulla kyseinen yhteiskuntapolitiittinen ohjelma toteutetaan. Kyse voi olla esimerkiksi taloudellisten panosten kanavoimisesta – pitäytyessämme samassa aluepolitiittisen ohjelman esimerkissä – kyseisille alueille, käynnistetyistä elinkeinorakennetta vahvistavasta pilottiprojektitoiminnasta, erilaisista paikallisille toimijoille suunnatusta koulutusinterventiosta jne.

Ajan takaa tällä ajattelukaavalla sitä, että ihmiset – tässä tapauksessa ohjelma-arvioiteja tekevät – tulkitsevat asioita eri tavoin. Kyse ei välttämättä ole havaintojen tekijöiden hyveistä tai pahansuopuudesta. Ihmiset yksinkertaisesti jäsentävät maailmaansa erilaisista horisonteista ja vaikuttavia tekijöitä tähän on lukuisia, kuten esimerkiksi tulkinnan tekijöiden käsitys hyvästä ja pahasta, poliittinen maailmankatsomus, ammattitaito, kokemus jne.

Kuten jo johdantojaksossa totesin, yhteiskuntatieteellisessä keskustelussa tätä on pohtinut kenties kaikkein pisimmälle Pierre Bourdieu, jonka ajatukset refleksiivisestä sosiologiasta tunnetaan, tunnustetaankin ja jopa ymmärretään Suomessa varsin yleisesti (ks. Bourdieu 1990 ja Bourdieu & Wacquant 1992).

Arviointitutkimuksessa tulkinnan tekijän ammattitaidon varassa on oivaltaa ero teon tuloksen ja vaikutuksen välillä. Formaalisesti määriteltynä teon tulos on "se asiointi, joka täytyy saavuttaa, jos haluamme perustellusti sanoa teon tulleen tehdyksi (...) Teon seuraukset taas ovat asiointiloja, jotka kausaalisen välttämättömyyden ansiosta astuvat voimaan kun teko on tehty" (von Wright 2001, 192).

Kuten jäljempänä yritän osoittaa, ei ole lainkaan itsestään selvää, että yhteiskuntapolitiittisten ohjelmien arvioinneissa pystytään kovinkaan yksiselitteisesti erittelemään yksittäisten ohjelmien vaikutuksia, useiden samanaikaisesti toimeenpantujen ohjelmien vaikutuksista puhumattakaan.

Hyvänä esimerkkinä jälkimmäisestä on Paavo Lipposen II hallituksen aluepoliittisten ohjelmien vaikutusten (mitä saatiin aikaan) ja vaikuttavuuden (mitä saatiin aikaan suhteessa ohjelmille asetettuihin tavoitteisiin) arviointi. Kun käynnissä on valtakunnallinen Osaamiskeskusohjelma, Aluekeskusohjelma, Seutuhanke ja joukko muita pienempiä (hallituksen ja muiden toimijoiden, esimerkiksi Sitran ja Raha-automaattiyhdistyksen, rahoituksella tehtäviä) uudistushankkeita, joita toteutetaan ympäri Suomea samoilla alueilla, on helppo ymmärtää arvioinnin tekijöiden ongelmat arviointien tekemisessä.

Edelleen voidaan pohtia, millaisia asioita eri arvioinnin tekijät joutuvat käytännössä ohjelma-arvioinneissa ratkaisemaan. Ovatko teot, tulokset ja vaikutukset esimerkiksi toivottavia, suotavia tai peräti vahingollisia. Voidaan ajatella, että ohjelma-arvioinneissa törmätään ohjelmiin, joissa itse ohjelman toimeenpano on sinänsä hyvin toivottu, samoin oletetut tulokset. Jos taas pelkkä toimeenpano on sinänsä toivottua, silloin sen harjoittaminen ("tekeminen") on sinänsä toivottu asia. Mihin arvioinnin tekijä siis kääntää katseensa ja millaisiin asioihin huomio kiinnitetään? Yksi varsin yleinen kysymys, johon arvioinnin tekijät joutuvat ratkaisemaan liittyy siihen, millainen yleisarvio jonkun ohjelman toteutuksesta, tuloksista ja vaikutuksista annetaan. Väitän, että poikkeuksetta yhteiskuntapoliittiset ohjelmat tuottavat erilaisia tuloksia ja vaikutuksia, joita arvioinnin tekijät luokittelevat joko hyviin tai huonoihin, toivottuihin tai ei-toivottuihin seurauksiin. Millainen yleisarvio tässä tapauksessa annetaan? Voidaanko laatia "arvoyhteenvetoa" ohjelman lopullisesta arvosta siihen tapaan, että lasketaan yhteen plussat ja miinukset? Entä miten nämä painotetaan? Jo pelkästään tällaisten kysymysten ääneen lausuminen riittää toivottavasti osoittamaan arviointien tekemisen vaikeuden.

Olen pahoillani, ratkaisua edellä esittämiini kysymyksiin ei ole olemassa, korkeintaan suuntaviittoja. Von Wright (2001, 233) puhuu esimerkiksi hyveen tiestä, joka sopii nähdäkseni ainakin muunnoksena arviointitutkimuksen tekijöillekin. Von Wrightille hyveen tie on via media aurea, kultainen keskitie kahden äärimmäisyyden välissä. Tämä on hieno käsitteellinen oivallus, joka itse asiassa palautuu Aristoteleen ajatteluun. Arviointitutkimuksen tekijän harteille jää tässäkin tapauksessa iso vastuu. Löydöksistä pitää raportoida rehellisesti ja tasapuolisesti, ja kuitenkin mitään

yleispätevää ohjetta ei ole olemassa, saati saatavilla.

## 5. OHJELMA-ARVIOINNIN TEKEMINEN

Avaan ja perustelen seuraavassa lyhyesti neljää näkökulmaa ohjelma- ja soveltuvien osin myös projektiarviointien tekemiseen. Näkökulmat ovat kiteytyneet kokemuksen kautta, ohjelma-arviointeja tekemällä. Samalla nämä näkökulmat ovat tavallaan vastausta koko artikkelin perustematikkaan – mistä ohjelma-arvioinnissa on kysymys ja mitä sen suorittamisessa tulisi ottaa huomioon?

(i) Riippumatta ohjelma-arviointien kulloisistakin motiiveista, arviointihankkeiden keskeisenä päämääränä tulee olla korkea hyödynnettävyys.

Kärjistäen voisi sanoa, että jos arviointeja ei voi hyödyntää, niitä ei kannata tehdä. Arviointien hyödynnettävyyden aste ratkeaa monen tekijän yhteisvaikutuksesta. Näitä tekijöitä ovat muun muassa arviointikysymysten määrittely kaikkien osallisten kanssa, evaluaattorin substanssiosaaminen arvioitavasta kohteesta, arviointiprosessin laadukas onnistuminen (ml. aineistohankinta ja analyysit) ja arviointiraportin keskeisten havaintojen, johtopäätösten ja toimenpidesuosituksen tarkoituksenmukaisuus (ks. tästä esim. Valovirta 2002). Keskityn näistä tekijöistä seuraavassa viimeksi mainittuun.

Jotta ohjelma-arvioinnin tulokset olisivat hyödynnettävissä, edellyttää tämä arviointitutkimuksen johtopäätöksenä syntyviä johtopäätöksiä tai toimenpidesuosituksia. Kysymys kuuluu, millaisia nämä suositukset ovat, mihin ne liittyvät ja mistä ne eivät puhu? Miksi evaluaattori vaikenee jostain asiasta? Hyvin harvoin näkee, että arviointitutkimusraporttia kirjoittava olisi todella syvällisesti pohtinut erilaisten toimenpidesuositusten tasoja. Tulkittaessa johtopäätösten ja toimenpidesuositusten muotoilua normien kielellisen muotoilun tapoina, voidaan ensininkin todeta, että tapoja on useita ja että tavat ovat toisiinsa nähden limittyneitä (ks. mukailen von Wright 2001, 251-252, ja hänen tulkintansa normien kielellisistä vastineista).

Ensimmäinen normien kielellinen vastine koostuu imperatiivisen tapaluokan lauseista, kuten esimerkiksi "Muuta ohjelman kolmatta tavoitetta",



"lisää taloudellisia budjettivaroja kanavoitavia panoksia ohjelman toteutusvuonna 2003" jne. Tapaluokka muistuttaa käskyä. Voidaan perustellusti kysyä, onko tällainen toimenpidesuosituksen tapa järjestyksessä, koska evaluuatuori ei voi olla päätöksentekijä, vaan korkeintaan tuottaa tietoa ja toimenpidesuosituksia päätöksenteon pohjaksi?

Toista normien kielellisistä vastineista voi kutsua deonttisiksi lauseiksi (kreikan kielen sanasta *deón* "on tarpeen", "pitää"). Näissä tapauksissa toimenpidesuositukset sisältävät "deontisia" eli velvoittavia apuverbejä kuten "pitää", "voi" ja "ei pidä". Tällöin esimerkit ohjelma-arvioinnin maailmasta voisivat olla jotain sellaista kuin "ohjelmajohtajan ei pidä kuunnella osaprojektien projektipäälliköiden mielipiteitä resurssienjaossa siinä määrin kuin vuonna 2001 on tapahtunut", "ohjelman seurantaryhmän pitää jätävämmän ohjata ohjelmaa" ja "ohjelman osaprojektien vetäjät voivat osallistua paikalliseen poliittiseen päätöksentekoon, se on peräti toivottavaa". Aste-ero edelliseen tapaluokkaan on ilmeinen.

Kolmas normien kielellinen vastine ovat anakastiset lauseet (kreikan kielen sanasta *anankastós*, "välttämätön", "pakollinen"). Nyt toimenpide-ehdotusten sävy on määräävämpi, koska muotoilussa käytetään pakottavia apuverbejä: "Eduskunnan on annettava varoja kaksinkertainen määrä ohjelman toteuttamiseen vuonna 2003", "ohjelmajohtaja ei voi määrällä osaprojektien vetäjiä siinä määrin enää vuonna 2003 kuin on tapahtunut vuonna 2001" jne. Oleellista tässä von Wrightilta (mt.) lainatusta typologiassa on se, että arviointitutkimusraporttia kirjoitettaessa olisi kirjoittajan oltava tietoinen eri luokkien aste-eroista. Vallitseva tilanne näyttää suomalaisessa – ja tasapuolisuuden nimissä on syytä todeta, kansainvälisessäkin ohjelma-arviointia koskevassa raportoinnissa – melko sekava. Toimenpidesuosituksia laaditaan paljonkaan miettimättä niiden normatiivisten muotoilun tapoja.

Valinta deonttisten ja anakastisten muotoilujen välillä on tapauskohtaista ja riippuu pitkälti evaluuatuorin uskottavuudesta ja ammattitaidosta, mikä kielellinen vastine otetaan käyttöön. Olen vakuuttunut, että ohjelma-arvioinnin tekijän tulee joka tapauksessa aina pyrkiä täsmälliseen ilmaisuun toimenpidesuosituksia muotoillessaan.

Toimenpidesuositusten tulee olla ymmärrettäviä, realistisia ja toteutettavissa olevia. Mikään ei ole tuskastuttavampaa kuin lukea sinänsä hyvin

tehtyjä arviointiraportteja, joissa yhteenvedot ja raportin lopussa oleva toimenpidesuosituksen lista on pelkkää sananhelinää "ehkä...mutta...jos...toisaalta" –tyyliin. Esimerkiksi EU-rahoitteisten ohjelma-arviointien raportoinnissa tulee raportin alkuun liittää niin sanottu executive summary –osio, joka on laajempi kuin perinteinen A4 –yhteenvedo. Siinä tulee kertoa mitä on arvioitu, miten arviointi on suoritettu, mitkä olivat arvioinnin keskeiset tulokset ja havainnot sekä mitkä ovat evaluuatuorin toimenpidesuositukset.

(ii) Ohjelma-arvioinnissa pitää olla aina kolme ajallista ulottuvuutta ja ohjelman seurantajärjestelmä ja arviointi on rakennettava toisiaan hyödyntäviksi

Ohjelma-arviointi ei voi koskaan olla ajallisesti yksilulotteista. Arviointitutkimusta koskeva "common sense" kuitenkin on, että arviointi on pelkästään jälkikäteisarviointia – arvioidaan sitä, mitä tapahtui ja mitä opittiin.

Liioittelematta voitaneen sanoa, että yhteiskuntapoliittisten ohjelmien suunnittelu tapahtuu Suomessa vielä tätäkin artikkelia kirjoitettaessa enemmän tai vähemmän sporadisesti ja että arvioinnit rakennetaan niihin hyvin tapauskohtaisia oppeja soveltaen. EU-rahoitteiset ohjelmat ovat kuitenkin vuodesta 1995 tuoneet tiettyä menetelmällistä ryhtiä ainakin tiettyjen hallinnonalojen ohjelmasuunnittelukäytäntöihin.

Ohjelmien suunnittelussa on itse asiassa kaksi vaihtoa: varsinainen ohjelman muotoilu ja suunnittelun ohjelman kriittinen arviointi (esim. Vedung 1997). Ohjelma siis suunnitellaan ja ennen sen varsinaista käynnistämistä sen toteutettavuus analysoidaan. Tätä kutsutaan vaihtoehtoisesti joko etukäteisarviointiksi tai *ex ante* –arviointiksi. Kuten edellä viittasin, tekee tällainen ajattelutapa vasta tuloaan suomalaisen hallintokulttuuriin ja EU-rahoitteisten ohjelmien toimeenpano on osaltaan toiminut uuden ajattelun muurinmurtajana.

Ohjelma-arvioinnin toinen ajallinen ulottuvuus ajoittuu ohjelman toimeenpanovaiheeseen. Tapauksesta riippuen ohjelmien arvioinnissa keskitytään tällöin sen analysoimiseen, miten ohjelma on edennyt kohti sille asetettuja tavoitteita. Tätä arvioinnin vaihtoa kutsutaan vaihtoehtoisesti joko jatkuvaksi arviointiksi tai *ex nunc* –arviointiksi.

Usein on vaikea erottaa seurannan ja arvioinnin välistä eroa nimenomaan juuri jatkuvan arvioinnin kohdalla. Nähdäkseni Suomessa toteutettujen EU-rahoitteisten ohjelmien toimeenpano



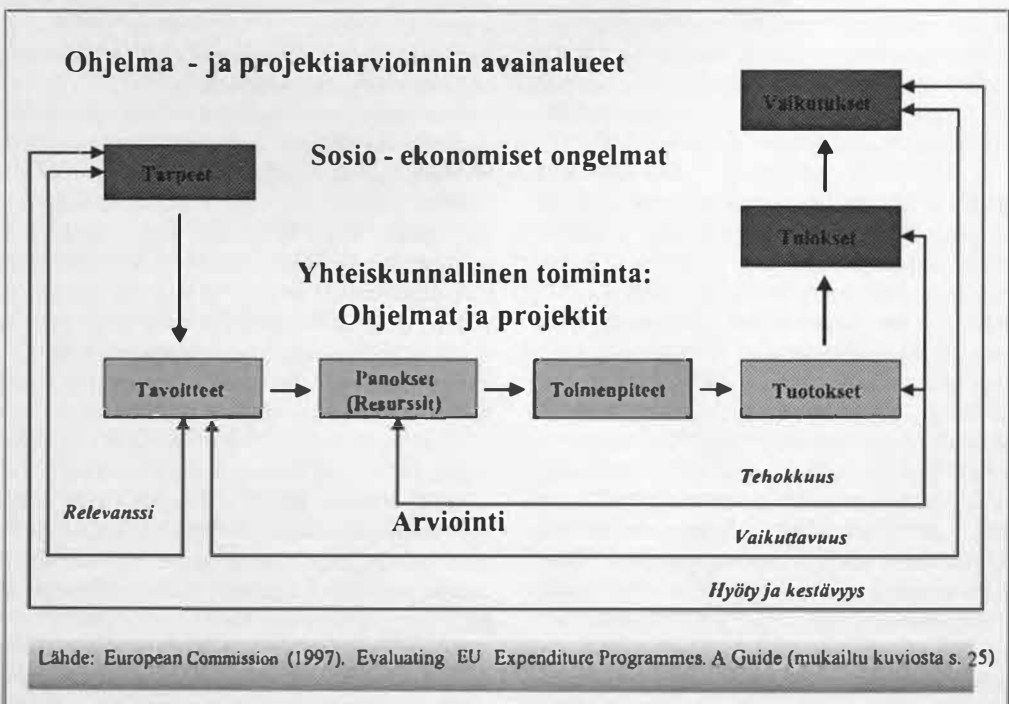
tarjoaa hyvän ja konkreettisen esimerkin seurannan ja arvioinnin välisestä erosta. Esimerkiksi käynnissä olevalla rakennerahastokaudella Euroopan sosiaalirahaston Tavoite 3 –ohjelman seurantarjestelmien kehittämisestä ovat vastuussa niin sanotut rakennerahastotoiminnan vastuuministeriöt (kuten työministeriö, kauppa- ja teollisuusministeriö, sisäasiainministeriö), jotka ovat rakentaneet ohjelmien seurantaan palvelevat tietojärjestelmät, joihin kerrytetään projektikohtaista tietoa kyseisen ohjelman suoritteista. Nämä tiedot – joihin siis sisältyvät ohjelman projekteja koskevat suoritteet (projekteissa aloittaneiden ja lopettaneiden määrät jne.) ja maksatustiedot – raportoidaan varsinaisten ulkopuolisten evaluaattoreiden raporttien kanssa ohjelman seurantakomitealle säännöllisin väliajoin. Ideaalinen tilanne on sellainen, jossa ohjelman evaluaattorit voivat hyödyntää seurantarjestelmän tuottamaa tilastotietoa, mikä puolestaan edellyttää seurantarjestelmien tuottaman tiedon luotettavuutta.

Kuviossa 1 on esitetty ohjelma-arvioinnin avainalueet ja kuvion johtopäätökset pätevät soveltuvin osin myös projektiarvioinnin kohdalla. Kuvion

käsitteistä panokset, toimenpiteet ja tuotokset ovat seurannan kannalta keskeisiä, kun taas arvioinnissa ollaan kiinnostuneita tuloksista ja vaikutuksista. Kuviossa mainitut relevanssin, tehokkuuden, vaikuttavuuden ja hyödyn käsitteet ovat puolestaan aivan keskeisiä jälkikäteisarvioinnin tarkastelu-ulottuvuuksia, joihin jälkikäteisarviointia tekevän evaluaattorin pitäisi kiinnittää huomionsa.

Ohjelma-arvioinnin kolmas ajallinen ulottuvuus liittyy siis edellä sanotun mukaisesti ohjelmien päättymisvaiheeseen. Tässä vaiheessa analysoidaan se, mitä ohjelmalla saatiin aikaiseksi. Tätä arvioinnin vaihetta kutsutaan joko jälkikäteisarviointiksi tai ex post –arviointiksi. Jälkikäteisarviointiin liittyy paljon ajatuksellisia haasteita, joista keskeisimpiä lienee ohjelman päättymisen ja arvioinnin ajoittumisen välinen ajallisen viiveen huomioon ottaminen. Milloin ylipäänsä on mahdollista järkevää tehdä arviointi? Puoli vuotta, vuosi tai kaksi vuotta ohjelman päättymisestä? Vastaus tähän kysymykseen riippuu ohjelmasta – mitään yleispätevää ohjetta ei voi käsittääkseni antaa.

Suomessa vallalla oleva ohjelma-arviointien



**Kuvio 1**

toteutuksen tilanne on sellainen, että jälkikäteisarviointien asema on ylivertainen suhteessa muihin ajallisiin ulottuvuuksiin. Kansallisissa yhteiskuntapoliittisissa ohjelmissa – siis pelkästään kansallisin budjettivaroin toteutetuissa – etukäteisarvioiteja ei juurikaan toteuteta ja jatkuvatkin arvioinnit ovat enemmän tai vähemmän tapauskohtaisesti viritettyjä asetelmallisessa tai metodisessa mielessä. Tässä on mielestäni iso kehityshaaste kansalliselle ohjelma-arvioinnille. Suomalaiset ohjelma-arvioinnit ovat tähän asti pääsääntöisesti olleet jonkinlaisia yhdistelmiä jatkuvasta tai jälkikäteisarvioinnista ja kokemukset arviointien hyödyntämisestä ovat vaihdelleet täysin hyödyttömästä erittäin hyödyllisiin. Viime vuosina on valtionhallinnossa alettu kiinnittää huomiota ohjelma-arviointien tilaamiseen liittyvään osaamiseen (Vuorela 1997; Valtiovarainministeriö 1999), mikä osaltaan vuosien saatossa edistää uudentyypin ohjelma-arviointikulttuurin syntyä.

**(iii) Arviointitiedon ja tuloksellisuusinformaation synergiahyötyjen paikantaminen ja arviointitoiminnan ja tuloksellisuusinformaation integroiminen ohjelmien ja projektien ohjaukseen**

Arviointitoiminnan kytkeminen osaksi ohjelmien – ja myös projektitasolla yksittäisten hankkeiden – ohjausta on oikeastaan jo tullut ääneen lausutuksi edellisissä kohdissa (i) ja (ii). Vaikka on tunnettua (ks. esim. Vakkurin 2001 havainnollinen esitys), että tuloksellisuusinformaation tiedonmuodostuksessa ja tiedon hyödyntämisessä on monenlaisia pulmallisuuksia, arviointitoiminnan kautta saadulla tiedolla pitää olla kiinteä yhteys ohjelmien ohjaukseen.

Käytännössä edellä sanottu tarkoittaa arviointitiedon saattamista sellaiseen muotoon, että ohjelmien ohjauksesta vastaava taho – seuranta-, ohjaus- johto- tai muu vastaava ryhmä – voi todella tehdä päätöksiä ohjelmien joko suunnan tarkistuksista tai peräti suunnan muutoksista. Miten tämä sitten voisi tapahtua ja minkä ajatuksellisen viitekehysten avulla? Vastaus löytyy lähempää kuin saattaa arvata. Tai oikeastaan paikasta, josta vastausta ei ole osattu hakea.

Suomessa nimittäin sovelletaan – ainakin vielä nykyisellään – valitettavan vähän niin sanottua Logical Framework (joskus näkee käytettävän lyhennettä LogFrame) –lähestymistapaa ohjelmatasoisessa suunnittelussa, seurannassa ja

arvioinnissa. (Erityisesti kehitysyhteishankkeiden) projektimaailmassa jo 1970-luvulla yleistynyt tapa mieltää hankesykli sopisi mitä mainoimmin myös ohjelmatasoiseen yhteiskuntapoliittisten interventioiden suunnitteluun, seurantaan ja arviointiin työkaluksi, joka varmasti toisi mukaan ohjelmien toteuttamiseen uutta käsitteellistä ja tilintekovastuuta huokuvaa ryhtiä. Tiedossani ei ole Suomessa montaakaan ministeriötä tai eri hallinnonalan tilivirastoa, joka soveltai Logical Framework –ajattelua ohjelmatyössään. Ulkoasiainministeriö – ja siellä erityisesti sen kehitysyhteistyösasto – edustaa tässä kuitenkin poikkeusta.

Kuviossa 2 esitetään Logical Framework –matriisin perusmalli, joka on tuttu useimmista aihealueen perusesityksistä. Tässä yhteydessä voidaan mainita vaikkapa Basil Edward Cracknellin (2000) mainio esitys lähestymistavan perusolettamuksesta ja soveltamisalueista. Ideana on siis eriyttää – tässä tapauksessa – ohjelmatason yleinen päämäärä, varsinainen tavoite, toiminnot, oletetut tulokset ja oletetut vaikutukset toisistaan ja miettiä näille validit ja tulosindikaattorit, niitä palvelevat tietolähteet (eli verifiointipohja) ja paikantaa yleistä päämäärää, varsinaista tavoitetta, toimintoja ja tuloksia koskevat olettamukset (siis tavallaan riskianalyysin tekeminen).

Näin menetellen ohjelmien ja projektien ohjausfunktio terävöityy ja samalla tulee myös pohdittavaksi ja päätettäväksi seurannan ja arvioinnin väliset erot ja roolit. En esitä Logical Framework –ajattelutapaa tässä yhteydessä mitenkään kevytmielisesti hetkellisen päähänpiston johdosta, vaan uskon olevani riittävän kokenut väittääkseni, että tällaisella yksinkertaisella suunnittelu-, seuranta- ja arviointimekanismilla moni yhteiskuntapoliittinen (ja varsinkin hallitusohjelmatasoinen) ohjelma ja hanke olisi ollut Suomessa paljon järkevämmällä ja uskottavammalla tolalla kuin mitä ne nyt ovat.

Lisäksi voidaan todeta, että näen edellä sanottulla myös laajempia julkisen hallinnon kehitysvaiheeseen liittyviä kontekstuaalisia implikaatioita. Suomen osalta voidaan sanoa, että julkishallinnossa ollaan viime vuosina niin kunnissa kuin valtionhallinnon organisaatioissa modernisoitu luotuneita toiminnan- ja tulosityhteisjärjestelmiä, mikä osaltaan on johtanut monenlaisiin uudistuksiin esimerkiksi tuloksellisuusajattelussa ja kirjanpitoikäytännössä (tästä esim. Oulasvirta 2001).

Vastaavasti esimerkiksi Euroopan komissio

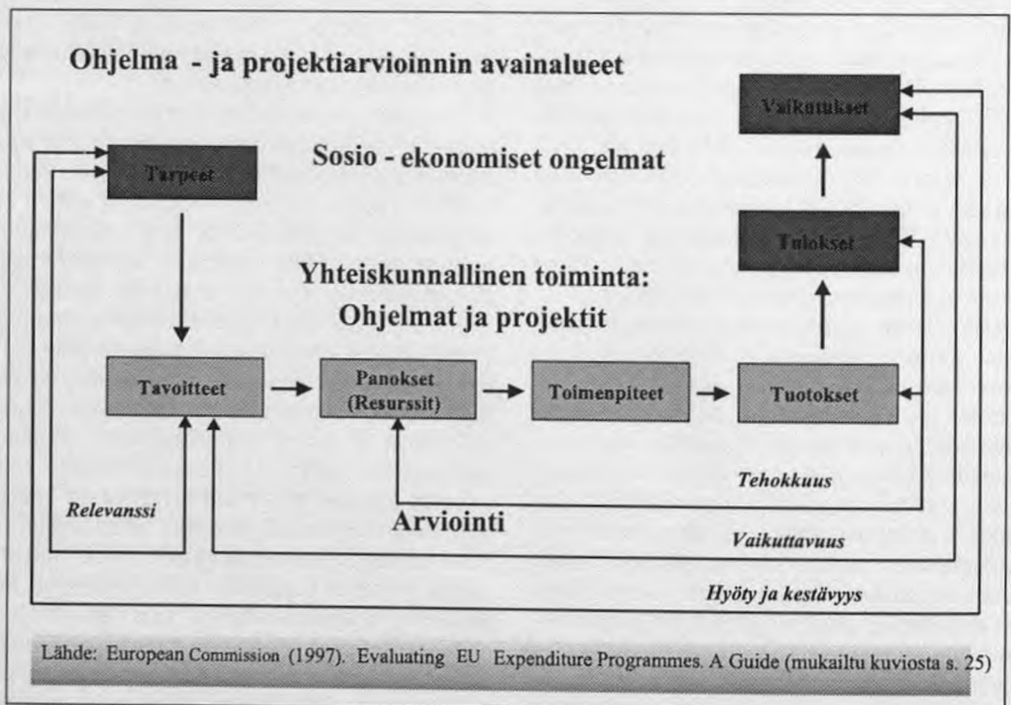
tarjoaa hyvän ja konkreettisen esimerkin seurannan ja arvioinnin välisestä erosta. Esimerkiksi käynnissä olevalla rakennerahastokaudella Euroopan sosiaalirahaston Tavoite 3 –ohjelman seurantajärjestelmien kehittämisestä ovat vastuussa niin sanotut rakennerahastotoiminnan vastuministeriöt (kuten työministeriö, kauppa- ja teollisuusministeriö, sisäasiainministeriö), jotka ovat rakentaneet ohjelmien seurantaa palvelevat tietojärjestelmät, joihin kerrytetään projektikohtaista tietoa kyseisen ohjelman suoritteista. Nämä tiedot – joihin siis sisältyvät ohjelman projekteja koskevat suoritteet (projekteissa aloittaneiden ja lopettaneiden määrät jne.) ja maksatustiedot – raportoidaan varsinaisten ulkopuolisten evaluaattoreiden raporttien kanssa ohjelman seurantakomitealle säännöllisin väliajoin. Ideaalinen tilanne on sellainen, jossa ohjelman evaluaattorit voivat hyödyntää seurantajärjestelmän tuottamaa tilastotietoa, mikä puolestaan edellyttää seurantajärjestelmien tuottaman tiedon luotettavuutta.

Kuviossa 1 on esitetty ohjelma-arvioinnin avainalueet ja kuvion johtopäätökset pätevät soveltuvin osin myös projektiarvioinnin kohdalla. Kuvion

käsitteistä panokset, toimenpiteet ja tuotokset ovat seurannan kannalta keskeisiä, kun taas arvioinnissa ollaan kiinnostuneita tuloksista ja vaikutuksista. Kuviossa mainitut relevanssin, tehokkuuden, vaikuttavuuden ja hyödyn käsitteet ovat puolestaan aivan keskeisiä jälkikäteisarvioinnin tarkastelu-ulottuvuuksia, joihin jälkikäteisarviointia tekevä evaluaattori pitäisi kiinnittää huomionsa.

Ohjelma-arvioinnin kolmas ajallinen ulottuvuus liittyy siis edellä sanotun mukaisesti ohjelmien päättymisvaiheeseen. Tässä vaiheessa analysoidaan se, mitä ohjelmalla saatiin aikaiseksi. Tätä arvioinnin vaihetta kutsutaan joko jälkikäteisarviointiksi tai ex post –arviointiksi. Jälkikäteisarviointiin liittyy paljon ajatuksellisia haasteita, joista keskeisimpiä lienee ohjelman päättymisen ja arvioinnin ajoittumisen välinen ajallisen viiveen huomioon ottaminen. Milloin ylipäänsä on mahdollista järkevää tehdä arviointi? Puoli vuotta, vuosi tai kaksi vuotta ohjelman päättymisestä? Vastaus tähän kysymykseen riippuu ohjelmasta – mitään yleispätevää ohjetta ei voi käsittääkseni antaa.

Suomessa vallalla oleva ohjelma-arviointien



Kuvio 1

toteutuksen tilanne on sellainen, että jälkikäteisarviointien asema on ylivoimainen suhteessa muihin ajallisiin ulottuvuuksiin. Kansallisissa yhteiskuntapolitiisissa ohjelmissa – siis pelkästään kansallisiin budjettivarojen toteutetuissa – etukäteisarvioiteja ei juurikaan toteuteta ja jatkuvatkin arvioinnit ovat enemmän tai vähemmän tapauskohtaisesti viritettyjä asetelmallisessa tai metodisessa mielessä. Tässä on mielestäni iso kehityshaaste kansalliselle ohjelma-arvioinnille. Suomalaiset ohjelma-arvioinnit ovat tähän asti pääsääntöisesti olleet jonkinlaisia yhdistelmiä jatkuvasta tai jälkikäteisarviointista ja kokemukset arviointien hyödyntämisestä ovat vaihdelleet täysin hyödyttömästä erittäin hyödyllisiin. Viime vuosina on valtionhallinnossa alettu kiinnittää huomiota ohjelma-arviointien tilaamiseen liittyvään osaamiseen (Vuorela 1997; Valtiovarainministeriö 1999), mikä osaltaan vuosien saatossa edistää uudentyypin ohjelma-arviointikulttuurin syntyä.

**(iii) Arviointitiedon ja tuloksellisuusinformaation synergiahyötyjen paikantaminen ja arviointitoiminnan ja tuloksellisuusinformaation integroiminen ohjelmien ja projektien ohjaukseen**

Arviointitoiminnan kytkeminen osaksi ohjelmien – ja myös projektitasolla yksittäisten hankkeiden – ohjausta on oikeastaan jo tullut ääneen lausutuksi edellisissä kohdissa (i) ja (ii). Vaikka on tunnettua (ks. esim. Vakkurin 2001 havainnollinen esitys), että tuloksellisuusinformaation tiedonmuodostuksessa ja tiedon hyödyntämisessä on monenlaisia pulmallisuuksia, arviointitoiminnan kautta saadulla tiedolla pitää olla kiinteä yhteys ohjelmien ohjaukseen.

Käytännössä edellä sanottu tarkoittaa arviointitiedon saattamista sellaiseen muotoon, että ohjelmien ohjauksesta vastaava taho – seuranta-, ohjaus- johto- tai muu vastaava ryhmä – voi todella tehdä päätöksiä ohjelmien joko suunnan tarkistuksista tai peräti suunnan muutoksista. Miten tämä sitten voisi tapahtua ja minkä ajatuksellisen viitekehityksen avulla? Vastaus löytyy lähempää kuin saattaa arvata. Tai oikeastaan paikasta, josta vastausta ei ole osattu hakea.

Suomessa nimittäin sovelletaan – ainakin vielä nykyisellään – valitettavan vähän niin sanottua Logical Framework (joskus näkee käytettävän lyhennettä LogFrame) –lähestymistapaa ohjelmatasoisessa suunnittelussa, seurannassa ja

arvioinnissa. (Erityisesti kehitysyhteishankkeiden) projektimaailmassa jo 1970-luvulla yleistynyt tapa mieltää hankesykli sopisi mitä mainoimmin myös ohjelmatasoiseen yhteiskuntapolitiisten interventoiden suunnitteluun, seurantaan ja arviointiin työkaluksi, joka varmasti toisi mukaan ohjelmien toteuttamiseen uutta käsitteellistä ja tilintekovastuuta huokuvaa ryhtiä. Tiedossani ei ole Suomessa montaakaan ministeriötä tai eri hallinnonalan tilivirastoa, joka soveltaisi Logical Framework –ajattelua ohjelmatyössään. Ulkoasiainministeriö – ja siellä erityisesti sen kehitysyhteistyösasto – edustaa tässä kuitenkin poikkeusta.

Kuviossa 2 esitetään Logical Framework –matriisin perusmalli, joka on tuttu useimmista aihealueen perusesityksistä. Tässä yhteydessä voidaan mainita vaikkapa Basil Edward Cracknellin (2000) mainio esitys lähestymistavan perusolettamuksesta ja soveltamisalueista. Ideana on siis eriyttää – tässä tapauksessa – ohjelmatason yleinen päämäärä, varsinaisen tavoite, toiminnot, oletetut tulokset ja oletetut vaikutukset toisistaan ja mieltää näille validit ja tulosindikaattorit, niitä palvelevat tietolähteet (eli verifiointipohja) ja paikantaa yleistä päämäärää, varsinaista tavoitetta, toimintoja ja tuloksia koskevat olettamukset (siis tavallaan riskianalyysin tekeminen).

Näin menetellen ohjelmien ja projektien ohjausfunktio terävöityy ja samalla tulee myös pohdittavaksi ja päätettäväksi seurannan ja arvioinnin väliset erot ja roolit. En esitä Logical Framework –ajattelutapaa tässä yhteydessä mitenkään kevytmielisesti hetkellisen päähänpiston johdosta, vaan uskon olevani riittävän kokenut väittäökseni, että tällaisella yksinkertaisella suunnittelu-, seuranta- ja arviointimekanismilla moni yhteiskuntapolitiinen (ja varsinkin hallitusohjelmatasoinen) ohjelma ja hanke olisi ollut Suomessa paljon järkevämmällä ja uskottavammalla tolalla kuin mitä ne nyt ovat.

Lisäksi voidaan todeta, että näen edellä sanottulla myös laajempia julkisen hallinnon kehitysvaiheeseen liittyviä kontekstuaalisia implikaatioita. Suomen osalta voidaan sanoa, että julkishallinnossa ollaan viime vuosina niin kunnissa kuin valtionhallinnon organisaatioissa modernisoitu luotuneita toiminnan- ja tulosohtausjärjestelmiä, mikä osaltaan on johtanut monenlaisiin uudistuksiin esimerkiksi tuloksellisuusajattelussa ja kirjanpitoikäytännöissä (tästä esim. Oulasvirta 2001).

Vastaavasti esimerkiksi Euroopan komissio

## Logical Framework – perusmatriisi

Toimintalogiikka	Indikaattori	Tietolähteet	Olettamukset
Mihin laajempaan tavoitteeseen politiikalla, ohjelmalla tai hankkeella pyritään?	Mistä tiedetään, millä mitataan?	Mistä tiedot saadaan?	
Mikä on politiikan, ohjelman tai hankkeen keskeinen tavoite, tarkoitus?	Mistä tiedetään, miten mitataan?	Mistä tiedot varmennetaan?	Mitä muuta tarvitaan laajemman tavoitteen edistämiseksi?
Mitä tuloksia tarkoituksen saavuttaminen edellyttää?	Mistä tiedetään, että tulokset on saavutettu? Kuka on tehnyt mitä, missä, kuinka usein?	Mistä tiedot varmennetaan?	Mitä muuta tarvitaan tarkoituksen saavuttamiseksi?
Mitä toimenpiteitä kukin tulos edellyttää?	Kuka tekee, paljonko aikaa ja muita resursseja?	Paljonko maksaa?	Mitä muuta tarvitaan, että tulokset saavutetaan?
			Mitä vaaditaan, että politiikka, ohjelma tai hanke käynnistyy?

### Kuvio 2.

on viime vuosina ajoittaisten skandaalienkin siivittämänä ollut pakotettu vahvistamaan omia toiminnanohjousjärjestelmiään ja hallinnollisia käytäntöjään. Tämä tulee arvatenkin vaikuttamaan EU:n varoin rahoitettujen ohjelmien ja projektien arviointiin. Euroopan komissio on kehittänyt 1990-luvun jälkipuoliskolta alkaen toimintopohjaiseen laskentamalliin (Activity Based Counting) pohjautuvaa johtamistapaansa, joka on alettu kutsua toimintopohjaiseksi hallinnoksi (Activity Based Management). Kyse on modernista tavasta ajatella perinteistä tulosohtausta eräänlaisina tulosketjujen hallintana (suomalaisesta keskustelusta ks. Uusikylä & Valovirta 2002): ABM-mallissa kytketään toisiinsa käsitteellisesti ja toiminnallisesti strateginen suunnittelu ja budjetointi, käytännön ohjelmointityö ja hallinnoiminen, seuranta, raportointi ja arviointi. Euroopan komission toiminnanohjousjärjestelmän painopisteen uudelleen suuntaaminen seurannan ja arvioinnin merkitystä alleviivaamalla tulee lähivuosina vahvistamaan vastaavaa ja itse asiassa jo 1990-luvulla alkanutta kehitystä myös Euroopan unionin jäsenvaltioissa.

(iv) Ohjelma-arviointien hyödynnettävyys paranee niiden vahvan empiirisen pohjan, metodologisen tarkkuuden ja situationaalisen tulkinnan myötä.

Neljäntenä näkökohtana nostan esille kysymyksen, jonka otsikoin väljästi empiiriseksi laaja-alaisuudeksi ja metodologiseksi tarkkuudeksi. Viitataan näistä ensimmäisellä nimenomaan empiiriseen näkökohtaan – siihen, että arvioinnit perustuisivat moninaisiin ja erilaisiin empiirisiin lähteisiin. Viime vuosina on ollut tapana puhua Suomessa yhteiskunta-, sosiaali- ja hallintotieteiden piirissä metodologisesta triangulaatiosta ja viitataan empiirillä laaja-alaisuudella juuri tähän aineistotriangulaatioon (ks. triangulaatiotyypin määrittelystä Denzin 1970). Suomessa arviointitutkimuksen osalta on taas nostettu esille monitahoarvioinnin näkökulma, jolla on nimenomaan haluttu korostaa arvioitavan ilmiön lähestymistä eri toimijoiden näkökulmasta. Kyse on tässäkin tapauksessa ollut aineistotriangulaation soveluksesta denziniläisessä hengessä.

On tietysti syytä kysyä, miksi eri aineistojen kerääminen ja analysoiminen sitten on vaivan-



näön arvoista? Oman kokemukseni perusteella tällainen toimintatapa tuo arviointiin uskottavuutta. Tuloksena on iteroituva kuva tarkasteltavasta ilmiöstä. Otan lyhyen esimerkin omasta arviointihankkeestani edelliseltä ESR-ohjelma-kaudelta. Tehtäväni oli arvioida kokemuksia nuorten työpajaprojekteista. Jos arvioinnissa olisi luottanut kaikkein helpoimmin saatavissa olevaan aineistoon eli kyseisten projektien vetäjien haastatteluun, käsitys kyseisten projektien merkityksestä olisi ollut hyvin myönteinen. Teimme tutkimusapulaiseni kanssa kuitenkin joukoin täydentäviä haastatteluja (itse asiassa haastatelimme ensinnäkin kahta isoa nuorten, EU:n slangissa "hyödynsaajia", ryhmää kyselylomakkeella, ja toisaalta kyseisten projektien keskeisiä yhteistyökumppaneita eli paikallisia pk-yrittäjiä) ja tulos kääntyi, jos ei nyt päälaelleen, niin ainakin kovin toisenlaiseksi.

Olen täysin tietoinen siitä, että kaikki eivät jaa käsitystäni monien aineistojen samanaikaiskäytöstä, koska siihen sisältyy lukuisia pulmia. Mitä tehdä, jos arvioitava kohde näyttää kovin erilaiselta kun sitä katsotaan eri näkökulmista? Kenen näkökulmaa uskotaan? Kenen näkökulmaa painotetaan? Entä miten eri tulkinnat raportoidaan? Vastaukset näihin kysymyksiin ovat aina tapauskohtaisia ja ne vaativat evaluaattorin ammattitaitoa ja tutkimuskohteen olemuksen tunnistamiseen liittyvää sensitiivisyyttä.

Metodologisella tarkkuudella viitataan yhtäällä siihen, että arvioinnissa selvitetään juuri niitä kysymyksiä, joita alun perin on ollut tarkoituskin analysoida. Tässä on keskeistä myös se, että evaluaattori on todella sisäistänyt sen, mitä toimeksiantaja – useimmiten ministeriöt ohjelma-arviointien ollessa kyseessä – on tilannut arviointihankkeen rahoittaessaan. Toisaalta eri arviointikäsitteiden kanssa tulee olla tarkkana. Jos arvioinnin tekijä ei ole varma, mitä hän tarkoittaa kirjoittaessaan panoksista, toimenpiteistä, tuloksista, tuotoksista, vaikutuksista, tarkoituksenmukaisuudesta tai pysyvyydestä (näistä käsitteistä esim. Ahonen 2001b, 75-77), on varmaan melko turhaa odottaa, että myöskään arviointiraportin lukija pysyisi perillä kirjoittajan ajatuksista. Kolmanneksi ei voida myöskään liikaa korostaa arviointikriteerien määrittämisen merkitystä. Evaluaattorin tulee määritellä hyvän, keskinkertaisen ja huonon tuloksen ja vaikutuksen arvo. Ilman tätä ei ole kysymys arvioinnista, vaan korkeintaan analyttisestä kuvailusta.

Kaiken kaikkiaan ohjelma- ja projektiarviointien uskottavuus on riippuvainen monesta tekijästä. Menen tässä nyt vähän kauemmaksi metodologisista kysymyksistä ja siirryn lähemmäksi – taas – ohjelma-arviointien tulkintaa. Arviointitutkimuksia tekevät ovat nimittäin paljon velkaa Ray Pawsonin ja Nick Tilley'n käsitteelliselle innovaatiolle heidän puhuessaan arviointitulosten kontekstisidonnaisuudesta. Kontekstin, mekanismin (toteutettava interventio) ja tuloksen välisen suhteen analyysi (niin sanottu Context – Mechanism – Outcome –suhde) on mielestäni yksi tärkeimmistä oivalluksista, mitä Pawson ja Tilley ovat ylipäänsä esittäneet. Heidän kirjoituksensa ovat herättäneet paljon innostusta ja he ovat saaneet paljon perässäkulkijoita, myös Suomessa. Muutoin realistinen evaluaatio ei ehkä sittenkään pystynyt tuomaan kovin paljon uutta ja ihmeellistä arviointitutkimukseen ja –toimintaan (tietysti heidän fanaattisimmat "opetuslapsensa" ovat tästä eri miellä, mikä on tietysti sallittua; en korrektiussyistä mainitse tässä yhtään lähdeviittausta) tai se ainakin jää vielä nähtäväksi. Arvioinnin tekijän on siis paitsi ymmärrettävä, miten jokin ohjelma ja projekti toimii jossain tietyssä ympäristössä, hänen pitäisi oivaltaa myös se missä olosuhteissa jokin projekti tai ohjelma ei toimi jossain muualla. Tätä voisi kutsua myös situationaaliseksi ohjelma-arvioinnin tulosten tulkinnaksi.

## 6. KATSE TULEVAISUUTEEN

Olen toivottavasti pystynyt esittämään, että ohjelma-arviointi (ja projektien arvioinnissa, kuten tässä artikkelissa olen kuljettanut ohjelma-arvioinnin rinnalla myös projektiarviointia) on vaikea taiteenlaji. Havaitut muutokset eivät ole nettomuutoksia, arvioinnilla on erilaisia motiiveja, arviointia tehdään ohjelman toteutuksen eri vaiheissa ja arvioinnissa hyödynnetään erilaisia empiirisiä tiedonkeruumenetelmiä, empiirisistä havainnoista pitää voida johtaa toimenpidesuosituksia, arvioinnin konkreettiset tutkimusmenetelmät edellyttävät laajaa osaamista, arviointien hyödyntäminen ei ole itsestään selvää ja niin edelleen.

Olen kuitenkin kohtalaisen vakuuttunut siitä, että ohjelma-arviointien tekijöiden kysyntä hallinnossa tulee lähivuosina kasvamaan. Suomen keskushallinnon käynnissä olevat uudistushank-

keet johtavat aikaa myöten uuden ohjelma-ajattelun lisääntymiseen ja poliittikkakeskeiseen (policy-oriented) ajatteluun hallinnon toimeenpanorakenteissa. Milloin muutos toden teolla tapahtuu – sitä en uskalla arvioida – mutta olen varma, että elämme tällaista kehitystä jo nyt parhaillaan julkishallinnossa.

Politiikkakeskeisyys on valtava haaste suomalaiselle julkiselle hallinnolle. Totutut budjettimomentit – ja erityisesti näiden budjettimomenttien "vartijat" – eivät välttämättä ole täysin ihastuneita tästä muutoksen tuulesta. Muutos kun merkitsee paitsi uutta ajattelua myös entisten valtasasetelmien uudelleen virittymistä ja virittämistä. Muutos on haaste myös politiikan tekijöille ja arviointien tekijöille. Miten tarjota uskottavaa, koeteltua ja sovellettavissa olevaa tietoa suomalaisesta yhteiskunnasta? Siinäpä kysymys.

## KIRJALLISUUS

- Ahonen, Pertti: *Evaluatio-oppi edistyville*. Hallinnon tutkimus 20(2001): 3, s. 102-117. 2001a.
- Ahonen, Pertti: *Evaluointi – evaluatio – arviointi: Oppia evaluaattoreille, evaluoittajille, evaluoituttajille ja evaluoitaville*. Teoksessa Myllymäki, Arvo & Vakkuri, Jarmo (toim.): *Tulos, normi, tiilivelvollisuus*. Vammala: Tampere University Press 2001, s. 62-88. 2001b.
- Bouckaert, Geert & Ormond, Derry & Peters, Guy: *A Potential Governance Agenda for Finland*. Research reports 8/2000. MoF: Helsinki 2000.
- Bourdieu, Pierre: *Homo Academicus*. Cambridge: Polity Press 1990.
- Bourdieu, Pierre & Wacquant, Loïc, J.D.: *An Invitation to Reflexive Sociology*. Cambridge: Polity Press 1992.
- Cracknell, Basil Edward: *Evaluating Development Aid. Issues, Problems and Solutions*. New Delhi ym.: Sage Publications 2000.
- Browne, Angela & Wildavsky, Aaron: *Implementation as Mutual Adaptation*. Teoksessa Pressman, Jeffrey L. & Wildavsky, Aaron: *Implementation*. Berkeley: University of California Press 1984, s. 206-231.
- Denzin, Norman: *The Research Act*. New York: Aldine Publishing Company 1970.
- European Commission: *Evaluation EU Expenditure Programmes. A Guide*. Brussels: European Commission, DG Budget. 1997.
- European Commission: *Focus on Results: Strengthening Evaluation of Commission activities*. Communication to the Commission SEC(2000)1051. 26.7.2000. 2000a.
- European Commission: *Reforming the Commission. A White Paper – Part II. Action Plan*. COM(2000) 200final. 2000b.
- Fischer, Frank: *Evaluating Public Policy*. Chicago: Nelson-Hall 1995.
- Gerholm, Tor Ragnar & Magnusson, Sigvard: *Ajatus, aate ja yhteiskunta*. WSOY: Porvoo 1983.
- Kekkonen, Sirpa: *Hallituksen yhteisen poliittisen johtamisen vahvistaminen – keinona ohjelmajohtaminen*. Valtion keskushallinnon uudistaminen. VM: Helsinki 2001.
- Oulasvirta, Lasse: *Valtion kirjanpituudistus ja tiliviraston tilinpäätöslaskelmat*. Teoksessa Myllymäki, Arvo & Vakkuri, Jarmo (toim.): *Tulos, normi, tiilivelvollisuus*. Vammala: Tampere University Press 2001, s. 193-212.
- Patton, Michael Q.: *Utilization-focused Evaluation*. Thousand Oaks ym.: Sage 1997.
- Pawson, R & Tilley, Nick: *Realistic Evaluation*. London ym.: Sage 1997.
- Poliitt, Christopher & Bouckaert, Geert: *Public Management Reform*. Oxford: Oxford University Press 2000.
- Pressman, Jeffrey L. & Wildavsky, Aaron: *Implementation*. Berkeley: University of California Press 1984.
- Uusikylä, Petri & Valovirta, Ville: *Osaoptimoinnista tuloksetkijuihin*. Helsinki: Nebooks. 2002.
- Vakkuri, Jarmo: *Organisaation tuloksellisuusinformaatio tarkastus- ja arviointitoiminnassa: Informaation tuotannon ja käytön näkökulmat*. Teoksessa Myllymäki, Arvo & Vakkuri, Jarmo (toim.): *Tulos, normi, tiilivelvollisuus*. Vammala: Tampere University Press 2001, s. 171-192.
- Valovirta, Ville: "Evaluation Utilization as Argumentation". *Evaluation 8 (2002): 60-80*.
- Valtiovarainministeriö: *Valtionhallinnon arviointityöryhmän loppuraportti*. Valtiovarainministeriön työryhmämuistioita 14/1999. Helsinki 1999.
- Vedung, Evert: *From Value-free Science to Empowerment of the Destitute*. Julkaisematon esitelmä Suomen Arviointiyhdistys ry:n seminaarissa 19.10.2001.
- Vedung, Evert: *Public Policy and Program Evaluation*. New Brunswick & London: Transaction Publishers 1997.
- Virtanen, Petri: *Arviointiosaamisen ulottuvuudet julkisella ja kolmannella sektorilla*. Hallinnon tutkimus 20(2001)2, s. 144-152.
- Virtanen, Petri: *Refleksiivinen evaluatio ja tiedon hyödyntäminen*. Hallinnon tutkimus 17(1998)4, s. 262-271.
- Virtanen, Petri & Uusikylä, Petri & Vartiainen, Saara: *Business Support Programmes. SME Policy Description, Compensation Mechanisms and Evaluation Systems in Five Countries and in the European Commission*. Ministry of Trade and Industry, Studies and Reports 8/2001. Helsinki 2000.
- Vuorela, Terho: *Arvioinnin tilaajan opas*. Helsinki: VM ja HAUS 1997.
- von Wright, Georg Henrik: *Hyvän muunnelmat*. Keuruu: Otava 2001.